Hospital working conditions, doctors' work-related wellbeing, and the quality of care provided : a multilevel perspective

Hospital Working Conditions, Doctors' Work-related Wellbeing, and the Quality of Care Provided: A Multilevel Perspective

Kevin Rui-Han Teoh

This thesis is submitted to Birkbeck, University of London

for the degree of Doctor of Philosophy

Submitted October 2017

Awarded March 2018

This is to confirm that the entire work presented in this thesis is the result of my own work.

Kevin Rui-Han Teoh

October 2017

Abstract

This thesis aims to examine the relationship between the psychosocial working conditions of hospital doctors in England and the quality of care being provided, with work-related wellbeing as a mediator. It applied the job demands-resources model to this occupational sample, and utilised a multilevel perspective to include trust-level demands and outcomes.

In the first of four studies - a systematic review and meta-analysis found that across 21 studies, 62% of the reported relationships between job demands and 64% of job resources' relationships with quality of care were significant; the presence of these relationships varied by the type of outcome measure used. A lack of theoretical grounding within these studies emphasised the need to frame these relationships within a theoretical framework.

The three subsequent empirical studies drew on composite scales from the 2014 NHS Staff Survey in England. Across these multilevel studies, job demands (insufficient work resources, workplace aggression) predicted negative work-related wellbeing (presenteeism, work-related stress), while job resources (manager support, job control, effective teams) predicted work engagement. Trust-level demands (number of emergency admission, bed occupancy rate) also predicted hospital doctors' work-related wellbeing. No interactions were observed between job demands and resources. Work-related wellbeing mediated most relationships between job demands and resources with individual self-rated quality-of-care measures. Some mediations involving patient satisfaction with doctors were found, but not for hospital mortality or patient safety incidents.

The research reported in this thesis highlights the complexity of work-related predictors to hospital doctors' work-related wellbeing and the quality of care provided. It further demonstrates that these outcomes are a product of their wider work context. Successful interventions should target the appropriate antecedent pathway, and recognise trust and system factors. The job demands-resources model can be useful in explaining individual-level relationships, but is limited when including trust-level measures. Further implications on research, practice, and policy are discussed.

Acknowledgements

In finishing this thesis, I am indebted to many.

To my supervisors Dr Juliet Hassard and Professor Tom Cox, thank you for your guidance, support, expertise, and friendship over the past four years. This extends not only in finishing this thesis, but in my all round development towards becoming a researcher.

I am grateful to the Department of Organizational Psychology, Birkbeck University of London, for the studentship that allowed me to pursue this thesis. Here, I would also like to thank for their encouragement and advice: Professor Philip Dewe, Professor Linda Trenberth, Dr Katrina Pritchard, Professor Pat Tissington, Dr Almuth McDowall, and Jean Devaney. My gratitude also goes to Dr Rebecca Whiting for the constant reminder that "the only good PhD is a finished PhD"; and to Dr George Michaelides and Dr Duncan Jackson for their statistical advice.

In addition, I must thank Dr Aditya Jain, Dr Sue Cowan, and Aaron Percival for acting as quality reviewers in the systematic review and scale development stages. I am also indebted to the many who assisted with proofreading in the final weeks: Andrew, Jules, Lucia, Liam, Jeremiah, Joanne, Chris, Asta, and Gintare.

More personally, I was blessed to have been surrounded by an excellent squad of peers on this PhD journey. From general banter, shared exasperation, resource sharing, and subject-matter support, it would have been very lonely without all of you. To Mum and Dad, I would not be here today without all your love, energy, sacrifices, and opportunities that you have made for me. To Jules and Dominique, I have always been proud of you both. You both have always inspired me with all that you do, and long may that continue. To my Dasha, thank you for putting up with the "eternal student", and for supporting me through this with love, laughter, and lameness. Lastly, Milo must be thanked, for being amusing at all hours of the day.

- Thank you - 谢谢 - Terima Kasih - Danke Schön - Спасибо -

Table of Contents

List of Tables

List of Figures

Table 0.1: *Table of Acronyms*

Acronym	**Definition**
A&E	Accident and Emergency
AIC	Akaike's Information Criterion
BMA	British Medical Association
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CQC	Care Quality Commission
df	Degrees of Freedom
EFA	Exploratory Factor Analysis
EWTD	European Working Time Directive
HSE	The United Kingdom Health and Safety Executive
ICC	Intraclass Correlation
JD-R	The Job Demands-Resources Model
MI	Modification Indices
ML	Maximum Likelihood estimator
MSEM	Multilevel Structural Equation Modelling
NHS	National Health Service
PCA	Principal Component Analysis
PSI	Patient Safety Incidents
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Model
SHMI	Standardised Hospital-level Mortality Indicator
TLI	Tucker-Lewis Index
WLS	Weighted Least Square estimator

Chapter 1 : Introduction

The aftermath of the Mid-Staffordshire scandal and the general review of patient care standards in the United Kingdom have led to the re-examination of the role that healthcare staff play in the delivery of care and the general state of their wellbeing (Francis, 2013; Keogh, 2013). Various studies highlight the vulnerable state of doctors' wellbeing, as they experience higher rates of psychological distress and burnout than the general population (Montgomery, Panagopoulou, Kehoe, & Valkanos, 2011; Prins, Gazendam-Donofrio, et al., 2007). This concern has given rise to a small but growing literature where doctors' wellbeing is seen as an outcome of challenging working conditions (Bernburg, Vitzthum, Groneberg, & Mache, 2016; Lee, Seo, Hladkyj, Lovell, & Schwartzmann, 2013; Scheurer, McKean, Miller, & Wetterneck, 2009), and where doctors are seen as under-resourced to work long hours with heavy workloads in difficult and even abusive environments. However, despite many advocating the improvement of hospital doctors' working conditions to improve quality of care provided (Royal College of Physicians, 2016), the reality is that there is very little evidence demonstrating this relationship. Instead, the underlying assumption relies on two separate research links: between working conditions and doctors' wellbeing; and between doctors' wellbeing and quality of care (Hall, Johnson, Watt, Tsipa, & O'Connor, 2016).

The purpose of this chapter is to introduce the thesis, which is aimed at understanding the relationship between psychosocial working conditions of hospital doctors and the quality of care being provided. To do so, this chapter is structured in four parts: (i) a description of the role of hospital doctors in the United Kingdom and the contemporary demands on the National Health Service (NHS); (ii) a review of the research gaps in the relationship between the psychosocial working conditions of doctors and quality of care; (iii) a clarification of the key constructs in this thesis; (iv) the presentation of the aims of this thesis; and (v) a description of the chapters in this thesis.

1.1 Hospital Doctors, What Do They Do?

The primary role of hospital doctors is to take responsibility and provide leadership for patient care (Godlee, 2008; McKay & Narasimhan, 2012). This involves complex decision making, patient communication, coordination of multidisciplinary teams, and extensive periods of medical training. The crucial difference between doctors and other healthcare professionals (e.g., nurses) is that ultimate responsibility for this care lies with doctors alone (Medical Schools Council, 2014). They are consequently a key source of influence within the healthcare sector, particularly considering their size as the smallest occupational group within the NHS (NHS Staff Survey Co-ordination Centre, 2015). Recognising the pivotal role doctors play in healthcare delivery, there have been considerable calls by various stakeholders for the need to better understand the multitude of factors that pertain to the performance of doctors (Bloor, Freemantle, & Maynard, 2012; Francis, 2013; Keogh, 2013).

Hospital doctors can broadly be distinguished into two groups: junior doctors and consultants. In the United Kingdom, junior doctors include both foundation year doctors and registrars. They would have completed medical school but are still undergoing postgraduate training; a period of time upwards of eight years post-medical school (Royal College of Physicians, 2016). These doctors form the backbone of the NHS and operate on the frontlines of healthcare alongside colleagues from other disciplines (Keogh, 2013). Junior doctors provide new perspectives, knowledge of contemporary medical curriculum developments (Victoria State Government, 2013), and have been identified as potential agents of change in the drive to improve patient care (Bagnall, 2012; Health Foundation, 2011) and develop distributive leadership in the NHS (Martin & Learmonth, 2012). However, their junior status and challenging working conditions means junior doctors' wellbeing, and potential for contributing to an improved health service, are frequently ignored or neglected (Iversen, Rushforth, & Forrest, 2009; Joyce, Schurer, Scott, Humphreys, & Kalb, 2011; McGowan, Humphries, Burke, Conry, & Morgan, 2013).

In comparison, consultants are the highest skilled group of doctors, representing the most experienced and trained section of the medical workforce (Academy of Medical Royal Colleges, 2012). The breadth, depth, and length of their training and experience allow consultants to make quick and relevant decisions for patient care. Their role as educators and supervisors of the next generation of doctors and nurses means they are imperative in the development of the healthcare workforce (Peadon, Caldwell, & Oldmeadow, 2010). Consultants' responsibilities extend beyond direct patient care, and include (Bloor et al., 2012): supporting professional activities (e.g., supervision, training), additional NHS duties (e.g., administrative and management roles), and external responsibilities (e.g., duties with the Royal Colleges). The experience and specialities of consultants have led to various national political initiatives towards "consultant-led" and "consultant-driven" services (Academy of Medical Royal Colleges, 2012; Goddard, 2010). Despite this, growing evidence highlights disillusionment among consultants about the various initiatives involving them, and concern about how to best improve quality of care through them (Bloor et al., 2012; Davies, Hodges, & Rundall, 2003; Russell, Wyness, McAuliffe, & Fellenz, 2010).

1.2 The NHS Context

According to the Royal College of Physicians (2016), contemporary national-level events have conspired to make an already inherently challenging job for hospital doctors even more difficult. The growing population places continual strain on healthcare services, while an aging population presents additional challenges through more complex health and support needs. Currently, a quarter of the population live with a long-term health condition (Department of Health, 2017), utilising 50% of all general practice appointments, occupying 70% of hospital beds, and costing 70% of the primary and hospital care budgets. The number of people with a long-term health condition are projected to continually rise (Department of Health, 2017). In addition, there have been substantial increases in acute hospitals admissions (30.4%) and operations (44%) between 2004 and 2014 (HSCIC, 2015a). Attendance in hospital accident and emergency (A&E) departments also increased by 39.9% from 14 million visits in 2004 to 19.5 million in 2015 (HSCIC, 2016). Dealing with these demands is not made easier by the chronic shortage of various healthcare staff (Buchan, Seccombe, & Charlesworth, 2016; Cylus et al., 2015).

In trying to manage these demands, the NHS in 2015-16 developed a budget deficit of £1.85 billion, a threefold increase from the previous year (Dunn, McKenna, & Murray, 2016). There is also concern that these figures were associated with drastic cost-saving measures, resulting in less wiggle room for additional savings that will have to be made (Dunn et al., 2016). Account reviews suggest that trusts are funnelling money from capital improvement projects to cover costs (King's Fund, 2016), meaning less investment for the larger projects needed to sustain current and future infrastructure. The NHS budget has effectively been frozen since 2010 (Gainsbury, 2016). Consequently, the discrepancy between resources and demands will continue to grow, as the annual budget increase of 0.2% does not compare to the annual 4% growth rate of demands. The maintenance of current funding levels may appear generous given cuts to other public services. However, it has been argued that cuts elsewhere have resulted in reduced capability to support vulnerable members of society, who instead have no option but to turn to the NHS instead (Royal College of Physicians, 2016).

1.3 Working Conditions and Quality of Care

Recent high-profile scandals, notably events involving Mid-Staffordshire (Francis, 2013) and Morecambe Bay Trusts (Kirkup, 2015), highlighted unacceptable failures in the delivery of care. Challenging working conditions, incapable leadership, a climate of bullying, and inadequate staffing have been identified as antecedents to these particular events. The Francis Inquiry (2013) advocated that junior doctors act as the "*eyes and ears*" (pg. 60) of the NHS, while the Keogh review (2013) wrote that the energy of junior doctors should be "*tapped and not sapped*" (pg. 5). More broadly, both reviews identified that doctors' concerns about their working environment should be taken as early indicators of serious underlying issues within hospitals. However, government initiatives over the last several years to reduce costs and to improve seven-day health coverage has raised significant discussion surrounding the working conditions of doctors and its ramifications for patient care (Bagenal, Moberly, & Godlee, 2015; Goddard, 2016). Concerned about their working conditions and the negative impact that poor conditions would have on quality of care, junior doctors in England took the unprecedented decision to strike, not once, but six times in 2016 (Pym, 2016).

Troubled by the lack of doctors and the difficult working environments that their doctors face, the Royal College of Physicians (2016) argued that together these "increase pressure on hard-working NHS staff, put patients at risk, and threaten the future of the NHS" (pg. 6). Researchers have sought to highlight the importance of understanding the consequences of challenging working conditions for doctors (Arnetz, 2001; Shackelton et al., 2010; Taylor, Graham, Potts, Richards, & Ramirez, 2005), while at the same time calling for more research into the antecedents of quality of care (Dollarhide et al., 2013). However, while this relationship may appear intuitive, in reality several weaknesses exist which test the assumption for this relationship.

First, research into the working conditions of doctors has traditionally focused on the structural aspects of their work, including: the number of hours worked (Moonesinghe, Lowery, Shahi, Millen, & Beard, 2011), caseload (Harley et al., 2013; Vree, Cohen, Chavan, & Einarsson, 2011), and staffing levels (Lang, Hodge, Olson, Romano, & Kravitz, 2004; Needleman, Buerhaus, Mattke, Stewart, & Zelevinsky, 2002). An alternate approach is to examine psychosocial working conditions. This focuses on the individual's perception of their work environment (Parkes & Sparkes, 1998). Although this approach has received less attention than structural aspects of work, psychosocial working conditions have consistently been found to be the better predictor of staff wellbeing and patient care outcomes (P. Tucker, Bejerot, Kecklund, Aronsson, & Åkerstedt, 2015; Visser, Smets, Oort, & de Haes, 2003).

Second, much of the evidence relating to this relationship for doctors is drawn from research involving their nursing counterparts (Krueger, Funk, Green, & Kuznar, 2013; Wong & Cummings, 2007) or from multidisciplinary samples (Hall et al., 2016; Hoff, Jameson, Hannan, & Flink, 2004). For example, Krueger et al.'s (2013) review found that high job demands and unfavourable work schedules for nurses were associated with increased patient mortality and complications. Alternatively, advocates for such a relationship explain this relationship using the separate evidence linking the psychosocial working conditions of doctors with their own wellbeing, and between doctors' wellbeing and quality of care. However, few have sought to examine all these constructs within the same framework, or to test the role of doctors' wellbeing as a mediator within this relationship (Weigl, Schneider, Hoffmann, & Angerer, 2015). As the roles of doctors differ from their health colleagues, there have increasingly been calls for more research to specifically test the relationship between doctors' psychosocial working conditions and quality of care (J. Klein, Frie, Blum, & von dem Knesebeck, 2011; Michtalik, Pronovost, Marsteller, Spetz, & Brotman, 2013; Wallace, Lemaire, & Ghali, 2009).

The third issue lies in the lack of theory explaining why, and how, psychosocial working conditions of doctors predict the quality of care. Despite the literature surrounding work-related stress and psychosocial working conditions being replete with studies introducing, testing, and validating theory (Cox, 1993; Cox, Griffiths, & Rial-González, 2000; Häusser, Mojzisch, Niesel, & Schulz-Hardt, 2010), this has not happened in relation to doctors. This may be due to most studies involving doctors being published in the medical literature. Traditionally, this literature has been more concerned with factual outcomes and less about theory (Alderson, 1998). This lack of theoretical framework hampers the development of appropriate interventions.

The fourth issue centres on the absence of testing this relationship from a multilevel perspective. Perceptions of working conditions, as well as reports of doctors' wellbeing, are constructs that operate at the level of the individual. However, quality of care is something that can be constructed at the individual (e.g., number of errors made, self-reported quality provided), or the trust-level (e.g., mortality rates). Similarly, there needs to be greater acknowledgement that all these constructs are situated within a wider system. This means that the demands a trust faces can have implications on both the work-related wellbeing of hospital doctors and the level of patient care being provided. A better understanding of how such constructs interact across different levels will further theoretical development with clearer practical implications; explaining how events at the trust-level influence individual doctors, and how individual doctors' perception of their work environment could influence care quality at the trust-level. This in turn opens up the possibility of evidence-based organisational interventions.

As such, there are a number of reasons as to why the assumption that better psychosocial working conditions of doctors are associated with better quality of care is a weak proposition. Addressing the identified research gaps forms the basis of this thesis, which will in turn enhance our understanding of this important relationship.

1.4 Clarifying Key Constructs

Before introducing the thesis' aims, research questions, and structure, the three core constructs that form the backbone of this thesis need to be defined. This is due to "psychosocial working conditions", "work-related wellbeing", and "quality of care" all being portmanteau constructs that are not consistently defined or operationalised in the literature. These three constructs are briefly introduced in the paragraphs below. Together with other key terms within this thesis, they are also defined in Table 1.1.

The focus of this thesis is on the psychosocial working conditions of hospital doctors, which refers to the perception of the individual of how their work environment is managed, organised, and designed, and the social environment context in which it is situated (Cox, 1993; Cox et al., 2000; Parkes & Sparkes, 1998). This approach does not focus on the structural aspects of work, such as number of hours worked (Moonesinghe et al., 2011) and staffing levels (Lang et al., 2004; Needleman et al., 2002). It is broader than the oft used psychosocial hazards (Cox & Griffiths, 1995), which focuses on how these aspects of work "may have the potential to cause psychological or physical harm" (pg. 69, Cox & Griffiths, 1995). The shortcoming of the term psychosocial hazards lies not only in that it can include structural aspects of work, but that it fails to recognise the potential positive aspects of the work environment (Cox, Karanika-Murray, Griffiths, Wong, & Hardy, 2009).

It has been argued that among the various taxonomies offered over the last 40 years (e.g., Cox, 1993; HSE, 2017), psychosocial working conditions can be divided into two broad categories: job demands and job resources (Bakker & Demerouti, 2007; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001). Job demands are the organisational, physical, social, and psychological aspects of the job that require sustained physical or mental effort that incur a physical or psychological cost. These include perceived workload, workplace bullying, job insecurity, and role ambiguity. In contrast, job resources encompass those work aspects that help attain work goals, stimulate personal development, and reduce the negative impact of job demands. These include social support, autonomy, and good leadership. Structural and psychosocial aspects of work are discussed in more detail in Sections 2.1.1 and 2.2 respectively.

The second core construct is that of work-related wellbeing (see Section 2.2 for an introduction to the construct). When situated in the occupational context, work-related wellbeing is defined as a multidimensional concept that includes affect, motivation, behaviour, cognition, and psychosomaticism (van Horn, Taris, Schaufeli, & Schreurs, 2004; Warr, 1994). The term "health" is not used here, as the World Health Organization's (1948) definition includes physical health, which is not the focus of this thesis. Work-related wellbeing therefore offers a much broader perspective than physical or mental health. It does not merely represent the absence of illness or infirmity, but exists on a continuum encompassing both negative and positive constructs (Bakker & Schaufeli, 2008). This includes burnout and ill-health on one end, and happiness, flourishing, and thriving on the other (Hall et al., 2016; Shanafelt, Sloan, & Habermann, 2003; Wallace et al., 2009). Although the focus within this thesis is on work-related wellbeing, the introductory chapters may use the more general wellbeing term when this includes reference to non-work related wellbeing measures (e.g., depressive symptoms). The studies within this thesis focus specifically on three work-related dimensions of wellbeing: affect, as measured by work-related stress; motivation, represented with work engagement; and behaviour, where presenteeism was the proxy. The definitions for these constructs are presented in Table 1.1.

The third core construct of this thesis is quality of care, which is reviewed in greater depth in Section 3.1. Not only are there different perspectives on what constitutes quality or a good outcome, but these are often dependent on the stakeholders involved (King's Fund, 2011; Wong & Cummings, 2007). The definition used within this thesis is that put forward by the Department of Health (2008; 2010). This defines quality of care as consisting of three core aspects: clinical excellence, patient safety, and the experience of patients. Clinical excellence is defined as preventing premature deaths, enhancing quality of life, and assisting recovery. The second core aspect - patient safety - aims to provide a safe care environment without avoidable harm. The final aspect of quality of care is patient experience, which refers to patient's experience of their personal care and treatment. All three are crucial in the delivery of good care.

Table 1.1: *Definitions of key constructs and terms*

Construct/Term	Definition
Acute trust	An NHS trust that delivers secondary care mainly through one or more hospitals.
Acute specialist trust	A NHS trust managing hospitals that provide specialist care (e.g., paediatrics, oncology, orthopaedics).
Clinical excellence	One aspect of quality of care. It is about preventing premature deaths, enhancing quality of life, and assisting recovery.
Hospital	An individual institution that provides medical or surgical treatment to sick or injured people. A hospital would be under the purview of a NHS trust.
Job demand	The organisational, physical, social, and psychological aspects of the job that require sustained physical or mental effort that incurs a physical or psychological cost.
Job resource	The organisational, physical, social, and psychological aspects of the job that help attain work goals, stimulate personal development, and reduce the negative impact of job demands.
Negative work-related wellbeing	Negative manifestations of work-related wellbeing (e.g., burnout, work-related stress, presenteeism).
[NHS] trust	Organisations within the NHS that provide a service to a geographical area, and can amongst others, be distinguished into acute, ambulance, and mental health trusts. These reflect the main services that are offered by the trust.
Quality of care	The Department of Health defines quality of care as meeting three core aspects: clinical excellence, patient safety, and patient experience.
Patient safety	The aim to provide a safe care environment without avoidable harm.
Patient experience	The quality of the patient's experience of their personal care and treatment.
Positive work-related wellbeing	Positive manifestations of work-related wellbeing (e.g., job satisfaction, work engagement)
Presenteeism	Coming to work despite the worker not feeling well enough to perform their duties.

Psychosocial working environment	Perception of the individual towards aspects of work design, the organisation and management of work, and their social and environmental context.
Work engagement	A positive, fulfilling, work-related state of mind characterised by vigour, dedication and absorption.
Work-related stress	The response people may have when presented with work demands and pressures that are not matched to their knowledge and abilities, and which challenge their ability to cope.
Work-related wellbeing	A multidimensional concept that includes affect, motivation, behaviour, cognition, and psychosomaticism.

1.5 Thesis Aims

This thesis has three core aims based on the research gaps identified above. The first is to enhance the understanding of the relationship between the psychosocial working conditions of hospital doctors and quality of care. This will be based on (i) systematically reviewing the literature to examine the current understanding of this relationship; and (ii) testing this relationship through a representative sample of hospital doctors drawn from every acute and specialist NHS hospital trust in England. This provides a systems perspective where both demands and outcomes at the trust-level are included in this model. Furthermore, the role of hospital doctors' work-related wellbeing as a mediator within this relationship will also be examined.

The second aim of the thesis is to validate this relationship within a theoretical framework, namely the job demands-resources (JD-R) model (Bakker & Demerouti, 2007; Demerouti et al., 2001). This represents a more contemporary wellbeing theory that builds on some of the limitations of earlier stress theories (See Section 5.1 for how the JD-R model is situated within the historical setting; Cox, 1993; Cox et al., 2000; Lazarus, 1966). More specifically, as an example of an interactional theory within the psychological approach (Cox & Griffiths, 2010; Cox et al., 2000), it focuses on the interaction between the individual and their work environment. Crucially, the JD-R model not only accounts for both job demands and resources in the workplace, but recognises that work-related wellbeing can manifest positively (e.g., work engagement) or negatively (e.g. burnout). At its core, this model postulates that every work environment contains psychosocial working conditions categorised as either a demand or a resource to the individual (Bakker & Demerouti, 2017). The model maps upon much of the research literature in this area, and explains how individual wellbeing mediates psychosocial working conditions' influence on individual and organisational-level outcomes.

The third aim is to test the proposed model by linking within the analyses, existing sources of data collected within the healthcare sector. The trend towards secondary data analysis is a reflection of research councils' desire for more economical research which are not only cheaper to conduct, but increase the utility of existing datasets (Teoh, 2016). Therefore, the findings of this thesis could add to the understanding, and potential value, of real-world healthcare data which informs the decision- and policy making of Government, the NHS, and various other relevant stakeholders.

Specific research questions and hypotheses are presented later in the individual chapters reporting on the respective studies. Based on the three aims above, the overarching research question asks whether hospital doctors' psychosocial working conditions (i.e., job demands and resources) influence the quality of care being provided to patients, and if work-related wellbeing (i.e., work-related stress, presenteeism, work engagement) functions as a mediator within this relationship. In validating the job demands-resources model within this sample of hospital doctors, the following specific research questions are asked (See Chapter Five):

i. Do hospital doctors' job demands uniquely predict negative work-related wellbeing; and do job resources uniquely predict positive work-related wellbeing?
ii. Will hospital doctors' job resources moderate the relationship between job demands and negative work-related wellbeing?
iii. Will hospital doctors' job demands moderate the relationship between job resources and positive work-related wellbeing?
iv. Does work-related wellbeing mediate the relationship between hospital doctors' psychosocial working conditions and quality of care provided?
v. Will trust-level demands have the same impact within the JD-R model as that of hospital doctors' job demands?
vi. Will hospital doctors' psychosocial working conditions and work-related wellbeing predict trust-level quality-of-care outcomes?

1.6 Thesis Structure

This section outlines each chapter within this thesis. It describes the purpose of each chapter and how it aligns with the overall aims of this thesis.

Chapter Two describes research on the contemporary working conditions faced by doctors. It distinguishes between structural and psychosocial aspects of work, and demonstrates that the latter is a stronger predictor of doctors' work-related wellbeing and quality-of-care outcomes. As the direct relationship between psychosocial working conditions and quality of care is not clear, this chapter focuses on first establishing the relationship between doctors' psychosocial working conditions and their own levels of burnout, work-related stress, job satisfaction, work engagement, and presenteeism.

Chapter Three develops the relationship described in Chapter Two by reviewing the relationship between doctors' work-related wellbeing and quality of care. It introduces what quality of care means, differentiates between individual and organisational-level quality-of-care outcomes, and examines their respective work-related wellbeing antecedents. This introduces the concept of multilevel analysis as specified in the thesis aims.

The first study of the thesis is described in **Chapter Four** (Study 1) – a systematic review and meta-analysis examining which psychosocial working conditions faced by doctors have been examined in the literature, and what impact they have on different quality-of-care outcomes. The review highlights the lack of theoretical grounding within the reviewed studies, with more work needed to examine potential mediators and moderators that can help understand the context of these relationships. This study has been presented at three conferences and a manuscript has been submitted to the *Work and Stress* journal.

Chapter Five introduces the job demands-resources model (JD-R; Bakker & Demerouti, 2007; Demerouti et al., 2001). The chapter defines what job demands and resources are and how they interact to predict work-related wellbeing and subsequently performance outcomes. The chapter concludes by examining the application of the JD-R model in the healthcare sector and from a multilevel perspective.

Chapter Six describes what secondary data research is, and its suitability to this thesis. This includes a discussion of ethical issues situated within this thesis. The chapter then introduces and describes the main dataset for the subsequent studies, the 2014 NHS Staff Survey for England. This is followed by discussing the utility in creating composite scales, including the importance of validity and reliability in measurement.

Chapter Seven describes the process and results of developing composite measures from the NHS Staff Survey and evaluating their psychometric properties. An exploratory and confirmatory factor analysis was respectively carried out on two separate samples of 7,033 hospital doctors, yielding a measurement model comprising eleven measures.

Chapter Eight (Study 2) examines the main propositions of the JD-R model (Bakker & Demerouti, 2007), and focuses on the relationship between psychosocial working conditions and doctors' work-related wellbeing. A multilevel model examines whether job demands and resources respectively predict negative (i.e., work-related stress, presenteeism) and positive (i.e., work engagement) work-related wellbeing. It also tests whether job demands and resources interact with each other in relation to these outcomes. The multilevel model included demands experienced by the trust, represented by bed occupancy rate and the number of patients in A&E. This study has been presented at two international conferences, and is currently being reviewed by the *Health Care Management Review* journal.

Chapter Nine (Study 3) expands the direct relationships observed in Chapter Eight - between job demands and resources with work-related wellbeing - to include quality of care. It also examines the mediating role of doctors' work-related wellbeing between psychosocial working conditions and three doctor-rated quality-of-care measures: the number of errors seen, quality of individual care, and quality of trust care.

Study 4 is described in **Chapter Ten**. This replaces the quality-of-care outcomes measured in Chapter Nine with trust-level outcomes. The three trust-level outcomes represent the three aspects of good quality of care, namely: clinical excellence (the summary hospital mortality indicator), patient safety (the number of patient safety incidents), and patient experience (patient satisfaction with their doctors).

Chapter Eleven represents the last chapter in this thesis, and draws together the outcomes of each study and how they relate to the original thesis aims. It discusses the validity of the job demands-resources model within this sample of hospital doctors, and what the implications of this thesis are from a policy, methodological, and practical perspective.

Chapter 2 : Working Conditions and the Work-related Wellbeing of Doctors

This chapter introduces the contemporary working conditions faced by doctors. It distinguishes between structural and psychosocial aspects of work, and demonstrates that the latter is a stronger predictor of doctors' work-related wellbeing. This is followed by a review on the current state of doctors' work-related wellbeing. Chapter One highlighted that the direct relationship between psychosocial working conditions and quality of care is not clear; the current chapter therefore focuses on first establishing the relationship between doctors' psychosocial working conditions and their own work-related wellbeing. Chapter Three will then link doctors' work-related wellbeing with quality of care.

2.1 Working Conditions of Doctors

The continuous exposure to human suffering and death can be a major source of distress and trauma for doctors and they can struggle to cope with these emotional demands (Baldisseri, 2007; Luthy, Perrier, Perrin, Cedraschi, & Allaz, 2004). While research has focused on mitigating these demands, much of the organisational research in this sector focuses on how work is structured; typically manifesting as working hours (Moonesinghe et al., 2011), shift work (Reed, Fletcher, & Arora, 2010), caseload (Harley et al., 2013; Vree et al., 2011), and staffing levels (Lang et al., 2004; Needleman et al., 2002). Both in the United Kingdom and beyond, the healthcare sector today is having to balance a reduction in resources allocated to it paired with additional demands expected from it (Goddard, 2016; Royal College of Physicians, 2016). This has resulted in more demanding structural aspects of work for doctors (Arnetz, 2001; Sargent, Sotile, & Sotile, 2004). From a staffing perspective, the UK already has a lower doctor-to-people ratio than the EU average, and the lowest ratio in Western Europe (Cylus et al., 2015). Doctor vacancies have also risen by 60% between 2013 and 2015 (Hughes & Clarke, 2016). Consequently, 70% of junior doctors and 84% of consultants report a permanent shortage on their work rotas. The issue is compounded by similar understaffing in nursing which results in increased workloads for doctors as they take on the duties of nurses while at the same time not receiving the support they need (Buchan et al., 2016).

2.1.1 The impact of structural aspects of work

Previous studies have sought to explore the effects of demanding structural aspects of work on doctors' wellbeing and patient care (Harley et al., 2013; Moonesinghe et al., 2011; Reed et al., 2010). However, the evidence on this is not conclusive. In terms of working hours, one systematic review of 34 studies found that in the majority of studies (*n*=28) no effect was observed between reducing working hours and quality of care (Moonesinghe et al., 2011). Further reviews link long working hours to increased risk of subcutaneous injuries and motor vehicle accidents, but not with general health or mood disorders (Reed et al., 2010; Rodriguez-Jareno et al., 2014). This inconsistent relationship is also seen with objective workload, where increased caseloads have related positively (Chen, Liu, Lin, & Lien, 2008; Yasunaga et al., 2009), negatively (Harley et al., 2013), or not at all (Ansmann et al., 2013; Vree et al., 2011) with better patient outcomes.

The ambiguity as to the presence of relationships between structural aspects of work with doctors' wellbeing and patient care does not mean that these do not exist; instead, other factors may be influencing these relationships (Moonesinghe et al., 2011). Researchers have become increasingly aware that far less attention has been given towards the experiences of doctors to their working conditions (Watt, Nettleton, & Burrows, 2008). How structural aspects of work are cognitively processed influences whether it is perceived as a positive or negative aspect of work to the individual (Jagsi & Surender, 2004; Williams et al., 2002). For example, Swedish doctors who did not feel they had control over their work patterns reported poorer sleep quality than those who perceived having that control (P. Tucker et al., 2015). The importance of this perception is underlined in a study of Dutch medical specialists (Visser et al., 2003). It found that perceived working conditions explained a substantially larger proportion of the variance in job satisfaction (34%) and job stress (24%) compared to personal and job characteristics (between 2%-6%). Hence, while structural aspects of work are still important to doctors, studies need to understand how these working conditions are perceived by doctors, and crucially, how they relate to doctors' work-related wellbeing and the quality of care being provided.

2.2 Psychosocial Working Conditions

The difficulties in establishing a direct relationship between structural aspects of work and outcome measures provides a rationale for examining how these work aspects are perceived (Jagsi & Surender, 2004; Williams et al., 2002). Parkes and Sparkes (1998) provide a useful distinction between socio-technical and psychosocial characteristics of work. The former addresses the structural aspects of work, including doctors' working hours (Moonesinghe et al., 2011), shifts (Reed et al., 2010), and caseload (Lin & Lee, 2008). Psychosocial working conditions within the context of this thesis refer to what Parkes and Sparkes consider psychosocial, that is the worker's perception of their work environment. This builds on to Cox and Griffiths' (1995) definition of psychosocial hazards by emphasising the perception of "aspects of work design and the organisation and management of work, and their social and environmental context" (pg. 69). However, the second part of this definition, "… which may have the potential to cause psychological or physical harm" is discounted as it only focuses on work's potential in creating harm, and not its ability to improve workers' motivation and affect.

Numerous taxonomies of psychosocial working conditions exist (e.g., Cox, 1993; HSE, 2017). One perspective is that psychosocial working conditions can be divided into two broad categories: job demands and job resources (Bakker & Demerouti, 2007; Demerouti et al., 2001). Job demands encompass the organisational, physical, social, and psychological aspects of the job that require sustained physical or mental effort. Examples include job insecurity, perceived workload, and role conflict. The literature has provided consistent and robust evidence demonstrating that in the general population, prolonged and chronic exposure to excessive job demands leads to poorer health and wellbeing, including mental disorders (Stansfeld & Candy, 2006), cardiovascular (Kivimäki et al., 2002), and coronary heart diseases (Kivimäki et al., 2006; Kuper & Marmot, 2003).

In contrast, job resources are those work aspects that help attain work goals, stimulate personal development, and reduce the negative impact of job demands (Demerouti et al., 2001). These include social support, autonomy, and good leadership. A large repertoire of resources would mean doctors are better equipped to handle the challenging aspects of their work, highlighting job resources as a health protecting factor. Moreover, job resources allow workers to feel supported, competent and in control of their working environment (Bakker & Demerouti, 2007; Schaufeli & Bakker, 2004).

Research has consistently identified two psychosocial working conditions that are salient in the hospital work environment (Buttigieg, West, & Dawson, 2011): perceived workload, and aggression from patients, the public, and colleagues. The prevalence of these two job demands is described in the section below, followed by a summary of other relevant psychosocial workings conditions.

2.2.1 Perceived workload

Unlike structural aspects of workload, such as number of hours worked, perceived workload quantifies how these are perceived by the doctor. The type of work, the experience and ability of the doctor, and the time frame in which it needs to be completed in, are factors that influence perception (Buttigieg et al., 2011; Linzer et al., 2009). In the 2016 National Training Survey, completed by 98.7% of the national junior doctor population (n=53,835), 44% of doctors described having a heavy or very heavy workload. In comparison, 52% described their workload as "about right" with the remaining reporting workload as light. Membership surveys from the Royal College of Physicians reveal feelings that staff shortages, funding cuts, lack of training time, and patients with increasing complex medical issues, were all contributory factors in 91% doctors reporting an increase in workload over the last five years (Royal College of Physicians, 2013, 2015a). At the consultant level, the 2014-15 national census revealed 50% still worked under excessive demands despite the number of consultants nationally increasing by 3.2% (Federation of the Royal Colleges of Physicians of the UK, 2016).

2.2.2 Workplace aggression

Aggression in healthcare is rife, and can originate from patients, their families, and the wider public, as well as from colleagues and superiors at work. An extensive literature base distinguishes different forms of aggression, including bullying, abusive supervision, incivility, and violence in the workplace (Hershcovis, 2011; Woodrow & Guest, 2012). This fragmented approach hinders our understanding of the topic, and instead workplace aggression has been proposed as an overarching construct to investigate negative interpersonal behaviours at work (Hershcovis, 2011). Aggressive behaviours include, but are not limited to, verbal abuse, threatening behaviour, physical violence, and obscene behaviours (Alexander & Fraser, 2004).

In the United Kingdom, 8% of junior doctors report being bullied at work and 13.6% witnessed such behaviours in the workplace (General Medical Council, 2016). The 2014 NHS Staff Survey highlighted that in the previous year, 10.1% of doctors experienced violence from patients and the public; 32% had been harassed by patients and the public; and 24.6% had been harassed by colleagues (Picker Institute Europe, 2015). Assessing accurate prevalence rates of workplace aggression is difficult, and there is considerable discussion of this elsewhere (Einarsen, Hoel, Zapf, & Cooper, 2010; General Medical Council, 2014; M. B. Nielsen, Matthiesen, & Einarsen, 2010). Briefly, the lack of clear definitions, valid measures, and a reluctance to speak out contribute to this range. For example, in one survey of UK doctors, although 84% had been exposed to bullying behaviours, only 37% reported being bullied in the previous year (Quine, 2002).

2.2.3 Other psychosocial working conditions

When national surveys of doctors have included other psychosocial working conditions, these focus on how doctors felt they were being supported. For example, the 2016 National Training Survey demonstrated that 88.7% of doctors were happy with the support they receive at work, and 84.2% experienced high quality supervision (General Medical Council, 2016). Similarly, doctors from acute hospitals responding to the 2014 NHS Staff Survey felt they were happy with support from their managers (71.7%) and felt that they worked in effective teams (83%; Picker Institute Europe, 2015). The numerous psychosocial working conditions that doctors experience are highlighted in a German study comparing psychosocial levels between doctors and a general sample of workers (Fuß, Nübling, Hasselhorn, Schwappach, & Rieger, 2008). This study found doctors experienced significantly more quantitative work demands, work-family-conflict, and role conflict than the comparison sample. Doctors also had less control at work and experienced poorer quality of leadership. No differences were found on other measures, including role clarity, emotional, and cognitive demands, developmental opportunities, sense of community, and relational justice.

2.2.4 Summary

It is evident that the healthcare environment is challenging and demanding to doctors who work in it. However, focusing on structural aspects of work (i.e., working hours, number of cases worked) ignores the role of the individual in perceiving their work environment (Jagsi & Surender, 2004; Williams et al., 2002). Considering the evidence reviewed so far, this psychosocial perspective is arguably more important, especially when trying to understand and predict the consequences of challenging working conditions. The link between doctors' psychosocial working conditions with their work-related wellbeing and with quality of care, form the basis of this thesis. This begins in Section 2.3 below by first describing the state of doctors' wellbeing and the psychosocial conditions that precede it.

2.3 Psychosocial Working Conditions and Doctors' Work-related Wellbeing

An estimated 10% of London's 30,000 doctors have sought treatment from the specialist occupational mental health unit for doctors (Roberts, 2016), illustrating a substantial problem with the psychological wellbeing of doctors. There is growing evidence suggesting that the wellbeing of doctors is not only influenced by individual characteristics, but is strongly linked to their occupational environment (Lee et al., 2013; Sibbald, Bojke, & Gravelle, 2003). The aims of this section are therefore two-fold: (i) to provide insight into the current state of doctors' work-related wellbeing; and (ii), to examine the psychosocial working conditions that precede them. This aligns with the broader thesis aims, as it focuses on the first two constructs in the relationships between psychosocial working conditions, work-related wellbeing, and quality of care.

Although a portmanteau construct, the World Health Organization's (1948) definition of health probably is the mostly widely accepted and used: "health is the state of complete physical, mental and social well-being, not merely the absence of disease or infirmity" (pg. 1). However, studies of doctors' physical health seldom find substantial differences between doctors and the general population (Fuß et al., 2008), suggesting that there are fewer occupational factors contributing to doctors' physical ill-health. In fact, doctors have been found to have lower mortality rates than national-norms, particularly on cardiovascular, diabetes, and lifestyle diseases (Carpenter, Swerdlow, & Fear, 1997; Frank, Biola, & Burnett, 2000). Instead, this thesis focuses on psychosocial working conditions, which are more strongly associated with psychological than physical morbidity (Leka & Jain, 2010). Moreover, doctors are particularly vulnerable to the three "Ds" – depression, drink, and drugs (Markwell & Wainer, 2009); which are indicators of affective and behavioural dimensions of wellbeing. Therefore, instead of using the term "health" which encompasses physical wellbeing, this thesis uses the term "work-related wellbeing".

As this thesis is set in an occupational context, there is a need to understand what wellbeing actually encompasses here. Work-related wellbeing traditionally has been construed as an affective component that is characterised through mood and emotion (Diener, Suh, Lucas, & Smith, 1999; Skakon, Nielsen, Borg, & Guzman, 2010; Soh, Zarola, Palaiou, & Furnham, 2016), and is often represented by measures such as job satisfaction and emotional exhaustion. It has, however, progressively expanded to include behavioural, motivational, cognitive, and psychosomatic dimensions as well (van Horn et al., 2004; Warr, 1994). Furthermore, there is strong agreement that wellbeing exists on a continuum ranging from ill-health to feelings of happiness (Bakker & Schaufeli, 2008; Diener et al., 1999; Hall et al., 2016), this includes both negative (e.g., burnout) and positive (e.g., work engagement) manifestations. The subsequent sections therefore review the available evidence pertaining to the prevalence of work-related wellbeing among doctors, namely negative (i.e., burnout, work-related stress) and positive (i.e., job satisfaction) affect, motivational (i.e., work engagement), and behavioural (i.e., presenteeism) dimensions.

2.3.1 Burnout

By far the most common measure of doctors' wellbeing is burnout, a syndrome characterised by emotional exhaustion, depersonalisation, and reduced levels of personal accomplishment (Maslach, Schaufeli, & Leiter, 2001). Inconsistencies in interpreting burnout measures means some studies class burnout as high scores on the aggregate of these three dimensions, or high scores on one, two, or all three dimensions (Eckleberry-Hunt, Kirkpatrick, & Barbera, 2017). Obviously studies utilising a more conservative approach yield lower prevalence rates than those that class burnout as high on one dimension alone (Marcelino et al., 2012; Ripp et al., 2011; Selič, Stegne-Ignjatović, & Klemenc-Ketiš, 2012). Hence, reviews on the prevalence of burnout amongst doctors has led to ranges of between 17.6% and 76% (Prins, Gazendam-Donofrio, et al., 2007; Thomas, 2004; Trufelli et al., 2008). This range is also explained by the evidence that doctors who were younger (Ožvačić Adžić et al., 2013), medical residents (Sargent et al., 2004), had frontline patient contact (Shanafelt, Boone, et al., 2012), and male (Trufelli et al., 2008), were susceptible to higher burnout rates. An alternate interpretation focuses on prevalence along each burnout dimension. One review of 25 studies found between 41% and 50% of medical residents were emotionally exhausted; between 34% and 70% experienced depersonalisation; and 23% scored low on personal accomplishment (Prins, Gazendam-Donofrio, et al., 2007).

Placing these figures into context, doctors consistently report higher burnout rates than comparison samples in Australia (Benson, Truskett, & Findlay, 2007), Germany (Fuß et al., 2008), the Netherlands (Prins et al., 2010), and the United States (Shanafelt, Boone, et al., 2012). In addition, the proportion of doctors in the United Kingdom (Taylor et al., 2005) and the United States (Shanafelt et al., 2015) experiencing burnout symptoms has increased, while no change was reported in a comparison sample of workers. Even within the healthcare sector and its high rates of burnout (Adriaenssens, De Gucht, & Maes, 2015), doctors have been found to report higher burnout rates than their colleagues from other disciplines (Grunfeld et al., 2000; Sargent et al., 2004; Visser et al., 2003).

Psychosocial working conditions and burnout. In a meta-analysis of burnout correlates involving 65 samples of doctors (Lee et al., 2013), emotional exhaustion positively correlated with six out of eight job demands: workload; organisation structure (e.g., inflexible work arrangements); professional values (e.g., compromising beliefs); position specific demands; inadequate resources; and insufficient input. Only one of four job resources exhibited the anticipated negative relationship with emotional exhaustion. Workload was the most commonly examined predictor of emotional exhaustion. Only four predictors of depersonalisation were found: organisation structure; position specific demands; professional values; and inadequate skills. Once again, workload was the most frequently examined predictor, although along with recognition, autonomy, support, and role conflict it was not significant. The stronger, and more frequent, relationships between job demands and emotional exhaustion (than with depersonalisation) is congruent with the assertion that emotional exhaustion is the core component of burnout (Maslach et al., 2001).

From their systematic review of 19 burnout studies in junior doctors, Prins et al. (2007) found individual risk factors were poor predictors of burnout and that instead, psychosocial work characteristics displayed stronger relationships with burnout. Studies not included in either review continue to support the impact workload has on emotional exhaustion and burnout, as seen in samples of Malaysian doctors (Ahmad, 2010), Dutch medical residents (Schaufeli, Bakker, van der Heijden, & Prins, 2009), Israeli specialists (Shirom, Nirel, & Vinokur, 2006), and Swiss general practitioners (Goehring, Gallacchi, Künzi, & Bovier, 2005). There is also growing evidence suggesting a relationship with social support from colleagues (Prins et al., 2008; Schaufeli, Bakker, van der Heijden, et al., 2009) and superiors (Prins et al., 2008; Prins, Hoekstra-Weebers, et al., 2007). It is also important to highlight the job demands not previously examined, namely frequency of difficult patient encounters (An et al., 2009, 2013) and work-family conflict (Ahmad, 2010; Fuß et al., 2008), which positively correlated with burnout amongst doctors.

In terms of longitudinal evidence, Bakker, Schaufeli, Sixma, Bosveld, and Van Dierendonck (2000) followed 207 general practitioners over a five-year period. They found that doctors who cope with emotional exhaustion by withdrawing from their patients inadvertently create more demanding patients, leading to subsequent increased feelings of emotional exhaustion. In a different survey of 1,668 UK doctors, workload and lack of support were related with burnout scores from six years earlier (McManus, Keeling, & Paice, 2004). This not only raises questions surrounding causality, but implies that the burnout levels of doctors influence how they subsequently perceive their work environment. It further suggests that doctors with burnout are likely to be trapped in a continuing downward spiral.

Theoretical models are important in understanding causal models, although Prins et al.'s (2007) review found not one study drew on existing theories surrounding burnout or psychosocial working conditions. One study since that has done so focused on the effort-reward imbalance model (Siegrist, 1996). In this study of German paediatricians (Weigl et al., 2015), both effort (e.g., time pressure, interruptions) and rewards (e.g., esteem, salary, job security) were significant predictors of emotional exhaustion. This effect was even more pronounced when high effort was paired with low rewards.

2.3.2 Work-related stress

It is beyond the scope of this thesis to review the debate surrounding the definition and measurement of work-related stress (see Section 5.1 for an overview). When doctors were asked if they could not cope with, or were feeling stressed in their work, 21.6% of 422 doctors (Linzer et al., 2009) and 22% of 2,326 doctors (Linzer et al., 2002) reported feeling very stressed. This is considerably lower than the 55% of highly stressed doctors observed in a survey of 1,573 Dutch medical specialists (Visser et al., 2003). In a different model of stress, the effort-reward imbalance model postulates that an imbalance between effort spent at work and rewards received results in emotional distress (Siegrist, 1996). Studies using this model report that effort was seen to exceed rewards for 25.1% of German surgeons; 28.4% of German paediatricians (Weigl et al., 2015); 28% of Italian radiologists (Magnavita et al., 2008); and between 22.7% and 37.1% of Swiss doctors (Buddeberg-Fischer, Klaghofer, Stamm, Siegrist, & Buddeberg, 2008).

Similar prevalence rates (between 22%-24%) were observed when work-related stress was represented as job strain, that is, being exposed to high demands and low control in the workplace (J. Klein et al., 2011; Magnavita et al., 2008; von dem Knesebeck, Klein, Frie, Blum, & Siegrist, 2010). What is evident across the different stress models, are the prevalence of high stress consistent between 20% and 30%, which are generally higher than that of the general population (Fuß et al., 2008; J. Klein et al., 2011).

Psychosocial working conditions and work-related stress. In regression based studies, personal, structural work conditions, and non-work factors have a minimal impact on work-related stress compared to psychosocial working conditions (Linzer et al., 2002; Shackelton et al., 2010). Testing the job demands-control-support model (Johnson & Hall, 1988) in a sample of 2,362 doctors in the United States, Linzer and colleagues (2002) found four measures of job demands, two measures of job control, and all three support measures, to significantly predict work-related stress. Two of the demands (work hours and solo practice) represented structural aspects of work, although none of the remaining 12 personal or structural predictors were significant.

As expected, work-home interference, restricted professional autonomy, and high workload predicted high work-related stress in Dutch consultants (Visser et al., 2003); while work-related stress was not predicted by any of the job resources examined, namely social support, effective management, feeling valued, intellectual stimulation, and job security. In a sample of 422 general practitioners in the United States (Linzer et al., 2009), work-related stress had a negative correlation with job control, and positive correlations with time pressure and work pace. Administrative and clinical autonomy were also found to be negatively correlated with work-related stress in a sample of 640 doctors drawn from Germany, the United Kingdom, and the United States (Shackelton et al., 2010).

2.3.3 Job satisfaction and work engagement

The preceding sections focused on negative wellbeing among doctors, reflecting the tendency in psychology to focus on negative instead of positive states (Bakker & Schaufeli, 2008). However, wellbeing exists on a continuum, with depression, anxiety, and ill-health on one end; and happiness, flourishing, and thriving on the other (Hall et al., 2016; Shanafelt et al., 2003; Wallace et al., 2009). One of the oldest, and most commonly, studied constructs in organisational research is job satisfaction (Locke, 1970). Studies that have reported on prevalence rates have found job satisfaction to be high, reflected by over 80% of Australian (Joyce et al., 2011), British (Federation of the Royal Colleges of Physicians of the UK, 2016), and Dutch (Visser et al., 2003) doctors reporting being moderately or highly satisfied with their work. A more recent construct with growing popularity is work engagement, defined as a positive, fulfilling, work-related state of mind characterised by vigour, dedication, and absorption (Bakker, Schaufeli, Leiter, & Taris, 2008). The construct has limited receivership with studies involving doctors, although one study did report that 27% out of 2,115 Dutch residents exceeded the threshold for being highly engaged at work (Prins et al., 2010).

Psychosocial working conditions, job satisfaction, and work engagement. Reflecting the lack of research surrounding positive occupational psychology (Bakker & Schaufeli, 2008), there has not yet been an examination of the psychosocial antecedents of work engagement in doctors. However, job satisfaction's long history means this is not the case for this construct. The most comprehensive examination of antecedents to doctors' job satisfaction is a systematic review of 97 articles from the United States (Scheurer et al., 2009). Nine out of the ten studies involved perceived pressure, and all four of the studies that used perceived workload, observed a significant correlation with job satisfaction. In comparison, only one of the four studies that used structural measures of workload was found significant. In terms of job resources, job control and autonomy (15/16 studies) and colleague support (5/5) were consistent predictors of job satisfaction. The importance of psychosocial predictors is underlined by income being the only consistent predictor out of all the other doctor (e.g., specialty, age, gender), practice (e.g., size, location), and patient characteristics examined. The strength of psychosocial predictors over personal and job characteristics is also evident in other studies (Joyce et al., 2011; Visser et al., 2003). For example, 34% of the variance in Dutch medical specialists' job satisfaction was predicted by their psychosocial work environment, and not number of hours worked, specialty, gender, age, or whether or not they had children (Visser et al., 2003).

The job demands-resources model (Bakker & Demerouti, 2007) postulates that job resources are closely linked with positive wellbeing. The studies of Laubach and Fischbeck (2007) and Joyce et al. (2011) support this proposition, as job resources (e.g., feeling valued, intellectual stimulation, job security) were the only predictors of job satisfaction, while this was not the case for any job demands examined. Further support is evident in Mache and colleagues' (2012) study, where all eight resources (influence at work, autonomy, developmental opportunities, leadership, social support, feedback, social relations, sense of community) correlated with job satisfaction. Not only did one of the three demands fail to exhibit this relationship, the correlation coefficients for the remaining two (quantitative, emotional demands) were lower than any of the job resources. Similar outcomes are seen in a survey of American doctors, where job control (i.e., resource) had a larger effect size than time pressure and work pace (Linzer et al., 2009). However, in Visser and colleagues' (2003) study, the best predictor of job satisfaction was the only significant job demand (feeling poorly managed and resourced), and not one of the four significant job resources.

2.3.4 Presenteeism

Unlike the preceding sections which represented the affective and motivational dimensions of wellbeing, presenteeism is a behavioural manifestation of wellbeing. More specifically, this refers to workers being present at work when they should instead be on sick leave (Jena, Baldwin, Daugherty, Meltzer, & Arora, 2010). Data from the 2013 NHS Staff Survey showed that 54.2% of doctors had engaged in presenteeism in the previous three months (NHS Staff Survey Co-ordination Centre, 2014b). This separated into 54.8% of consultants and 48.3% of junior doctors. In the United States, 57.9% of postgraduate trainees reported presenteeism at least once in the previous year, while 31.3% had done so more than once (Jena et al., 2010). In New Zealand (Chambers, Frampton, & Barclay, 2017), of 1,806 senior doctors and dentists, 88% had come into work in the previous two years even though they were not well enough to do their jobs, with 75% being present while they had an infectious illness. On average, each respondent had three presenteeism days in this time period.

Psychosocial working conditions and presenteeism. To date there have been limited attempts to explore the psychosocial antecedents of doctors' presenteeism. Qualitative explorations have identified high workloads, not wanting to let down colleagues, inability to adapt to a patient role, and competitive pressure as contributing factors (Chambers et al., 2017; Jena et al., 2010; Oxtoby, 2015). In addition, some doctors believe that going on sick leave is a sign of weakness or reflection of incompetence (Jena et al., 2010). These links with demands and support in the workplace is evident in a survey of Dutch residents, where presenteeism was found to be positively correlated with both job demands and resources (Schaufeli, Bakker, van der Heijden, et al., 2009). These were respectively represented by seven (e.g., work overload, work-home conflict, role conflict) and six (e.g., coaching, supportive colleagues, job control) measures.

2.4 Conclusion

2.4.1 The state of doctors' work-related wellbeing

This chapter reviewed the evidence surrounding the level of burnout, work-related stress, job satisfaction, work engagement, and presenteeism among doctors. It was beyond the scope of this thesis to discuss the research on other measures of doctors' wellbeing (e.g., anxiety and depressive symptoms) or health-related behaviours. This is despite some of these health-related behaviours being salient to doctors, including motor vehicle accidents (Gander, Purnell, Garden, & Woodward, 2007; C. P. West, Tan, & Shanafelt, 2012), sickness absenteeism (Kivimäki et al., 2001), substance abuse (Sebo, Gallacchi, Goehring, Künzi, & Bovier, 2007; Wischmeyer et al., 2007), and poor sleeping patterns (Dollarhide et al., 2013; General Medical Council, 2016). Doctors have also been found to have higher suicidal ideation and suicide rates (Shanafelt, Balch, & Dyrbye, 2012; Stack, 2004) and depressive symptoms (Fahrenkopf et al., 2008; Shen et al., 2012) than the general population.

One final issue worth noting is that surveys of doctors typically suffer from low response rates. For example, one review of burnout correlates found the average response rate across 62 samples of doctors to be only 19% (Lee et al., 2013). It may be that doctors who most struggle with their work demands, or are suffering from burnout or depression, may not find the energy and motivation to complete these surveys (Gander et al., 2007; Taris & Schreurs, 2007). Furthermore, doctors are often reluctant to identify themselves as being ill or struggling (Grant, Rix, Mattick, Jones, & Winter, 2013; B. Hayes, Prihodova, Walsh, Doyle, & Doherty, 2017), which may lead to further underreporting. Therefore, despite the concerning prevalence rates highlighted, almost all were drawn from self-reported surveys and are vulnerable towards underestimating the true state of wellbeing in doctors.

2.4.2 Psychosocial working conditions and doctors' work-related wellbeing

By far the most common psychosocial predictors examined were workload, job control, and social support. This may be informed by these constructs representing the core components of one of the more popular work-related stress theories: the job-demand-control-support model (Johnson & Hall, 1988). Despite this, few studies integrated theoretical models or hypotheses development into their study designs. Although this was raised as an issue in relation to burnout (Prins, Gazendam-Donofrio, et al., 2007), this is also the case when psychosocial working conditions predicted job satisfaction (Mache et al., 2012) and work-related stress (Linzer et al., 2002). This may be a side-effect of most studies involving doctors being published as short articles in medical journals, where "medical journals and research funders are mainly concerned with practical factual research, not with research that develops theories" (pg. 1007; Alderson, 1998). This highlights the need to not only demonstrate relationships between psychosocial work conditions and work-related wellbeing (and eventually quality of care), but to be able to explain why this occurs.

As expected, examination of psychosocial antecedents reveals that job demands better predicted negative wellbeing (e.g., work-related stress, burnout) than job resources. In contrast, job resources exhibited stronger relationships with job satisfaction than job demands. This is consistent with the job demands-resources model's dual-process hypothesis (Bakker & Demerouti, 2007; Demerouti et al., 2001), which postulates that job demands and resources respectively have separate and independent effects on workers' negative and positive wellbeing.

2.4.3 Looking forward

This chapter provided a snapshot into the current state of doctors' work-related wellbeing. What is evident is not only that doctors experience higher levels of burnout and work-related stress than the general population (Grunfeld et al., 2000; Shanafelt, Boone, et al., 2012), but these levels are continuing to increase (Shanafelt et al., 2015). While studies often examine individual characteristics (e.g., demographics) and structural aspects (e.g., hours worked, caseload) as antecedents to these wellbeing measures, the evidence is clear that perception of these as demands or resources have stronger implications on doctors' work-related wellbeing (Linzer et al., 2002; Visser et al., 2003).

As the main purpose of the healthcare sector is to provide effective and safe care to the public, it is important to appreciate how psychosocial working conditions and doctors' work-related wellbeing, ultimately predict quality of care. The next chapter serves to introduce what quality of care means, before exploring these relationships in more detail.

Chapter 3 : Quality of Care in Healthcare and Its Antecedents

The primary purpose of the healthcare sector is for the provision of safe and effective care to society (Department of Health, 2008; Keogh, 2013). As described at the start of this thesis, the quality of care provided by the NHS has drawn national attention, primarily due to junior doctor strikes (Bagenal et al., 2015) and the Mid-Staffordshire and Morecambe Bay scandals before that (Francis, 2013; Kirkup, 2015). Although numerous publications have advocated the improvement of working conditions of doctors to increase their work-related wellbeing, and in turn, raise quality of care; there is little evidence advocating an empirical relationship between doctors' working conditions and quality of care provided.

This section therefore aims to introduce what quality of care means, and what some of the main outcome measures are. It then extends the relationship built in the preceding chapter to link psychosocial working conditions with quality of care. While the previous chapter focused on psychosocial working conditions with doctors' work-related wellbeing, this chapter links wellbeing with quality of care.

3.1 Quality of Care

The variable nature of patient care makes the definition and assessment of "quality" within the healthcare sector a complex issue (King's Fund, 2011; J. Klein et al., 2011). This is compounded in that the value placed on different outcomes is dependent on the stakeholder, with governments, hospitals, doctors, and patients having different perspectives on what constitutes quality or a good outcome (King's Fund, 2011; Wong & Cummings, 2007).

The *NHS Next Stage Review* (Department of Health, 2008) defined quality in the NHS as consisting of three core areas: clinical excellence, patient safety, and the experience of patients. Clinical excellence is preventing premature deaths, enhancing quality of life, and assisting recovery. The second core aspect: patient safety, aims to provide a safe care environment without avoidable harm. The final aspect of quality of care is patient experience, which refers to the quality of the patient's experience towards their personal care and treatment. Related to this is the NHS Outcome Framework emphasising five domains of quality and outcomes (Department of Health, 2010). The first three domains map onto the core area of clinical excellence: (i) preventing people from dying prematurely; (ii) enhancing quality of life for people with long-term conditions; and (iii) helping people recover from episodes of ill-health or injury. The core area of patient experience pertains to the domain (iv) ensuring people have a positive experience of care. The fifth domain of (v) treating and caring for people in a safe environment and protecting from avoidable harm, matches the core area of patient safety. With the Institute of Medicine having a similar definition in the United States (Aspden, Corrigan, Wolcott, & Erickson, 2004), it is evident that the measure of quality through patient outcomes and patient safety are conceptually and operationally interwoven with healthcare quality.

Assessing quality and safety is not straightforward as different measures illustrate different aspects, and possibly non-comparable indicators of quality (Hall et al., 2016; Vincent, 2010). It may be useful to distinguish different quality measures according to the three core areas of the *NHS Next Stage Review* (Department of Health, 2008). However, this is not always practical or straightforward. For example, a prescription error falls short of enhancing patient safety, but if it results in a lengthened stay in hospital or even death then it does not meet clinical excellence. Nevertheless, the subsequent section reviews the state of the evidence surrounding quality of care; first at the level of the hospital or nationally, and second at the level of the individual doctor.

3.1.1 Quality of care at the hospital and national-level

The first core area of quality in the NHS, namely clinical excellence, aims to prevent premature deaths, enhance quality of life, and assist recovery (Department of Health, 2008, 2010). Focusing on preventable deaths, a review of 1,000 adults deaths in ten English hospital trusts reported that 5% of deaths could have been prevented (Hogan et al., 2013). Preventable was attributed when the death was a result of an incorrect diagnosis or treatment plan, or by treatments deemed unsafe. Extrapolated to England, this equalled 11,859 preventable adult deaths nationally in 2009. A larger follow up study reviewed 100 deaths from each of 34 trusts and found the rate of preventable deaths had declined to 3.6% (Hogan et al., 2015), although this was not statistically different to the results from the 2009 study.

Interlinked with the first core area is patient safety, which aims to provide a safe care environment without avoidable harm (Department of Health, 2008, 2010). When 1,006 hospital admissions in England were examined, adverse events were identified in 8.4% of cases (Sari et al., 2007). Of these, 31% were deemed avoidable, with 15% leading to an impairment exceeding six months and a further 10% contributing to patient death. These rates were slightly lower than the 10.8% noted in an earlier examination of acute hospital admission (Vincent, 2001). Beyond the United Kingdom, a systematic review of eight studies from five countries (de Vries, Ramrattan, Smorenburg, Gouma, & Boermeester, 2008) found the median adverse events rate for acute hospital admissions was 9.2%. Nearly half of these were preventable (43.5%), with 7.4% contributing to patient death and 7% to permanent disability. Adverse outcomes have also been linked with longer hospital stays (Baker et al., 2004), post-discharge complications (Forster, Murff, Peterson, Gandhi, & Bates, 2003), and rehospitalisation (Moore, Wisnivesky, Williams, & McGinn, 2003).

The measurement of patient experience, the third core area of quality, is not as developed as the outcomes discussed above. In fact, it has attracted criticism due to the inconsistency in specifying what this represents, the difficulty in establishing the validity of measures, and its poor links with other more objective quality measures (Coyle & Williams, 1999; Crow et al., 2002; Mehta, 2015; Salisbury, Wallace, & Montgomery, 2010). The NHS' Friends and Family Test asks patients whether they would recommend the service used based on their experience of it (NHS England, 2016), and has over the last three years been consistent in its satisfaction rates: 77% in 2014/15, 79% in 2015/16, and 80% in 2016/17. This is congruent with the 2015 NHS Inpatient Survey, where 85% of overnight patients were satisfied with their overall hospital experience (Care Quality Commission, 2016). However, lower satisfaction was found in the 2015 British Attitude Survey; only 66% of respondents were satisfied with the service received from the NHS in the preceding year (Appleby & Robertson, 2016).

The difficulties in accurately capturing healthcare quality, particularly when reliant on self-report data, is highlighted in a review of 5,879,954 patient safety incident reports collected in England over a ten year period (Howell et al., 2015). No relationships were found between acute hospitals' number of incidents reported, hospital mortality, and patient satisfaction. These outcomes could respectively represent the NHS' core areas of patient safety, clinical excellence, and patient experience. Interestingly, hospitals with high number of incident reports also had lower levels of litigation claims, a possible reflection that high rates of voluntary reporting could be a manifestation of a culture that is safety conscious.

3.1.2 Quality of care at the individual-level

The breadth of quality measures at the level of the doctor focus mainly on prescription errors, self-reported medical errors, and patient-rated experience. In a systematic review of 65 international studies of doctors' prescription errors, Lewis et al. (2009) concluded that 7% of all prescriptions contained an error. This translated to 52 errors per 100 hospital admissions, and involved 24 errors per 1000 patients. Since then, a census on 124,260 medication orders across 19 English hospitals found 8.9% of them contained some form of an error (Dornan et al., 2010). Not surprisingly, doctors in their first two years of postgraduate training, who have less experience than consultants and have more contact with patients, had higher error rates than their senior colleagues (8.4% in year 1; 10.3% in year 2; 5.9% for consultants). These findings were congruent with a review of patient charts in eight Scottish hospitals, where second year postgraduate doctors had the highest error rate (8.6%), followed by first year postgraduates (7.4%) and then consultants (6.3%; Ryan et al., 2014).

In one postal survey of British junior doctors, all of the 114 respondents were able to recall making a medical mistake in the previous year (Wu, Folkard, McPhee, & Lo, 2003). However, only 45% of invited doctors here returned their survey. For 90% of these medical residents, the medical mistake had a significant negative impact on the patient, including emotional distress (27%), prolonged hospital stay (24%), and death (31%). High prevalence of doctors acknowledging having made mistakes was also seen among 222 anaesthesiologists (84%) across eleven Scottish hospitals (Flynn, Fletcher, McGeorge, Sutherland, & Patey, 2003).

Finally, in terms of patient experience, Feddock and colleagues (2005) studied 168 patient-medical resident dyads, where 73% of residents had at least one patient not satisfied with the interaction with the residents. However, this is despite mean patient satisfaction scores indicating the patients as a sample were satisfied with their visit. In Germany, 65.8% out of 1,844 breast cancer patients rated the support they receive from their doctor as high quality, in contrast to 20.9% who felt the support was poor quality (Ansmann et al., 2013). This approximates the 70% of German patients satisfied with the quality of care provided by their surgeons (Mache et al., 2012).

Three points are worth noting in relation to the research around the quality of care, and in particular errors, provided by doctors. First, while errors and poor care should be avoided, it should not be ignored or penalised (Vincent, 2010; Wu et al., 2003). Doing so facilitates a culture of blame and not one that facilitates openness and learning. Second, errors can create "double victims" – the patient and the doctor (Shanafelt et al., 2010; Wu et al., 2003). Doctors who are not able to cope in a healthy manner with mistakes made are vulnerable to impaired psychological wellbeing, which increases their susceptibility towards more mistakes and traps them in a continuous downward spiral (Wu et al., 2003). Third, fear of reprisal, a refusal to acknowledge mistakes made, and less contact with patients, are all factors that could affect the prevalence of quality-of-care scores reported by doctors (Hall et al., 2016; Prins et al., 2010).

3.1.3 Summary

Substantially less data exists on quality of care at the level of the individual doctor than at the hospital or national-level. This is not surprising given the infrastructure for ongoing data collection, and interest in healthcare as a sector (Appleby & Robertson, 2016; King's Fund, 2011; NHS England, 2016). As this section outlines, attempts to conceptualise and measure quality of care are not straightforward. Where this has been done, the statistics indicate a pattern where errors and poor quality do exist, and crucially could be avoided. The challenge therein lies in fostering a healthcare environment where doctors can flourish and provide better standards of care than currently is being done.

3.2 The Link between Doctors' Wellbeing and Quality of Care

The relationship between work-related wellbeing and quality of care is more commonly studied in healthcare professionals aside from doctors. This is seen in a recent systematic review (Hall et al., 2016) and meta-analysis (Salyers et al., 2016), where respectively only 17% and 25% of included studies focus specifically on doctors. For burnout, the meta-analysis of 102 studies reported significant negative relationships between burnout and quality (r=−.26) and safety (r=−.23) outcomes (Salyers et al., 2016). Effect sizes were strongest amongst nurses, followed by multidisciplinary samples, and then doctors. Relationships at the level of the individual were also stronger than when burnout was aggregated to the unit level and related with organisational care outcomes. Similarly, the systematic review found that 16 studies of poor wellbeing (represented by depression, job stress, mental health, and distress; n=27) and 21 studies on burnout (out of 30) observed a negative relationship with patient safety (Hall et al., 2016). Interestingly, the majority of studies that did not observe significant relationships utilised objective measures of patient safety.

For doctors, the quote by Galen (130-200 A.D.) that the "physician will hardly be thought very careful of the health of his patients if he neglects his own" (pg. 305, Huth & Murray, 2006), emphasises the long-held view that the wellbeing of the doctor influences the care that they are able to provide. The section below reviews the evidence for this relationship: first with individual-level outcomes, before exploring organisational-level outcomes. As only one known study has examined a relationship between presenteeism and quality of care among doctors, presenteeism is not covered within the individual-level outcome discussions below. This study, set among Dutch residents, observed a positive correlation between presenteeism and self-rated medical performance (Schaufeli, Bakker, van der Heijden, et al., 2009).

3.2.1 Work-related stress and quality of care

The core aspects of quality of care most commonly examined here centres around the clinical excellence and patient safety aspects, but not patient experience. In terms of clinical excellence, a qualitative account of UK doctors' revealed that half of the sample reported reduced standards of patient care due to feelings of stress, including 7% which led to serious mistakes (Firth-Cozens & Greenhalgh, 1997). Inconsistent findings were found amongst German hospital doctors (J. Klein et al., 2011), where job strain predicted all five measures of self-rated service and performance measures. However, a negative imbalance between doctors' efforts and rewards only predicted three outcome measures.

Findings involving observed performance were more consistent. Arora et al. (2010) found that surgeons' subjective and objective measures of stress were correlated with time taken, dexterity, and number of errors, during simulations. Subjective stress was measured via an anxiety scale, while heart rate was a proxy for objective stress. In a different simulation of clinical reasoning, medical students who were placed in a high stress simulation had a negative relationship between stress scores and differential diagnosis performance (Pottier et al., 2013). No significant relationships were observed between stress measures and reaching a correct diagnosis, suggesting that stress did not influence students' ability to accurately treat the patient. Instead, students were not evaluating the entire situation's context, which may be detrimental in patient care. The study used validated questionnaires to measure subjective stress while salivary cortisol represented objective stress indicators. Interestingly, not one of the three studies that examined work-related stress reported a significant relationship with doctors' self-reported errors (J. Klein et al., 2011; Linzer et al., 2009; Winefield & Veale, 2002).

3.2.2 Burnout and quality of care

The popularity of burnout within research involving doctors results in more studies linking it with quality of care, compared with those relating to other measures of negative wellbeing. For the first core aspects of care (Department of Health, 2008) - clinical excellence, high burnout scores have been found to correlate negatively with self-reported quality of care amongst Israeli specialists (Shirom et al., 2006), German paediatricians (Weigl et al., 2015), and American general practitioners (Williams, Manwell, Konrad, & Linzer, 2007). In addition, 53% of American residents who experienced burnout reported providing suboptimal patient care on a monthly basis (Shanafelt, Bradley, Wipf, & Back, 2002), while Mexican junior doctors who reported high burnout levels were 5.5 times more likely to report providing poor care than their colleagues not experiencing burnout (Toral-Villanueva, Aguilar-Madrid, & Juárez-Pérez, 2009).

In terms of patient safety, most studies demonstrated a relationship between burnout and the number of self-reported errors made, although these almost exclusively involved doctors based in the United States (Shanafelt et al., 2010; C. P. West et al., 2006; C. P. West, Tan, Habermann, Sloan, & Shanafelt, 2010; Williams et al., 2007). An exception of this was a survey of 2,115 Dutch residents where all three dimensions of burnout were moderately positively-related with errors due to lack of time (Prins et al., 2009). However, only emotional exhaustion and depersonalisation were observed to relate, albeit weakly, with errors due to inexperience. Focusing specifically on the burnout dimensions (Prins et al., 2009; C. P. West et al., 2006, 2010) the largest effect sizes are reported between emotional exhaustion and self-reported errors, followed by depersonalisation and reduced personal accomplishment. This reinforces the role of emotional exhaustion and depersonalisation as the main dimensions of burnout, with stronger ramifications for both antecedents and outcomes (Maslach et al., 2001). As was the case with depressive symptoms, burnout was also found to exist in a negative downward spiral with self-reported errors (Shanafelt et al., 2010; C. P. West et al., 2006). Comparing depressive symptoms with burnout, studies involving surgeons (Shanafelt et al., 2010) and medical residents (C. P. West et al., 2010) both found depressive symptoms to have the larger effect size with error .

The expectation is that doctors with high level of burnout will withdraw from the patient relationship, resulting in patients feeling unsatisfied. This is evident in a study of 178 matched doctor-patient dyads (Halbesleben & Rathert, 2008). More specifically, emotional exhaustion predicted depersonalisation, which in turn predicted patient satisfaction and recovery time. Similarly, emotional exhaustion and depersonalisation, but not reduced personal accomplishment, was found to relate with patients' satisfaction with interaction and communication from their Greek doctors (Anagnostopoulos et al., 2012). Despite this, the evidence suggests an inconsistent relationship between burnout and patient experience. No relationship was observed between burnout in Croatian family practitioners and patient-rated care (Ožvačić Adžić et al., 2013); in British general practitioners' and patient-rated interpersonal skills or observed patient-centeredness (Orton, Orton, & Pereira Gray, 2012); or in US doctors' with patient experience (Ratanawongsa et al., 2008).

3.2.3 Explaining the negative wellbeing and quality-of-care relationship

According to the conservation of resources theory (Hobfoll, 1989; Hobfoll & Shirom, 2001), workers carefully manage their motivational resources. Consequently, doctors with burnout or depression could cease their motivational process and withdraw from aspects of work which burden their resources. As doctors choose carefully how to engage with their work environments, they may provide adequate instead of good levels of care by withholding all unnecessary attention towards the patient (Halbesleben & Rathert, 2008). At the same time, doctors who are depressed or experiencing burnout could compromise patient safety by their reduced problem solving and decision making capability (Pottier et al., 2013; Sargent et al., 2004).

That depressive symptoms and burnout, but not work-related stress, relate with quality of care suggests that these wellbeing constructs may operate differently. Doctors experiencing depression or burnout may already be struggling with the effects of poor wellbeing (Joules, Williams, & Thompson, 2014; Maslach et al., 2001), which might undermine their ability to provide care. In comparison, work-related stress revolves around the perceived inability to cope, and has actually been suggested to be antecedent towards burnout and depression (Hillhouse, Adler, & Walters, 2000; Weigl et al., 2015). Therefore, doctors experiencing work-related stress may still be able to absorb the negative influence of working conditions on their health while trying to provide the same level of care quality, although this could have ramifications on their own long-term health (Linzer et al., 2009; Rabatin et al., 2015).

It is further possible that the relationship between poor doctors' wellbeing and quality of care is dependent on the outcome measure utilised, evidenced by most non-relationships observed in relation to patient experience or observer-rated performance. Errors may be more salient to doctors experiencing burnout or depression, resulting in more of them being recalled within this group of doctors (Fahrenkopf et al., 2008; Welp, Meier, & Manser, 2015). It could also be that doctors' professional standards mean they still deliver, or appear to deliver, appropriate levels of service despite feelings of burnout (Larson & Yao, 2005). These doctors may also be aware of their limitations and may attempt to overcompensate in their delivery (Ratanawongsa et al., 2008). For example, doctors with burnout have been found to have longer consultations with patients (Zantinge, Verhaak, de Bakker, van der Meer, & Bensing, 2009), while stressed doctors use more patient-centred styles of communication during consultation (Howie, Hopton, Heaney, & Porter, 1992).

Finally, the relationship between wellbeing and quality of care could be reciprocal. Medical errors, an indicator of patient safety, actually create two victims- the patient and the doctor themselves (Wu et al., 2003). While the former experiences the physical (and mental) influence being harmed by their carer, for the latter the event may be a significant source of distress. This potentially traps doctors in a vicious cycle. Using research on depressive symptoms among doctors, West et al.'s (2006) three-year study of medical residents found that self-reported error was associated with increased odds for subsequent positive depression screening. This in turn increased the likelihood of making an error. Similar findings found depressive symptoms to predict surgeons reporting a major medical error in the preceding three months, and vice-versa (Shanafelt et al., 2010).

3.2.4 Positive wellbeing and quality of care

Scheepers and colleagues (2015) systematically reviewed 18 studies of doctors' wellbeing, arguing that insufficient attention was paid to positive manifestations of wellbeing. Of the included studies, job satisfaction was the most commonly examined predictor (*n*=14), followed by career satisfaction (*n*=3) and work engagement (*n*=1). This echoes the extant research, highlighting the dominance of job satisfaction within positive wellbeing research. Overall, wellbeing was associated with better patient experience (i.e., patient satisfaction, interpersonal relationships) and clinical excellence (patient adherence to treatment, quality of overall care process). However, the relationship with technical aspects of care yielded contrasting results. In a similar narrative review of 44 doctors' job satisfaction studies, eleven studies considered patient outcomes where in total 20 out of 23 correlations were significant (Williams & Skinner, 2003). More specifically, the outcomes were congruent with those examined by Scheepers et al. (2015).

The explanation for these relationships could lie with doctors who are happier and healthier being able to better process information and make better decisions, which could translate to more accurate diagnoses and treatment (Flinn & Armstrong, 2011; Scheepers, Boerebach, Arah, Heineman, & Lombarts, 2015). More engaged and satisfied doctors could also share their positive attitudes with their patients. This is particularly the case for work engagement which is characterised through feelings of energy, dedication, and passion for work

(Bakker, Schaufeli, et al., 2008). This may motivate patients to engage with the consultation process, feel positive, and adhere to the treatment plan (Grol, Mokkink, Smits, & Van, 1985; Scheepers et al., 2015). Finally, doctors with high levels of positive wellbeing may be willing to put in extra effort into their work to improve quality of care (Prins et al., 2009).

Surprisingly, when quality of care was measured through more technical aspects, results ceased to be consistent. It may be that quality of care represents a broader construct than just single dimensions of technical performance (Scheepers et al., 2015). The variety of possible measures could also mean that chosen dimensions may not be the most accurate, or sensitive, representation of technical performance. This is an issue that is also evident when doctors' negative wellbeing is examined, and warrants further research to understand the phenomenon and to develop better proxy measures.

3.3 The Link between Doctors' Wellbeing and Organisational-level Outcomes

In contrast to individual-level quality-of-care outcomes, there has been significantly less attention towards linking healthcare staff wellbeing and organisational-level quality-of-care outcomes (Pinder, Greaves, Aylin, Jarman, & Bottle, 2013; Welp et al., 2015). Organisational-level outcomes can include hospital mortality rates, infection rates, admission duration, and patient satisfaction, amongst others. Much of this data is routinely collected, highlighting its importance as indicators of quality of care (Powell, Dawson, Topakas, Durose, & Fewtrell, 2014; Topakas, Admasachew, & Dawson, 2010c). As such, identifying antecedents to these not only reinforces the value of these measurements, but provides areas that can be targeted to improve quality of care being provided. The dearth in this literature means the following sections are broadened to review healthcare staff in general and not only doctors. Moreover, the lack of studies means organisational-level patient safety indicators are not available. Instead, this section highlights relationships between healthcare staff work-related wellbeing with clinical excellence and patient experience aspects of quality of care.

3.3.1 Clinical excellence

In the United Kingdom, higher MRSA infection rates were reported in hospitals with poor staff wellbeing than those with better wellbeing (Boorman, 2009). The report used data on work-related stress, job satisfaction, turnover intention, and injury rates as collective indicators of staff wellbeing. Similarly, work-related stress, characterised by an imbalance between efforts and rewards, was linked with a 2.47 fold increase in infection rates (Virtanen et al., 2009). However job strain (i.e., high demands and low control) did not relate with infection rates in this study of 60 bed-wards across six Finnish hospitals. This was attributed to the effort-reward imbalance perspective encompassing a broader range of psychosocial working conditions that is more reflective of the workplace, for example distributive justice, in comparison the narrower job demand-control perspective. Similarly, an examination of Norovirus outbreaks on 11 German wards found a negative correlation between outbreak duration and the time after infected staff returned to work (Jansen et al., 2004), suggesting that presenteeism can facilitate the transfer of viruses in the healthcare environment.

Focusing on hospital mortality rates, a study involving 1,425 doctors and nurses from 54 Swiss intensive care units found emotional exhaustion was the only burnout component that predicted units' standardised mortality ratios (Welp et al., 2015). Although none of the three burnout components predicted patients' length of stay they all predicted clinicians' self-rated patient safety. Analysis based on the 2009 NHS Staff Survey using the work engagement related items found it predicted trusts' financial performance and absenteeism rates, but not mortality rate (Topakas, Admasachew, & Dawson, 2010a). Similar findings were found in a different study with presenteeism and general health as predictors (Topakas et al., 2010c). When a third study looked at longitudinal outcomes two-years later, neither work engagement, work-related stress, job satisfaction, or presenteeism were significant predictors of MRSA infection rates, *C. difficile* infection rates, and patient mortality in 2011 (Powell et al., 2014). However, the measurement of mortality rates has been argued to be a poor indicator of quality of care as it is too blunt a measure that is influenced by multiple factors (Bottle, Jarman, & Aylin, 2011), possibly making it not sensitive enough to relate with staff wellbeing (Powell et al., 2014).

3.3.2 Patient experience

The Boorman Review (2009) into staff wellbeing in the NHS grouped trusts into three bands based on their staff's work-related wellbeing, which consisted of work-related stress, job satisfaction, turnover intention, and injury rates. When matched with the national inpatient survey, trusts with better wellbeing had on average higher rates of patient satisfaction. In general, stronger effect sizes were observed within nurses, followed by doctors. Studies that drew data from the 2009 NHS Staff Survey aggregated individual-level wellbeing to the trust level, and found work engagement to correlate strongly with patient satisfaction scores in the same year (Topakas et al., 2010a), although the same was not observed for work-related stress, presenteeism, and general health as predictors (Topakas et al., 2010c). Aggregated job satisfaction from the 2007 Staff Survey had moderate positive correlations with all eight aspects of care drawn from the Inpatient Survey from the same year (Dawson, 2009). From a longitudinal perspective, work engagement and job satisfaction scores from the 2009 NHS Staff Survey predicted patient satisfaction scores in 2011 across 347 trusts (Powell et al., 2014). The same was not observed for work-related stress and presenteeism.

In terms of burnout, a large scale study involving 488 European and 617 American hospitals found that higher rates of nurse burnout was positively associated with patients reporting being less satisfied with hospitals (Aiken et al., 2012). More specifically, patients were less likely to rate their hospital experience as good; less likely to recommend the hospital; and to report less favourable nurse communication. Similarly, Garman, Corrigan and Morris (2002) found emotional exhaustion of healthcare staff from 31 rehabilitation teams to correlate with three out of four patient satisfaction aspects: environment, treatment, and preparation for autonomy. The relationship involving the remaining two burnout dimensions observed inconsistent findings with the patient satisfaction measures.

3.3.3 The multilevel perspective

Linking the wellbeing of doctors with hospital-level quality of care may appear straightforward, but actually presents a methodological issue that is often ignored altogether in the healthcare literature. Wellbeing exists as an individual-level measurement while hospital-level quality-of-care outcomes operate at the group level. To overcome this, researchers typically either: (i) aggregate all the individuals' responses within a team to create one team score; or, (ii) use the same team performance score for each member of that specific team (Heck & Thomas, 2015). This violates the compatibility principle which requires all variables within a model to operate at the same level of specificity (Ajzen, 2005). A multilevel perspective is important as it allows us to have an interaction between individual and the organisational/ systems perspective.

One study that considered a rationale for a team-level examination instead of the individual-level (Garman, Corrigan, & Morris, 2002), as the rehabilitation which patients undertook involved various group-focused treatments with different members of staff. The salient nature of the team therefore meant it did not make conceptual sense to match patients and staff at the individual-level. Garman and colleagues supported a multilevel perspective by examining the degree of in-group variance within units compared to the overall observed and theoretical variance in the sample. The comparison against observed sample variance was calculated with intra-class coefficients while comparison against theoretical sample variance was with an index of agreement. Despite advances in research designs, most studies aggregate wellbeing at individual-level to the unit or trust level with no consideration as to whether this is methodologically or conceptually justified.

3.4 Conclusion

In general, the evidence suggests a relationship between doctors' work-related wellbeing and quality of care, although there are a number of factors to consider. Firstly, this relationship is more evident with severe forms of negative wellbeing, namely depressive symptoms and burnout (Shanafelt et al., 2010; Weigl et al., 2015; C. P. West et al., 2010). This has led to the argument that based on these relationships, doctors' wellbeing should be included as a proxy indicator of high-quality healthcare (Wallace et al., 2009). In contrast, work-related stress, work engagement, and job satisfaction's role are more inconsistent, and may represent doctors still being able to function and perform adequate levels of service. Secondly, observer or patient-rated outcomes were also less consistently involved in significant relationships (Scheepers et al., 2015). It is plausible that these measures are not sensitive enough, although common method bias could also explain the consistent significant findings involving self-rated outcome measures (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Third, there is a distinct lack of studies examining this relationship from a multilevel perspective involving hospital-level data (Garman et al., 2002). This is despite these outcome data continually being collected. Finally, underpinning nearly all studies is a lack of theoretical exploration. This is important, as theory provides a framework that allows the explanation as to why doctors' wellbeing predicts, or not, quality of care. Clearly, the wellbeing of doctors and its relationship with quality of care is extremely complex and certainly warrants further exploration (Ožvačić Adžić et al., 2013).

Within the context of this thesis, this chapter completes the second part of the relationship that began with psychosocial working conditions relating to doctors work-related wellbeing (Chapter Two). However, the main aim of this thesis is to examine the relationship between psychosocial working conditions of doctors and quality of care. This forms the basis of the systematic review and meta-analysis in the next chapter.

Chapter 4 : A Systematic Review of the Relationship between Doctors' Psychosocial Working Conditions and Quality of Care Provided (Study 1)

The two preceding chapters broke down the relationships between doctors' psychosocial working conditions and quality of care into two parts. The former established a relationship between psychosocial working conditions and work-related wellbeing (Lee et al., 2013; Prins et al., 2007), while the latter described the work-related wellbeing and quality-of-care relationship (Scheepers et al., 2015). This chapter contains the first study of this thesis, which aims to conduct a systematic review and meta-analysis on the relationships between the psychosocial working conditions of doctors and the quality of care delivered. It begins by providing a rationale to conduct a systematic review on this relationship, before explaining what a systematic review actually is, then introducing the research questions. This is followed by the method section which encompasses the search and inclusion strategy. Lastly, the twenty one studies found are reviewed and discussed.

4.1 Introduction

Much of our understanding of the relationship between doctors' psychosocial working conditions and quality of care provided is extrapolated from the healthcare literature involving non-doctors. However, evidence for a direct relationship can also be drawn from the safety literature involving other sectors (Phipps, Malley, & Ashcroft, 2012). In the most comprehensive examination of this topic to date (Nahrgang, Morgeson, & Hofmann, 2011), the job demands-resources (JD-R) model (Bakker & Demerouti, 2007; Demerouti et al., 2001) was used to frame the relationships between psychosocial working conditions (i.e., job demands and resources), burnout, work engagement, and safety outcomes. This meta-analysis of 203 independent samples (*N*=186,440) found support towards the health-impairment process, whereby job demands (e.g., risks & hazards, complexity) positively related to burnout. In comparison, and according to the motivation process, job resources (e.g., knowledge, autonomy and supportive environments) positively predicted work engagement. As expected, both burnout and work engagement were related to safety outcomes. The meta-analysis also highlighted differences across industries, including that task complexity was the largest predictor of burnout in the healthcare and manufacturing sectors, while risk and hazards were the largest predictors in the construction and transportation sectors. Nahrgang and colleagues concluded that still little is known about psychosocial working conditions in relation to safety, and the validity of the JD-R model needs to be explored across different sectors.

Similar models drawing together the work environment, staff wellbeing, and quality-of-care outcomes have also been proposed within the healthcare sector (Lowe & Chan, 2010; Powell et al., 2014). For example, Lowe and Chan (2010) developed a conceptual model describing the relationship between the work environment and organisational outcomes. The model proposes that contextual factors (e.g., economy, governmental policy) influence job and work environment factors. In turn, this influences workers' work life and their levels of wellbeing. These then impact workforce stability, costs and productivity, care quality, and patient safety. Focusing specifically on doctors' experience of burnout, Montgomery and colleagues (2011) proposed that this would mediate the relationship between working conditions and organisational factors on one hand, with patients' experience of their hospital stay on the other.

Few attempts have been made to empirically test these models within the healthcare sector, especially in relation to doctors; although, qualitative explorations on this topic lend credence towards these models. Interviews with twenty Irish doctors found that doctors felt that difficult working conditions compromised the care they were able to give towards patients, mainly by rushing the amount of time they had with the patient (McGowan et al., 2013). Moreover, doctors reported that they did their best to prevent patient care from being impacted, often at the expense of their own health. Similarly, German oncologists named increasing workloads and time pressure as the greatest threat to patient care (Groß et al., 2014). In comparison, collaboration with colleagues was an important aspect of work that negated some of the detrimental impact of challenging working conditions.

At the hospital-level, a study involving 61,168 nurses across 488 European and 617 American hospitals reported that poor working conditions had strong effect sizes with nurse and patient-rated outcomes (Aiken et al., 2012). More specifically, working conditions here represented the amount of control nurses had, the support they received from management, and the quality of relationship with doctors. In the United Kingdom, surveys of hospital staff have revealed that manager support was associated with hospital-level patient satisfaction scores (Raleigh, Hussey, Seccombe, & Qi, 2009); or, that job-related training, effective appraisals, working in effective teams, and supportive managers, predicted patient satisfaction and mortality rates of the hospital (Topakas, Admasachew, & Dawson, 2011).

Reflecting on this, it is clear that there is reason to believe that doctors' psychosocial working conditions should influence the quality of care provided. However, this evidence either stems from non-healthcare sectors or is not specific enough to doctors. Given the different roles and working conditions experienced, it is important to examine what the relevant evidence for this relationship is.

4.1 Why the Need for a Systematic Review?

Considering the practical and policy interest in the topic of psychosocial working conditions and quality of care, an accurate representation of the current state of the evidence will inform a more meaningful discussion. To this student's knowledge there have not been any reviews on this relationship. This absence is accentuated by the presence of such reviews in nursing (Kazanjian, Green, Wong, & Reid, 2005) and general healthcare (Hoff et al., 2004). The presence of systematic reviews on burnout in medical residents (Prins, Gazendam-Donofrio, et al., 2007), doctors' shift work (Reed et al., 2010), and the impact of physicians' positive wellbeing on care provided (Scheepers et al., 2015), all indicate an interest around this as well as a gap in the academic literature.

Within the context of the thesis, this systematic review carries two key functions. Firstly, it provides a comprehensive review of the current state of the literature in relation to the broader doctoral research question. Such a review is valuable in understanding the quality of the evidence, the research designs, the psychosocial working conditions considered, quality-of-care measured, and the theoretical frameworks used. Secondly, the findings of the review inform the development of the subsequent quantitative studies within this thesis. Where possible, the following studies will build on the strengths of past research and consider the limitations in the area to provide a stronger and more meaningful contribution to our understanding of this topic.

4.1.1 What is a systematic review?

The progression of research and science has developed a large body of knowledge and information. Collating multiple studies is more likely to lead to concluding appropriate findings that are valid, defendable, and reliable (Khan, Riet, Glanville, Sowden, & Kleijnen, 2009; Petticrew & Roberts, 2006); the ability to do so is therefore essential to inform researchers, practitioners, and decision-makers. Moreover, comprehensive reviews of the literature allow researchers to identify existing gaps and limitations that can direct future research (Oliver, Peersman, Harden, & Oakley, 1999). However, traditional literature reviews lack transparency and rigour; this opens the possibility for researchers to, consciously or unconsciously, present a bias perspective of a particular research area (Petticrew & Roberts, 2006). Instead, systematic reviews offer researchers a mechanism by which to coherently understand a large body of research whilst at the same time reducing the effect of reviewer bias (Rojon, McDowall, & Saunders, 2011).

What distinguishes systematic reviews from traditional literature reviews is its *explicit* and *systematic* approach in identifying, evaluating, and synthesising appropriate studies relevant to the research question (Khan, Kunz, Kleijnen, & Antes, 2003). This is conducted through the adherence of pre-set guidelines and methodology that allow scrutiny and replication. The basis for a systematic review is a well-defined research question; a comprehensive search strategy; a clear screening process to include and exclude studies; a structured process in extracting information; the appraisal of the quality of studies; and a coherent manner to synthesise the findings (Khan et al., 2009; Petticrew & Roberts, 2006; Rojon et al., 2011). When carried out correctly, systematic reviews provide a crucial method in evidence-based research and practice.

4.1.2 Review aim and research question

The literature reviewed in the preceding chapters suggests it is possible that improved psychosocial working conditions lead to improved doctor health, and subsequent better quality of care provided. However, the vast majority of literature in this area has mainly involved healthcare workers in general. As such, this systematic review asks:

1. What are the psychosocial working conditions faced by doctors? And,
2. How are these working conditions associated with different types of quality-of-care outcomes?

4.2 Methods

4.2.1 Review protocol

A review protocol was developed specifying the research question, search terms, search strategy, data extraction, and study appraisals (Khan et al., 2009). This was informed by guidelines and best practice on systematic reviews (Khan et al., 2003, 2009; Oliver et al., 1999; Petticrew & Roberts, 2006; Rojon et al., 2011; Rush, Shiell, & Hawe, 2004). In order to maintain objectivity, the review protocol was registered with PROSPERO – an international prospective register of systematic reviews (Teoh, Hassard, & Cox, 2015).

4.2.2 Scoping review

Before the systematic review proper was undertaken, a scoping review was conducted to assess whether any reviews related to the research question already existed in the literature (Khan et al., 2003). It also helped refine the research question, initial search strategy, and collection of relevant studies (Petticrew & Roberts, 2006). The scoping review was conducted in four research databases (Medline, Web of Science, EBSCO, and Science Direct). Search terms were based on the key elements of the research question: the emphasis on psychosocial working conditions ("job demands" OR "job control" OR support OR "job resources"), quality of care ("care quality" OR "patient safety" OR "patient outcomes"), and to the profession of interest (doctor OR physician). The Cochrane database, Google, and Google Scholar were also reviewed for relevant reviews or protocols.

No reviews on the proposed research question were found, although two related reviews were found. Neither focused on doctors; the first examined psychosocial working conditions and quality of care within the healthcare sector (Hoff et al., 2004), while the second restricted this to nursing studies (Kazanjian et al., 2005). Related reviews found involving doctors instead focused on burnout (Prins, Gazendam-Donofrio, et al., 2007), prescription error rates (Ross, Bond, Rothnie, Thomas, & Macleod, 2009), salary (Gosden, 1999), and working hours (Fletcher et al., 2005; Moonesinghe et al., 2011; Reed et al., 2010).

4.2.3 Data sources and search terms

Search terms. In line with guidance developed by the Centre for Reviews and Dissemination (Khan et al., 2009), a step by step process was used to develop the search terms in this review. First, the research question was broken down according to its different elements: psychosocial working conditions, quality of care, and doctors. Next, free text terms were generated for each of these elements. These included synonyms and spelling variants (Khan et al., 2009). The free text variants were developed through discussions with supervisors and information from the scoping review. These were then collated, discussed, and refined. In addition, keywords from the studies found in the scoping review were examined to ensure relevant studies would be captured using these search terms. The 'AND' Boolean operator was used between each of the three elements to ensure that at least one free text variant from each element was represented in the search hits.

The psychosocial literature has developed different taxonomies containing various working conditions (Cox et al., 2000; Dewe & Trenberth, 2004), leading to numerous possible terms and variants. Pilot searches revealed that using a large variant of specific psychosocial working conditions yielded a substantially larger amount of search hits without appearing to increase the number of relevant hits. Moreover, numerous search terms generated text fields which were too long for the search text boxes on some electronic databases. Consequently, in discussion with supervisors it was decided to restrict search terms to the broader components that constitute the key work-stress theories; namely, the job-demand-control-support (Johnson & Hall, 1988), the effort-reward imbalance (Siegrist, 1996) and the job demand-resources models (Demerouti et al., 2001). This emulates other reviews involving psychosocial stress (Kivimäki et al., 2006; Stansfeld & Candy, 2006). It substantially reduced the number of search hits obtained during pilot searches to a more manageable number. As an additional control measure, two relevant studies were earmarked and the search strategy implemented to assess whether the targeted studies would be retrieved. The elements and their respective free text variants are presented in Table 4.1 below. Search terms were joined within each element by the 'OR' Boolean operator.

Table 4.1: *Search terms used for each element*

Element	Free text variant
Psychosocial working conditions	"effort-reward" OR "effort reward" OR "job demand" OR "job control" OR "decision latitude" OR "decision authority" OR "job strain" OR "social support" OR "job resources" OR "job stress" OR "work stress" OR "work strain" OR "work-related stress" OR "occupational stress" OR stressor OR "working environment" OR "working conditions" OR "psychosocial risk" OR "psychosocial factor" OR "psychosocial hazard" OR workload
Quality of care	"quality of care" OR error OR "patient safety" OR "patient outcomes" OR "patient satisfaction" OR "adverse impact" OR "adverse event"
Doctor	"house officer" OR physician OR "medical officer" OR "medical resident" OR "medical trainee" OR doctor OR surgeons

Data sources. The search terms were used between the 26th and 31st of January 2015 in seven electronic academic databases. It was updated again on the 19th of April 2017 for any publication since the first search was carried out. Four of these were considered medical databases: Medline; the Health Management Information Consortium (HMIC); the Cumulative Index to Nursing and the Allied Health Literature (CINAHL Plus); and EMBASE. The remaining three databases were EBSCO (including Academic Search Premier, Business Source Premier, PsychArticles and PsychINFO); Science Direct; and Web of Science. Where possible, articles were restricted to journal articles published in the English language. No restrictions were applied in terms of dates.

To increase the scope of the search field, reference review, citation tracking, and internet searches were also carried out (Khan et al., 2009; Petticrew & Roberts, 2006). Reference review is a past-orientated search strategy where relevant references found during full-text article reviews are included into the search process. In contrast, citation tracking encompassed looking up the final included articles on Google Scholar to examine subsequent publications that have referenced the article of interest. This typically allows reviewers to obtain other articles located within a similar conceptual cluster (Khan et al., 2009). Finally, both Google Scholar and Google were used as databases in addition to the traditional academic databases listed above.

4.2.4 Selection of studies

A set of inclusion and exclusion criteria were developed to structure and inform the selection of studies in this review. These criteria were defined through the research question, and drawn from systematic review guidelines and discussion papers (Khan et al., 2003, 2009; Rush et al., 2004). The inclusion and exclusion criteria were reviewed and discussed with supervisors, and are presented in succinct form in Table 4.2. Specifically, these entail the predictor variable, outcome measures, research design, the profession of interest, and language.

Table 4.2: *Inclusion and exclusion criteria*

Inclusion criteria	Exclusion criteria
Predictor variable	
• A psychosocial working condition which can be classed as a job demand or a job resource	• Not a measure of stress symptoms, perceived stress, burnout, or other health variant • Not socio-technical measures of the work environment (i.e., objective measures such as hours worked, shift work, night shifts, surgeon caseload, number of patients seen)

<u>Outcome measure</u>	

• Patient outcomes should be applicable to one of the following types (Department of Health, 2008):
  o Clinical excellence
  o Experience of patients
  o Patient safety
• No restrictions are made as to the source (e.g., doctor, third-party or patient rated; clinical data and records)
	• Not a measure of safety culture
	• Not set within a training, simulation, or experimental settings

<u>Research design</u>	
• Study design should be quantitative based • There must be a direct relationship present between psychosocial working conditions and quality of care	• Literature reviews • Qualitative designs
<u>Profession</u>	
• The occupation being examined must be qualified doctors that are in practice, be it in hospital or in the community	• Not in the allied health profession setting, including care homes and rehabilitation centres • Not students in training that are still in medical school
<u>Language</u>	
• English	• Non-English studies

Predictor variable. Relevant articles had to contain a psychosocial working condition as a predictor variable. These include aspects of work and the management of work, perceived by the employee, that have the potential for psychological or physical harm (Cox & Griffiths, 1995; Parkes & Sparkes, 1998; see Section 2.2 for a detailed review on defining psychosocial working conditions). Variables which refer to doctor health (e.g., burnout), stress or its symptoms, or the socio-technical aspect (e.g., caseload, shift work, hours worked) are not relevant and were excluded.

Outcome measure. Considering the broad definition of care quality, patient outcomes could encompass one of the following: clinical excellence, the experience of the patients, or patient safety. This had to occur within work setting, and consequently outcomes measured during simulation, training, or experimental settings were excluded. The measures could be objective measures or subjective ratings made by the doctor, the patient or an appropriate third-party. A detailed review of quality of care is available in Section 3.1.

Profession of interest. The heterogeneity of the profession means emphasising a particular qualification (in-training/consultant), sector (public, private), specialty, or location (community, acute) would place additional restrictions in an already narrow area; hence, these characteristics were not restricted. Nurses, medical students, and the allied health professions do not fit within the scope of this review and were excluded.

Research design. Only studies that demonstrated a quantitative relationship between the predictor variable and the outcome measure were included. Qualitative studies and literature reviews do not provide a quantitative examination of this relationship.

Language. Included studies were restricted to the English language. This restriction is a potential source of bias that could reduce the generalisability of the findings (Petticrew & Roberts, 2006) and is an acknowledged limitation in this review.

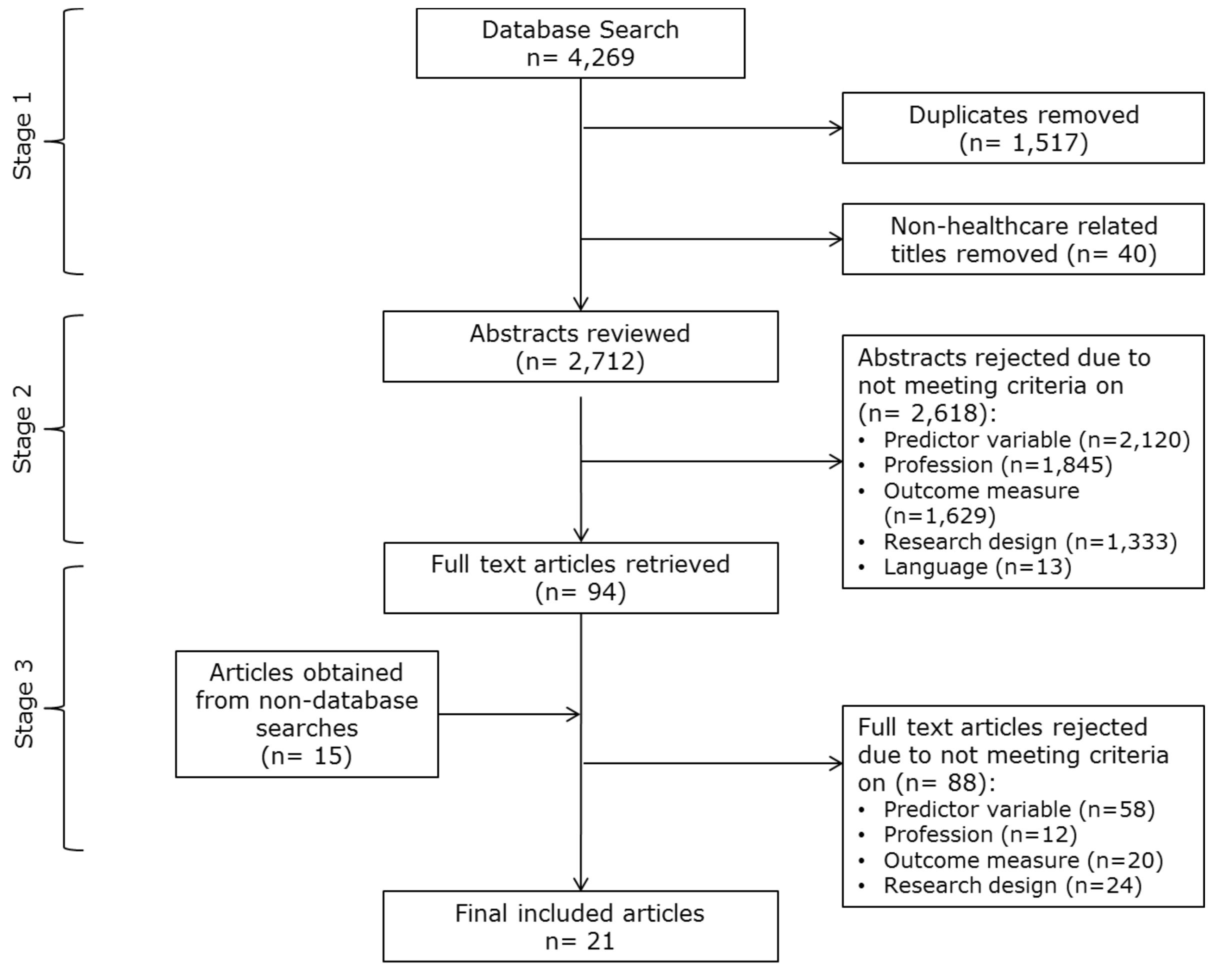

Figure 4.1: Flow Chart of Study Selection Process

4.2.5 Search strategy

The strength of a systematic review lies in the systematic process that guides the search process, allowing consistency, and reducing bias (Khan et al., 2009; Petticrew & Roberts, 2006). This process consisted of three stages: (1) the title review stage, (2) the abstract review stage, and (3) the full-text review stage. The flowchart detailing these steps is presented in Figure 4.1.

Stage 1 - Title review. All hits from the database search were extracted and organised into Endnote Web. In total, 4,269 hits were recorded across the seven databases. After duplicates were removed, 2,752 articles remained. The title of each article was reviewed to evaluate whether the article referred to the healthcare sector. To prevent relevant articles from being excluded, any uncertainty resulted in the article included in the next stage. A further 40 articles were removed, leaving 2,712 articles for the next stage.

Stage 2 - Abstract review. The author, publication year, title, and database source of the remaining articles were extracted into Microsoft Excel for recording purposes. Each abstract was subsequently reviewed on Endnote Web against the criteria set out in Table 4.2: predictor variable, outcome measure, profession, research design, and language. Abstracts that met all five criteria were marked as 'yes' for full-text to be retrieved. To prevent relevant articles from being excluded, the benefit of the doubt was given to any criteria where there was uncertainty. This eased the inclusion of the abstract into the next stage, where review of the full-text article allowed a more informed decision.

In total, 94 abstracts met all five inclusion criteria and were included for full-text review. For the excluded abstracts, the most common criteria failed was related to the predictor variable (n=2,120) and profession (n=1,845). These were followed by outcome measure (n=1,629) and research design (n=1,133). The lowest number of abstract exclusions were those associated with language criteria (n=13).

As an additional control measure, a random selection of 18% of all abstracts (n=465) were independently assessed by a second reviewer who is a safety practitioner. There were 460 agreements and 5 disagreements between both reviewers, a 98.9% agreement rate. It has been argued that a simple agreement percentage between raters is susceptible to chance (J. Cohen, 1960), hence Cohen's kappa was also calculated for a more robust measure of inter-observer agreement. Cohen's kappa between both raters was .833, which was considered "strong".

Stage 3 - Full-text review. The full-text articles for the remaining 94 abstracts were subsequently retrieved and reviewed. A further 15 articles referenced within the full-text articles that appeared relevant were also retrieved and added into the review list. A Microsoft Excel spreadsheet, similar to the one used in Stage 2 was developed to record the assessment of each article against the inclusion/exclusion criteria. Of the remaining 88 articles, 58 were rejected for not referring to a psychosocial predictor variable, 12 were rejected for not focusing on doctors as a sample, 20 for not having a suitable outcome measure, and 24 for not having an appropriate research design. All of the articles were in English. The number of articles that met all the inclusion criteria, and therefore were included in this systematic review was 21. The same second reviewer from stage 2 reviewed a random sample of 16 (17%) full-text articles. There was 100% agreement between both raters; a perfect Kappa score of 1.

4.2.6 Data extraction

A standardised data extraction form (Appendix I) was developed that allowed data extraction to be conducted in a transparent and consistent manner which provides rigour and consistency, thereby improving validity and reliability (Higgins & Deeks, 2011). This form was piloted with three of the included studies. For consistency, extractions were carried out by the first reviewer, and reviewed by the second independent reviewer. The data extraction form was divided into the six sections outlined below:

Study information. Questions in this section elicited information around the data extraction process, including article title, authors, and the date of the extraction.

Study background. This section extracted information on the background of the study being reviewed, with questions on the country the study was set in, the research question, the study design, the (psychosocial) predicting variable(s), the outcome variable, any mediating variables, and what the theoretical framework of the study was.

Sample and measures. The first part of this section pertained to the sample size, the number of recruiting sites, the sampling method, and the response rate. The second part related to the measures used to assess each construct/variable. The extraction form allowed the researcher to record the instrument name, number of items, whether the measure is self-report, and the internal reliability of the instrument.

Key findings. The main findings of the study related to the psychosocial and outcome measure relationship was extracted in this section. The predictor and dependent variable, the effect size, and the sample size for that particular relationship, were all recorded in a table. Where appropriate, findings could be elaborated under the "other findings" header.

Limitations. This section listed all the limitations identified by the study authors.

Study quality. Systematic review guidelines advocate the inclusion of higher quality research methodologies and design (Petticrew & Roberts, 2008; Oxman, 1994). The purpose of this is twofold (Khan et al., 2003; Rush et al., 2004). Firstly, it restricts articles to those which are likely to have higher internal validity. And secondly, it allows for easier comparisons and synthesis of findings between the selected studies. However, the methodologies in the social sciences and applied setting vary significantly and are unlikely to meet the standards advocated for clinical and medical based reviews. Hence, developing exclusion criteria based on methodological design and quality would result in a reduction in the number of included studies. Instead, the inclusion of weaker studies presents a more realistic answer based on the available evidence, and helps to understand the methodological limitations that exist within this area (Hassard, Teoh, Visockaite, Dewe, & Cox, 2017a).

The review utilised the Medical Education Research Study Quality Instrument (MERSQI; Reed et al., 2007), a ten item checklist measuring study design, number of institutions, response rate, type of data, internal structure, content validity, criterion validity, appropriateness of data analysis and outcome levels. The ten items are organised into six domains, each with a maximum score of three points. Possible MERSQI scores range from 5 to 18. Items which are not applicable to the study are discounted, and that study's MERSQI score is subsequently extrapolated to represent an 18 point scale. The MERSQI has been validated against journal impact factors, funding received, and expert quality ratings, as well as having high inter-rater reliability and internal consistency (Reed et al., 2007, 2008).

4.2.7 Data synthesis

The final aspect synthesised the data extracted from the study. This was based on the concepts introduced in the research question and the data extraction process. The data extraction forms were used to inform the aggregation and integration of the included studies which are reviewed in the Results section below.

4.2.8 Statistical analysis

To supplement the summary of results from included studies, *r* coefficients were extracted to allow examination of effect sizes. Where more than one effect size was present in a relationship these were meta-analysed using Hedges and Vevea's (Hedges & Olkin, 1985; Hedges & Vevea, 1998) random effect model. This allowed mean effect sizes that are weighted in favour for studies with larger sample sizes. Studies that reported results as mean differences or as odds ratios were first converted into *r* coefficients (Borenstein, Hedges, & Higgins, 2009; J. Cohen, 1960). Where *r* coefficients were not reported or available from authors, standardised regression values were used instead. These are strongly correlated with *r* coefficients and are a suitable replacement in meta-analysis (Borenstein et al., 2009; Bowman, 2012). To prevent double-counting, average coefficients were used for multiple estimates of the same relationship within the same study (Borenstein et al., 2009; Nahrgang et al., 2011).

4.3 Results

The details of the 21 included studies are outlined in Table 4.3. Eight studies came from Germany (Ansmann et al., 2013, 2014; Bernburg et al., 2016; Krämer, Schneider, Spieß, Angerer, & Weigl, 2016; Loerbroks, Weigl, Li, & Angerer, 2016; Mache, Danzer, Klapp, & Groneberg, 2013; Mache et al., 2012; Weigl et al., 2015) and six from the United States (An et al., 2013; Bertram et al., 1992; Bertram, Hershey, Opila, & Quirin, 1990; Dollarhide et al., 2013; Feddock et al., 2005; Linzer et al., 2009). The remaining countries included: Israel (*n*=3; Naveh et al., 2015; Shirom et al., 2006; Stern et al., 2008), the United Kingdom (*n*=2; Baldwin et al., 1997; McKinstry et al., 2007), the Netherlands (*n*=1; Zwaan, 2012), and Sweden (*n*=1; Tucker et al., 2012).

Medical disciplines represented included ambulatory care (Bertram et al., 1992), general practice (McKinstry et al., 2007), hospital physicians (Bernburg et al., 2016; Dollarhide et al., 2013; Krämer et al., 2016), internal medicine (Bertram et al., 1990; Feddock et al., 2005), oncology (Ansmann et al., 2013, 2014), paediatrics (Weigl et al., 2015), surgery (Mache et al., 2013, 2012), multiple disciplines (An et al., 2013; Linzer et al., 2009; Shirom et al., 2006; Stern et al., 2008; P. Tucker et al., 2012), as well as residents (Naveh et al., 2015; Zwaan, 2012) and junior doctors (Baldwin et al., 1997; Loerbroks et al., 2016).

Table 4.3: *Key characteristics of included studies*

Author	Study size	Country	Psychosocial working condition	Quality-of-care measurement	Outcome rated by	Theoretical framework	Data analysis
An *et al*. (2013)	422 general internists and family physicians, and 1384 patients	USA	Burden of difficult encounters (eight items)	Quality of care for hypertension (Chart audit of successful blood pressure control); Quality of care for diabetes (Chart audit for successful control of haemoglobin A1c and blood pressure); Errors for hypertension and diabetes (Chart audit of guideline non-adherence and missed opportunities for prevention or management)	Observer	None	Latent cluster analysis
Ansmann *et al*. (2013)	864 oncologists and 1462 patients	Germany	Work overload (one item); Time pressure (one item)	Patient satisfaction (Cologne Patient Questionnaire; three items)	Patient	None	Multilevel modelling
Ansmann *et al*. (2014)	348 hospital physicians and 1844 patients	Germany	Decision latitude, psychological job demands, physical demands, and work postures demands (Job Content Questionnaire, unspecified number of items) Social support from colleagues (Unspecified number of items)	Patient satisfaction with support (Cologne Patient Questionnaire; three items)	Patient	None	Multilevel modelling
Baldwin *et al*. (1997)	142 junior doctors	United Kingdom	Feeling overwhelmed (Four items from the Attitude to Work Scale) Effective learning and skill use (Four items from the Attitude to Work Scale)	Subjective work performance (Number of mistakes made in the previous year)	Self	None	Correlations
Bernburg et al. (2016)	435 hospital doctors	Germany	Job demands (Quantitative demands, emotional demands); Job resources (Influence at work, possibilities for development, degree of freedom at work, sense of community, feedback, quality of leadership, social support, social relationships); assessed by the Copenhagen Psychosocial Questionnaire (items not reported)	Work ability (7 items)	Self	Job demands-resources	Correlations and Hierarchical Regressions

Bertram *et al*. (1990)	48 internal medicine physicians	USA	Task mental workload (Ten items)	Physician satisfaction with care provided (one item); Physician self-rated quality of care (one item)	Self	None	Correlations
Bertram *et al*. (1992)	22 residents in ambulatory care	USA	Task mental workload (Six items)	Physician self-rated satisfaction with care provided (one item); Physician observer-rated quality of care (one item); Personal interaction performance score (Chart audit); Technical performance score (Chart audit)	Various	None	Correlations
Dollarhide *et al*. (2013)	185 hospital physicians	USA	Workload (NASA Task Load Index with six items)	Medical events (self-reporting electronic tool which collects data on type and severity of a medication event.	Self	None	Regression analyses
Feddock *et al*. (2005)	42 internal medicine residents with 168 matched patient resident dyads	USA	Workload (One item)	Patient satisfaction (Seven items)	Patient	None	Regression analyses
Krämer et al. (2016)	95 hospital physicians	Germany	Patient demands (5 items); time pressure (5 items); social stressors (3 items)	Quality of care (Three items)	Self	Job demands-resources	Path models
Linzer *et al*. (2009)	422 general internists and family physicians and 1795 patients	USA	Time pressure (Recorded average time allocated for examinations vs. estimated time needed to provide quality care); Office pace (One item); Work control (14 item Physician Worklife Study)	Quality of care: Control of blood pressure for hypertension, control of haemoglobin A l c and blood pressure for diabetes, stability of signs and symptoms for heart failure (audio-recorded visits) Treatment errors: missed treatment opportunities, inattention to behavioural factors, guideline non-adherence and defined prevention errors (audio-recorded visits)	Observer	None	Multilevel modelling

Loerbroks et al. (2016)	416 junior physicians	Germany	Effort and rewards (23 item)	Self-reported perceived quality of care (8 items)	Self	Effort-reward imbalance	Regression analyses
Mache *et al*. (2012)	98 surgeons and 122 of their patients	Germany	Job demands (Quantitative demands, emotional demands and demands hiding emotion); Job resources (Possibilities for development, degree of freedom, influence at work, sense of community, social support, quality of leadership, feedback at work); assessed by the Copenhagen Psychosocial Questionnaire (items not reported)	Patient satisfaction (12 items)	Patient	None	Correlations
Mache *et al.* (2013)	123 surgeons	Germany	Job demands (Quantitative demands, emotional demands, cognitive demands and demands for hiding emotion); Job resources (Possibilities for development, degree of freedom, influence at work, social relationships, social support, quality of leadership, feedback at work); assessed by the Copenhagen Psychosocial Questionnaire (items not reported)	Work ability (7 items)	Self	None	Correlations and Hierarchical Regressions
McKinstry *et al*. (2007)	198 GPs and an average of 49.6 patients per GP	United Kingdom	Work control and support (each measured by one item from the 13 item Morale Assessment in General Practice Index)	Patient satisfaction dimensions on quality of communication and enablement (measured by the General Practice Assessment Questionnaire (items not reported)	Patient	None	Correlations
Naveh *et al.* (2015)	142 residents	Israel	Autonomy (three items); Consultation with physicians (two items)	Error rate (four items)	Department head	None	Hierarchical linear regression
Shirom *et al*. (2006)	890 specialists in cardiology, dermatology, general surgery, gynaecology, ophthalmology and otolaryngology	Israel	Autonomy (Ten items) Overload (Nine items)	Quality of care (15 items from the original 22 item Service Quality Scale)	Self	Person environment fit & Conservation of resources	Structural equational modelling

Stern *et al*. (2008)	123 residents	Israel	Autonomy (Four items)	Treatment errors rated by senior nurse (Number of 12 different types of mistakes)	Nurse	None	Multilevel modelling
Tucker *et al.* (2012)	1534 doctors	Sweden	Work time control (One item)	Concerns on patient safety (One item)	Self	None	Mediation
Weigl *et al*. (2015)	88 paediatricians	Germany	Effort and rewards (23 item)	Prevention and disease management performance (11 items) Self-reported perceived quality of care (Two items)	Self	Effort-reward imbalance	Regression analyses
Zwaan (2012)	210 patients and their attending resident	Netherlands	Subjective workload (one item)	Number of patient harm incidents or diagnostic errors in patient charts	Observer	None	Logistical regression

Table 4.4: *MERSQI criteria for included studies*

Author (Year)	Study design (Number of Institutions)	Response rate	Validity of psychosocial measure: Internal structure	Content validity	Criterion validity	Type of output data	Data analyses: Appropriateness	Sophistication	Outcomes	MERSQI score
Mache *et al*. (2012)	Single group cross-sectional (7 hospitals)	55%	Reported	Reported	Reported	Assessment by patient	Appropriate	Correlations	Satisfaction, attitudes, perception	13.5
Linzer *et al*. (2009)	Single group cross-sectional (119 practices)	59.6%	Reported	NR	Reported	Chart audits by researchers	Appropriate	Regressions	Behaviours	13.5
Stern *et al*. (2008)	Single group cross-sectional (2 teaching hospitals)	80%	Reported	NR	Reported	Assessment by nurse	Appropriate	Multilevel modelling	Behaviours	13.5
Naveh *et al.* (2015)	Single group cross-sectional (2 teaching hospitals)	80%	Reported	NR	Reported	Assessment by senior physicians	Appropriate	Regressions	Behaviours	13.5
Bertram *et al*. (1992)	Single group cross-sectional (1 clinic)	100%	Reported	NR	Reported	Assessment by evaluator & doctor	Appropriate	Correlations	Behaviours	13
Zwaan (2012)	Single group cross-sectional (5 hospitals)	80.4%	NA	NR	Reported	Chart audits by researchers	Appropriate	Regressions	Behaviours	12.7
An *et al*. (2013)	Single group cross-sectional (119 practices)	59.8%	NR	NR	Reported	Chart audits by researchers	Appropriate	Cluster analysis	Behaviours	12.5
Dollarhide *et al*. (2013)	Non-randomised two groups (4 hospitals)	75.8%	NR	Reported	NR	Assessment by doctor	Appropriate	Regressions	Behaviours	12
Ansmann *et al*. (2013)	Single group cross-sectional (31 hospitals)	46.4%	NA	Reported	NR	Assessment by patient	Appropriate	Multilevel modelling	Satisfaction, attitudes, perception	11.6
Shirom *et al*. (2006)	Single group cross-sectional (Multiple settings)	63%	Reported	Reported	Reported	Assessment by doctor	Appropriate	Structural equation modelling	Satisfaction, attitudes, perception	11.5

Krämer *et al.* (2016)	Single group longitudinal design (2 hospitals)	Time 1 53%; Time 2 47%	Reported	Reported	Reported	Assessment by doctor	Appropriate	Path models	Satisfaction, attitudes, perception	11.5
Loerbroks *et al.* (2016)	Single group cross-sectional (Multiple hospitals)	69%	Reported	Reported	Reported	Assessment by doctor	Appropriate	Path models	Satisfaction, attitudes, perception	11.5
Mache *et al.* (2013)	Single group cross-sectional (10 hospitals)	63%	Reported	Reported	NR	Assessment by doctor	Appropriate	Regressions	Satisfaction, attitudes, perception	11.5
Ansmann *et al*. (2014)	Single group cross-sectional (35 hospitals)	46%	NR	Reported	NR	Assessment by patient	Appropriate	Multilevel modelling	Satisfaction, attitudes, perception	11
Baldwin *et al*. (1997)	Single group cross-sectional (1 hospital)	95%	Reported	NR	Reported	Assessment by doctor	Appropriate	Correlations	Behaviours	11
Bernburg *et al.* (2016)	Single group cross-sectional (12 departments)	61.8%	Reported	Reported	NR	Assessment by doctor	Appropriate	Regressions	Satisfaction, attitudes, perception	10.5
McKinstry *et al*. (2007)	Single group cross-sectional (Multiple practices)	62%	NR	Reported	NR	Assessment by patient	Inappropriate	Correlations	Satisfaction, attitudes, perception	10.5
Weigl *et al*. (2015)	Single group cross-sectional (1 hospital)	73.8%	NR	Reported	Reported	Assessment by doctor	Appropriate	Regressions	Satisfaction, attitudes, perception	10.5
Bertram *et al*. (1990)	Single group cross-sectional (2 clinics)	98%	Reported	NR	Reported	Assessment by doctor	Appropriate	Correlations	Satisfaction, attitudes, perception	10.5
Tucker *et al.* (2012)	Single group cross-sectional (Multiple settings)	53.1%	NA	NR	Reported	Assessment by doctor	Appropriate	Regressions	Satisfaction, attitudes, perception	10
Feddock *et al*. (2005)	Single group cross-sectional (1 clinic)	NR	NA	NR	NR	Assessment by patient	Inappropriate	Regressions	Satisfaction, attitudes, perception	8.5

4.3.1 Study quality

The quality score of the included studies on the MERSQI ranged from 8.5 to 13.5 (Table 4.4). The observed mean of 11.63 out of 18 (*SD*=1.35) is slightly higher than the majority of mean MERSQI scores in reviews using this indicator (Reed et al., 2007, 2008, 2010; Scheepers et al., 2015). Nineteen studies used cross-sectional designs involving a single group of participants. One further study compared two non-randomised groups (Dollarhide et al., 2013), while another used a longitudinal design (Krämer et al., 2016). Four studies sampled doctors from a single institution, with the remaining studies recruiting from two (*n*=4) or more (*n*=13) institutions. Seven studies had a high response rate (>75%). Medium (50-74%) and low (>50%) response rates were reported by 11 and three studies respectively.

Any measure used should be valid for its intended purpose, and this is reflected in the items that focus on the internal, content, and criterion validity of the outcome measures being used. Only five studies reported on all three. Validity of internal structure did not apply to four of the studies (Ansmann et al., 2013; Feddock et al., 2005; P. Tucker et al., 2012; Zwaan, 2012) where single item measures of patient satisfaction were utilised, while a further five studies did not report any information on this. Where this was done, Cronbach's alpha was used (*n*=11). However, it is worth noting that the measures of technical performance (α=.52) and personal interaction quality (α=-.51) for observed physicians in Bertram et al.'s (1992) study were both lower than the commonly accepted score of α>.70. Content validity was addressed in ten studies (Ansmann et al., 2013, 2014; Dollarhide et al., 2013; Mache et al., 2013, 2012; McKinstry et al., 2007; Shirom et al., 2006; van den Hombergh et al., 2009; Weigl, Hornung, Angerer, Siegrist, & Glaser, 2013; Weigl et al., 2015), with all these studies describing the measure as being established or validated elsewhere.

Quality-of-care outcomes were derived from doctor self-reports (*n*=9), patient ratings (*n*=6), chart audits (*n*=3), colleague ratings (*n*=2), or a combination of methods (*n*=1). No study considered clinical outcomes. Instead, 13 studies measured satisfaction, attitudes, and perceptions; and eight measured behaviours.

4.3.2 Psychosocial working conditions and quality of care

Only five studies utilized a theoretical framework, namely: the job demands-resources model (*n*=2; Bernburg, Vitzthum, Groneberg, & Mache, 2016; Krämer et al., 2016), the effort-reward imbalance model (*n*=2; Loerbroks, Weigl, Li, & Angerer, 2016; Weigl, Schneider, Hoffmann, & Angerer, 2015), and the person-environment fit model (*n*=1; Shirom, Nirel, & Vinokur, 2006). All five studies that did examine mediation tested whether negative wellbeing mediated the relationship between job demands and quality of care. More specifically, burnout functioned as a mediator in two (Shirom et al., 2006; Weigl et al., 2015) of the three studies that tested for this indirect effect. Depressive symptoms (Loerbroks et al., 2016) also functioned as a mediator, while job satisfaction (An et al., 2013) and irritation (Krämer et al., 2016) did not. The one study that did examine an interaction effect (Stern et al., 2008) demonstrated that high job autonomy, when occurring in an environment which did not encourage learning, was associated with an increase in the number of treatment errors made.

All included studies tested for a direct relationship between job demands or job resources with quality of patient care. Within each section below, the meta-analyses for these relationships are first presented, followed by a brief description of the results from each individual study. The extracted data were categorised across two dimensions. *Explanatory measures* were categorised as a job demand or resource. *Outcome measures* were grouped by their foci of care quality (Department of Health, 2008): (i) clinical excellence (including, subjective work performance, chart audits, and self-rated care quality of care provided); (ii) patient safety, represented by the number of self-reported or observer-assessed errors; and (iii) patient experience (e.g., patient satisfaction, patient-rated quality of care).

4.3.2 Summary of evidence examining psychosocial working conditions and clinical excellence

Job demands and clinical excellence. Ten studies (An et al., 2013; Bernburg et al., 2016; Bertram et al., 1992, 1990; Krämer et al., 2016; Linzer et al., 2009; Loerbroks et al., 2016; Mache et al., 2013; Shirom et al., 2006; Weigl et al., 2015) examined 15 relationships between six types of job demands and clinical excellence (See Appendix II for a list of individual relationships). Table 4.5 presents the meta-analyses of these relationships. However, the studies examining *demanding patients* (An et al., 2013) and *time pressure* (Linzer et al., 2009) reported percentage changes and unstandardized regression coefficients respectively; neither allowed meta-analysis. Of the relationships involving more than one study, the largest effect sizes were observed for *higher-order job demands* (r=-.30; CI:-.37, -.22; k=3), followed by *emotional demands* (r=-.23; CI:-.30, -.16; k=2) and *perceived workload* (r=-.21; CI:-.26, -.16; k=5). *Time pressure* (r=-.62; CI:-.73, -.48; k=1) and *social* *conflict* (r=-.27; CI:-.53, -.22; k=1) also correlated negatively with clinical excellence, however only one study tested each of these relationships.

Table 4.5: *Job demands effect sizes mapped to outcomes*

Job demand	Definition	Clinical excellence	Patient safety	Patient experience
Perceived workload	Perception of additional or excessive work demands	**r=-.21** (CI: -.26, -.16) k=5	**r=.10** (CI: .02, .18) k=3	r=.02 (CI: -.25, .28) k=2 (¤[c])
Demanding patients	Frequency of challenging patient behaviours	◊[a]	×[a]	
Time pressure	The difference between time allocated for treatment compared to the estimated time needed to provide quality care	**r=-.62** (CI: -.73, -.48) k=1 (¤[d])	¤[b]	r=-.24 (CI: -.55, .13) k=1
Physical load	Experience of the continuous physical exertion			r=-.12 (CI: -.35, .12) k=1
Emotional demands	How emotionally demanding the job is and how emotionally involved doctors become	**r=-.23** (CI: -.30, -.16) k=3		
Social conflict	Conflicting relationships with direct colleagues, supervisors, and co-workers	**r=-.37** (CI: -.53, -.18) k=1		
Higher-order job demands	The composite of multiple facets of job demands, and typically exist as a second-order factor	**r=-.30** (CI: -.37, -.22) k=3		**r=-.38** (CI: -.47, -.29) k=1

Note. r: correlation effect size; CI: Lower and upper 95% Confidence Interval; k: number of studies; Bold denotes significant relationships; ◊ expected findings found; ¤ predicted results not supported; ×results opposite to that predicted
[a]An *et al.* (2013) only reported percentage change between high physicians experiencing high and low patient demands
[b]Linzer *et al.* (2009) reported unstandardized regression coefficients.
[c]excludes study by Feddock *et al.* (2005) who only reported correlation coefficients for significant items.

Perceived workload and clinical excellence. Five studies showed perceived workload to be associated with clinical excellence. Specifically, perceived workload negatively correlated with: US residents' satisfaction with care provided (r=-.46; Bertram et al., 1990); German surgeons' (r=-.24; Mache et al., 2013) and hospital physicians' (r=-.20; Bernburg et al., 2016) self-rated work ability; and with self-rated quality of care among Israeli specialists (β=-.15; Shirom et al., 2006). Shirom and colleagues (2006) also found that the link between overload and quality of care was partially mediated by burnout. In the fifth study (Bertram et al., 1992), perceived workload of US ambulatory residents correlated with both self-rated (r=-.67) and observer-rated (r=-.38) performance. It also showed curvilinear relationships between perceived workload and both forms of performance.

Demanding patients and clinical excellence. Only one study focused on demanding patients and clinical excellence. A review of medical charts found US family physicians who experienced high exposure to difficult patient demands had a 7.68% lower overall care quality score than those with low exposure (An et al., 2013). However, when focusing specifically on diabetes or hypertension management, no difference on quality of care was observed between the high and low exposure groups. Furthermore, neither burnout nor job satisfaction was observed to function as a mediator here.

Time pressure and clinical excellence. Krämer et al.'s (2016) longitudinal study found that German hospital physicians' perceived time pressure predicted self-rated quality of care one year later (β=-.19). Irritation was not found to mediate this relationship. However, a cross-sectional study of US doctors found only three out of nine relationships between time pressure (during the first examination, during follow-up, and general office pace) and three observer-rated care measures (total, hypertension-related, and diabetes-related) were significant (Linzer et al., 2009).

Emotional demands and clinical excellence. Two out of the three studies that examined emotional demands found it negatively associated with clinical excellence, including among German surgeons (r=-.21; Mache et al., 2013) and residents (r=-.20; Bernburg et al., 2016). In the third, how emotionally demanding their interaction with patients was did not predict German hospital physicians' self-reported quality of care one year later (Krämer et al., 2016).

Social conflict and clinical excellence. The only study to examine social conflict found that perceived difficulties working with colleagues and supervisors by German hospital physicians predicted lower self-reported quality of care scores one year later (β=-.15; Krämer et al., 2016). This relationship was not mediated by doctors' feelings of irritation.

Higher-order job demands and clinical excellence. Three studies, all from Germany, examined higher-order job demands. This represents a latent second-order factor of multiple facets of job demands. In the first, job demands (comprising quantitative, emotional, demands for hiding emotions, and cognitive demands) explained 10% of the variance of surgeons' work ability (Mache et al., 2013). The remaining two studies both measured time pressure, interruptions, physical demands, and long working hours, and found that these collectively were negatively associated with self-rated quality of care in paediatricians (β=-.49; Weigl et al., 2015) and hospital physicians (β=-.24; Loerbroks et al., 2016). These were respectively mediated by emotional exhaustion and depressive symptoms.

Job resources and clinical excellence. Six studies examined the association between types of job resource and clinical excellence (Bernburg et al., 2016; Linzer et al., 2009; Loerbroks et al., 2016; Mache et al., 2013; Shirom et al., 2006; Weigl et al., 2015). Positive correlations between the examined types of job resource and clinical excellence were present in 13 out of 15 relationships (Appendix II). The largest effect sizes were observed (Table 4.6) for higher-order job resources (*r*=.28, CI:.21, .35, *k*=3) and autonomy (*r*=.29, CI:.24, .34, *k*=2), followed by job control (*r*=.21, CI:.12, .28, *k*=2), learning and development (*r*=.17, CI:.09, .25, *k*=2), social support from colleagues (*r*=.17, CI:.089, .25, *k*=2), and supervisor support (*r*=.13, CI:.04, .21, *k*=2). This excluded Linzer et al.'s (2009) relationship between job control and clinical excellence as this only reported unstandardized regression coefficients.

Table 4.6: *Job resource effect sizes mapped to outcomes*

Job resource	Definition	Clinical excellence	Patient safety	Patient experience
Autonomy	The freedom to make decisions on how to perform work tasks	***r*=.29** (CI: .24, .34) *k*=3	*r*=.-.02 (CI: -.14, .11) *k*=2	
Job control	How much influence doctors have over their work environment	***r*=.21** (CI: .12, .28) *k*=2 (¤[a])	***r*=-.180** (CI: -.23, -.13) *k*=1 (¤[a])	*r*=.17 (CI: -.18, .47) *k*=1 (¤[b])
Learning & development	The opportunities to learn and develop professionally	***r*=.17** (CI: .09, .25) *k*=2	***r*=.-16** (CI: -.27, -.04) *k*=2	
Social Support - Colleagues	The emotional, informational, and tangible support from colleagues	***r*=.17** (CI: .09, .25) *k*=2		*r*=.14 (CI: -.12, .38) *k*=1 (¤[b])
Supervisors support	The emotional, informational, and tangible support from supervisors	***r*=.13** (CI: .04, .21) *k*=2		*r*=.14 (CI: -.12, .38) *k*=1
Higher-order job resources	The composite of multiple facets of job resources, and typically exist as a second-order factor	***r*=.28** (CI: .21, .35) *k*=3		***r*=.42** (CI: .33, .50) *k*=1

Note. *r*: correlation effect size; CI: Lower and upper 95% Confidence Interval; *k*: number of studies; Bold denotes significant relationships; ¤ predicted results not supported
[a]excludes Linzer et al. (2009) who reported unstandardized regression coefficients
[b]excludes McKinstry *et al.* (2007) who did not report *r* coefficients for insignificant relationships

Autonomy and clinical excellence. Autonomy was assessed by three studies. It correlated positively with Israeli specialists (β=.37; Shirom et al., 2006), German medical residents' (r=.10; Bernburg et al., 2016) and German surgeons' (r=.32; Mache et al., 2013) self-rated quality of care.

Job control and clinical excellence. In terms of job control, two studies showed it to correlate positively with self-reported work ability among German surgeons (r=.39; Mache et al., 2013) and medical residents (r=.15; Bernburg et al., 2016). However, in the United States this relationship was not evident; Linzer et al. (2009) reported that only one of three relationships between physicians' job control and observer-rated quality-of-care outcomes was significant.

Learning and development, colleague support and supervisor support with clinical excellence. The individual relationships between the other three job resources with clinical excellence were examined in two German studies. Increased supervisor support (r=.25), support from colleagues (r=.30), and opportunities for learning and development (r=.32) were associated with higher surgeons' self-reported work ability (Mache et al., 2013). Among medical residents (Bernburg et al., 2016), opportunities for learning and development (r=.13) and social support from colleagues (r=.13), but not supervisor support (r=.09), positively correlated with self-reported work ability.

Higher-order job resources and clinical excellence. Three studies focused on higher-order job resources. Mache et al. (2013) examined the relationship between higher-order job resources (consisting of eight types of job resources) and self-rated performance in Germany surgeons. They found a positive correlation (r=.42). Also in Germany, Weigl et al. (2015) and Loerbroks et al. (2016) showed higher-order job resources (which included perceived salary, promotion prospects, esteem, job security) to respectively predict paediatricians' (β=.44) and hospital doctors' self-rated care quality (β=.20). Mediation analyses demonstrated that these were respectively mediated by emotional exhaustion and depressive symptoms.

4.3.3 Summary of evidence examining psychosocial working conditions and patient safety

Job demands and patient safety. Five studies examined the association between three types of job demands and doctors' error rates (An et al., 2013; Baldwin et al., 1997; Dollarhide et al., 2013; Linzer et al., 2009; Zwaan, 2012; Appendix III). Table 4.5 reports a positive relationship between perceived workload and errors (r=.10; CI: 02, .18; k=3). However, the studies using *demanding patients* and *time pressure* did not report effect sizes that allowed meta-analyses.

Perceived workload and patient safety. Three studies reported a significant relationship between perceived workload and patient safety. Higher levels of perceived workload reported by Dutch and British residents, respectively were associated with more patient harm incidents or diagnostic errors in patient charts (*OR*=1.10; Zwaan, 2012) and self-reported mistakes in the previous year (*r*=.22; Baldwin et al., 1997). Lastly, when US hospital physicians were prompted to complete surveys on a handheld device at random intervals, results indicate that on days where a medical event (e.g., administration error, near miss) occurred, higher workloads was recorded than on non-event days (Dollarhide et al., 2013).

Demanding patients and patient safety. The only study to examine demanding patients and patient safety found US physicians in the high exposure group (i.e., high patent demands) reported lower error rates (5.57%), compared to those in the low exposure group (An et al., 2013). However, when specific errors in relation to diabetic or hypertension care were compared, no difference was observed between the high and low exposure groups.

Time pressure and patient safety. In the one study of US physicians that tested relationships between time pressure and patient safety. None of the 12 relationships between the three types of time pressure (during the first examination, during follow-up, and general office pace) and four error measures (total, prevention-related, hypertension-related, diabetes-related; Linzer et al., 2009) were significant.

Job resources and patient safety. Five studies tested the job resources and patient safety relationship (Baldwin et al., 1997; Linzer et al., 2009; Naveh et al., 2015; Stern et al., 2008; P. Tucker et al., 2015; Appendix III). Table 4.6 shows that both *job control* (*r*=-.18, CI:-.23, -.13, *k*=1) and opportunities for learning and development (*r*= -.16, CI: -.27, -.04, *k*=2) negatively correlated with errors made. No relationship involving *autonomy* was observed (*r*=-.02, CI:-.14, .11, *k*=2). Linzer et al.'s (2009) study was excluded from Table 4.6 as the unstandardized regression coefficients reported for *job control* were unsuitable for meta-analysis.

Autonomy and patient safety. Neither of the two Israeli studies, both examining residents, found autonomy to be related to either nurse-reported (Stern et al., 2008) or self-reported (Naveh et al., 2015) errors.

Job control and patient safety. Although job control correlated with Swedish physicians' concern for patient safety (β=-.18; Tucker et al., 2015), a second study of US physicians reported no relationship between job control and four different error outcomes (Linzer et al., 2009).

Learning and development and patient safety. In terms of learning and development, self-reported error rates negatively correlated with residents' perception of consulting physicians and familiarity with the medical literature in Israel (r=-.14; Naveh et al., 2015), and with working in an effective learning environment in the United Kingdom (r=-.18; Baldwin et al., 1997).

4.3.4 Summary of evidence examining psychosocial working conditions and patient experience

Job demands and patient experience. Four studies tested job demands in relation to patient experience (Ansmann et al., 2014, 2013; Feddock et al., 2005; Mache et al., 2012; Appendix IV). Feddock et al. (2005) did not report nonsignificant coefficients and their study was excluded from the meta-analyses in Table 4.5. Higher-order job demands (r=-.38; CI:-.47, -.29; k=1), but not perceived workload (r=.02; CI:-.25, .28; k=1) nor time pressure (r=-.24; CI:-.55, .126; k=1), demonstrated a significant negative association with patient experience.

Perceived workload and patient experience. Although three studies tested for a relationship between perceived workload and patient experience, none reported significant results. The outcomes measured here were patient satisfaction towards US residents (Feddock et al., 2005), and patient satisfaction with support from German doctors (Ansmann et al., 2013, 2014).

Time pressure and patient experience. Only one study linked time pressure to patient experience; when German oncologists perceived high time pressure their patients reported lower levels of satisfaction with the support provided by their doctor (OR=0.41; Ansmann et al., 2013).

Physical demands and patient experience. In Ansmann et al.'s (2014) study, perceived physical job demands reported by doctors from German breast cancer centres' was measured as physical activity and work postures. The former showed a significant correlation with patient satisfaction, while the latter did not.

Higher-order job demands and patient experience. When Mache et al. (2012) utilised a higher-order measure of job demands comprising of quantitative demands, emotional demands, and demands for hiding emotion in a sample of German surgeons, it correlated negatively with patient satisfaction (*r*=-.38).

Job resources and patient experience. Four types of job resources from three studies (Ansmann et al., 2014; Mache et al., 2012; McKinstry et al., 2007) were tested for correlations with patient experience (Appendix IV). However, the study by McKinstry et al. (2007) was excluded from the meta-analyses (see Table 4.6) as it did not report correlation coefficients for insignificant results involving job control and colleague support. Results from the meta-analysis observed patient experience to positively correlate with *higher-order job resources* (*r*=.42, CI:.33, .503, *k*=1), but not with the individual types of job resources: *colleague support*, *supervisor support*, or *job control*.

Job control and social support with patient experience. Two studies tested whether job control and social support were associated with patient experience. Neither job control nor colleague support was found to predict patient-rated satisfaction with quality of care being provided by British general practitioners (McKinstry et al., 2007). Similarly, job control, colleague support, and supervisor support did not correlate with patient-rated satisfaction with the support received from their oncologists in Germany (Ansmann et al., 2014).

Higher-order job resources and patient experience. Mache et al. (2012) measured higher-order job resources (comprising influence at work, degree of freedom of work, possibilities for development, quality of leadership, social support, feedback at work, social relations, and sense of community) among German surgeons. This positively correlated with patient satisfaction (*r*=.42).

4.4 Discussion

This systematic review and meta-analysis indicate that the relationship between doctors' psychosocial working conditions and quality of patient care provided is not as clear as expected, although it does indicate that complex differential effects exist. While most of the studies examined showed that aspects of job demands and resources predicted quality of care, the evidence suggests these pertain mainly to clinical excellence and patient safety, and not patient experience. It is important to acknowledge that these conclusions are drawn from only 21 fairly heterogeneous studies from a small number of developed countries, with most of the meta-analyses based on a few studies. Moreover, the reliance on cross-sectional convenience studies, along with the absence of theoretical considerations makes it implausible to conclude that any causal relationships exist.

4.4.1 Theoretical and methodological considerations

It is important here to recognise that the relationship between doctors' psychosocial working conditions and quality of care is likely influenced by a variety of complex and dynamic systems, and their interactions. The complexity of heterarchical organisational contexts and the nature of their inter-relationships mean other factors potentially affect this relationship, including: curvilinear effects, and moderating and mediating variables. For example, curvilinear properties were observed in the present studies for mental workload (Bertram et al., 1992) and autonomy (Stern et al., 2008), where increasingly high scores on either corresponds with an initial increase followed by a progressive decline in performance. The one study that did examine an interaction effect demonstrated that high job autonomy, when occurring in an environment which did not encourage learning, was actually associated with an increase in the number of treatment errors made (Stern et al., 2008). Similarly, other constructs within psychosocial research that are prevalent in the healthcare sector (e.g., job insecurity, role conflict) were not examined by the included studies in this review. This, in turn, limits our understanding of how all aspects of the perceived work environment potentially relate with quality-of-care indicators.

It is also plausible that working in environments with lower standards of care leads to doctors perceiving the environment as more demanding and less resourceful. This dynamic system operates in parallel to the complex system above, with measures representing psychosocial working conditions and quality of care reciprocally influencing one another. Longitudinal studies demonstrate support for the possibility of reverse causality where doctors' error rates predicted future levels of depressive symptoms (Shanafelt et al., 2010). Within this review, Krämer et al. (2016) did find that quality of care predicted time pressure, but not social conflict or emotional demands, one year later. Aside from this, none of the included studies considered a reverse or cyclical relationship. However, Dollarhide et al.'s (2013) findings that hospital physicians reported higher levels of task load prior to medical events occurring provides some evidence that job demands precede quality-of-care outcomes.

The absence of any significant consideration to these two factors could be attributed to the lack of theoretical consideration from the included studies, as only five studies utilised a theoretical framework (Bernburg et al., 2016; Krämer et al., 2016; Loerbroks et al., 2016; Shirom et al., 2006; Weigl et al., 2015). Where studies included mediators within this relationship, wellbeing measures such as burnout and depressive symptoms were used. In line with the JD-R model (Demerouti et al., 2001) and similar models in healthcare (Lowe & Chan, 2010; Montgomery et al., 2011), the results lend some support to the finding that job demands are associated with poorer wellbeing and in turn, with lower standards of patient care.

Theory is essential, not only in explaining how or why doctors' psychosocial working conditions influences quality of care, but also to account for confounding factors and the possibility of reverse causality, as well as to inform the design of interventions. Meta-analytical reviews indicate a large proportion of unaccounted variance within the psychosocial working conditions and performance relationship (Gilboa, Shirom, Fried, & Cooper, 2008), highlighting the necessity of considering behavioural, cognitive, motivational, and physiological mechanisms within this context. Consequently, as researchers strive to test and understand this relationship, it is imperative that appropriate theoretical frameworks be used to structure their investigations and explain their data. Without theory, there is a danger of oversimplifying our understanding of this relationship and of developing inadequate interventions.

4.4.2 The relationship between doctors' psychosocial working conditions and quality of patient care

Acknowledging the shortcomings above, the majority of studies did observe some relationships between doctors' psychosocial working conditions with clinical excellence and patient safety. This is in line with suggestions that overloaded doctors waste energy and time coping with their working conditions, diverting limited personal resources away from performance-related behaviours (Jex, 1998). In turn, this may lead to them ignoring important contextual cues and information. Furthermore, doctors struggling with emotionally demanding work or patients may be more prone to burnout, which subsequently, can reduce the quality of care provided (Hall et al., 2016; Weigl et al., 2015). It has also been suggested that stressors are linked with physiological responses that inadvertently disrupt performance (Lazarus, 1999; Spector, Dwyer, & Jex, 1988).

In terms of job resources, job control and autonomy can function to provide more opportunities to cope with challenging situations (Bakker & Demerouti, 2017; Shirom et al., 2006). Similarly, opportunities for learning and development allow doctors to become better trained and equipped to do their work, increasing their ability to provide better care. Providing learning and development opportunities also reinforces the value of the worker to the organisation and fosters meaning and purpose within workers; both of which are associated with better wellbeing at work (Panari, Guglielmi, Simbula, & Depolo, 2010; Schaufeli & Bakker, 2004). Social support is also useful as a source of information and emotional support (Deci & Ryan, 1985; Teoh, Coyne, Devonish, Leather, & Zarola, 2016), particularly from supervisors who are typically better placed to influence work patterns and access to resources. These findings reinforce the argument that interventions should encompass the strengthening of job resources in the workplace and not only focus on the reduction of job demands (Knight, Patterson, & Dawson, 2017; K. Nielsen et al., 2017).

The most consistent predictors of quality of care, with the largest effect sizes, were the measures of higher-order job demands and resources. This is not surprising considering these capture a wider and more comprehensive picture of what the work environment is like (van Vegchel, de Jonge, Bakker, & Schaufeli, 1999; Wellens & Smith, 2006). It has also been argued that the specificity of an outcome should match that of the predictor (Ironson, Smith, Brannick, Gibson, & Paul, 1989), meaning a narrower and more specific measure of job demands (or resources) would require an equivalent measure of quality of care to demonstrate an effect. This suggests that quality-of-care initiatives that target specific job demands or resources may fail to address the underlying problems within the system or may only yield improvements on specific outcomes. Therefore, any changes to the working conditions of doctors should consider how they influence the job demands and resources perceived by doctors. However, few studies (e.g., Benning et al., 2011) have evaluated workplace-based psychosocial interventions in healthcare, highlighting the need to complement the growing literature on interventions targeting the individual.

4.4.3 Does the type of outcome measures matter?

Surprisingly, none of the studies sought to examine hard data on clinical and health outcomes, which would have provided clearer practical significance (Reed et al., 2007). Instead, studies used behavioural or attitudinal outcome measures that were self, observer or patient-rated. These in turn were separated into those representing perceived clinical excellence, patient safety, and patient experience. The lack of relationships involving patient experience in this review is congruent with the inconsistency of patient experience in the wider research literature, where some studies observe it to relate with staff wellbeing and psychosocial working conditions (Powell et al., 2014; Salisbury et al., 2010), while others fail to do so (Ratanawongsa et al., 2008). These results lend themselves towards a separate argument about the utility of measuring patient experience measures, with issues including inconsistency in conceptualising what this represents, and its poor links with other forms of quality measures (Crow et al., 2002; Salisbury et al., 2010).

However, sufficient evidence elsewhere demonstrates the validity of some aspects of patient experience as a proxy of care quality (Powell et al., 2014; Salisbury et al., 2010), but this review highlights that its relationship with doctors' psychosocial working conditions is not clear cut. It is plausible that relationships involving patient experience are more complex than those involving clinical excellence or error outcomes. For example, patient satisfaction scores arguably capture the patient's attitudes and expectations about the service received (Crow et al., 2002). Patients accustomed to poor practice over time may perceive this as standard practice, thereby blunting the usefulness of such a measure (McKinstry et al., 2007). The absence of a relationship involving patient demands could also be attributable to emotional labour (Hochschild, 1983), a process where doctors are expected to express and regulate desired emotions during patient interactions. Similarly, it is possible that doctors' professional standards mean they are aware of their limitations and attempt to overcompensate in their delivery (Ratanawongsa et al., 2008) to still deliver, or appear to deliver, appropriate levels of care. These factors that predict patient experience need to be better understood, because if doctors are overexerting themselves to maintain adequate levels of care, this has serious ramifications for their long term wellbeing (Mann, 2005).

4.4.4 Limitations of current review

Within this review, doctors were treated as a homogenous group. The reality is that doctors are part of a heterogeneous profession. The studies included here involved doctors from various specialties and levels. This heterogeneity is compounded by the representation of multiple countries that operate different health systems. These have implications for the nature of the work being conducted and the types of working conditions that doctors are exposed to. Moreover, how quality of care is perceived across different specialties and nations may also confound the relationship. The effect sizes included in the meta-analysis came from a small number of studies published in the English language and did not account for study quality or publication bias. Not all studies reported *r* values, and, where possible, standardised regression values were used instead. While these are suitable replacements in meta-analysis (Borenstein et al., 2009), the consistent use of *r* values would strengthen validity of the findings. In addition, correlational analysis fails to account for cluster effects of doctors' groupings in department or hospitals (Ansmann et al., 2014). Finally, any conclusions drawn must consider the limitations of the included studies that are covered above (e.g., the reliance on cross-sectional data, the absence of clinical outcome measures, and lack of theory).

4.5 Conclusion

4.5.1 Summary of the systematic review

The consistent observation in this review was that higher-order measures of job demands and resources have a relationship with quality of care. This suggests that interventions need to target a range of psychosocial factors; focusing on specific demands or resources may fail to address the underlying problems within the system, or may only yield improvements on specific outcomes (Benning et al., 2011; Weigl et al., 2013). Any changes to the working conditions of doctors (e.g., organisational restructuring, contract negotiations) should consider how it potentially influences the demands and resources perceived by doctors, as collectively these have some relation with poorer quality of care. From a research perspective, future studies should utilise theoretical frameworks to explain and structure their investigations. The lack of consistency in observing effects of specific job demands and resources can be attributed to methodological issues of measurement, particularly on patient-rated measures. The absence of theoretical frameworks makes it difficult to fully understand and explain these relationships. However, it is likely that focusing on addressing higher-order job demands and resources, rather than specific aspects, will lead to the best improvements for quality of care.

4.5.2 Implications for thesis

The review highlights two key implications relevant to this thesis. First, that there still is ambiguity towards a relationship between psychosocial working conditions of doctors and quality of care. From a design perspective, consideration should be given to examining clinical outcomes alongside attitudinal and affect outcomes. Second, and perhaps more importantly, is the need to situate these relationships within a theoretical framework. As such, the next chapter introduces the job demands-resources model (Bakker & Demerouti, 2007; Demerouti et al., 2001), which provides the theoretical background for this thesis.

Chapter 5 : Introducing the Job Demands-Resources Model

The preceding chapters identified the lack of theory explaining the relationship between doctors' psychosocial working conditions and the quality of care being provided. Consequently, this chapter serves to introduce the job demands-resources (JD-R; Bakker & Demerouti, 2007; Demerouti et al., 2001) model. It first provides a brief review of the different approaches to stress theories, before situating the JD-R model within it. The chapter then introduces the model's dual processes that predict motivation (i.e., positive wellbeing) and strain (i.e., negative wellbeing). This is followed by exploring how job resources can buffer the negative impact of job demands, as well as interacting with high level of job demands to predict positive wellbeing. The chapter proceeds to explore how the JD-R model explains performance, which is followed by considering the model from a multilevel perspective. Finally, some of the key limitations of the JD-R model are reviewed before the chapter ends with applying the model to the healthcare sector.

5.1 A Brief Introduction into the Theories of Stress

Theories of work-related wellbeing lie primarily within the stress literature. The broad consensuses of reviews into the historical progression of stress theories explain that these theories can be divided into three different approaches (Cox, 1978; Cox & Griffiths, 1995; Cox et al., 2000; Lazarus, 1966). The first is the engineering approach, which conceives stress as a threating, noxious, or aversive element of the work environment (Cox, 1978). Here, stress is seen as the level of demand placed on an individual. In contrast, the second approach, known as the physiological approach, sees stress as the response to the environment (Cox, 1993; Selye, 1956). Therefore, stress is the psychophysiological changes that occur within the individual. Finally, the psychological approach focuses on the interaction between the individual and their work environment (Cox, 1993; Cox & Griffiths, 1995; Cox et al., 2000). It focuses on the emotional reactions and cognitive processes that stem from any mismatch that happens here. More specifically, the psychological approach can be further separated into interactional and transactional theories. The former encompass the individual's interactions with the structures of their work environment. The latter refer to the cognitive appraisal and coping that explain the perception of the work environment (Dewe & Trenberth, 2004).

It is beyond the scope of this chapter to review in-depth the various stress theories; the interested reader is directed to resources elsewhere (Cox & Griffiths, 1995, 2010; Cox et al., 2000;

Häusser et al., 2010). Psychological theories of stress, and in particular interactional theories, have dominated the research literature. This is partially due to their simplicity in understanding their models and the availability of standardised assessment tools. As such, an extensive literature base has developed around some of the main interactional theories (e.g., Häusser et al., 2010; Kivimäki et al., 2006; Luchman & González-Morales, 2013; Van der Doef & Maes, 1999), namely the job-demand-control-support model (Johnson & Hall, 1988) and the effort-reward imbalance model (Siegrist, 1996).

For the purpose of this thesis, however, it was decided that a more contemporary interactional theory – the job demands-resources model (Bakker & Demerouti, 2017; Demerouti et al., 2001) was better suited. The reason for utilising the JD-R model in this thesis is three-fold. First, it addresses the shortcomings found in other psychological theories that use narrow or selective definitions of the psychosocial work environment. For example, the job-demand-control support model only focuses on three aspects of the psychosocial work environment – job demands, job control, and support (Johnson & Hall, 1988; Karasek & Theorell, 1990). This is particularly limiting considering the range of job demands and resources identified in the systematic review (Chapter Four). Second, the design of the JD-R model inherently matches with the conceptual relationships examined in the healthcare literature, whereby work-related wellbeing mediates the psychosocial working conditions and performance relationship. Finally, and perhaps most importantly, the JD-R model does not identify as a stress theory nor does it focus solely on negative wellbeing (Bakker, Demerouti, & Sanz-Vergel, 2014; Demerouti & Bakker, 2011). Instead, it embraces a positive psychology approach that psychosocial working conditions can have a positive impact on work-related wellbeing, allowing it to account for positive constructs such as job satisfaction and work engagement within the model.

5.2 The Job Demands-Resources Model

The job demands-resources (JD-R) model was first published in 2001, and has over the last 15 years been revised, extended, and updated as new developments and research emerged (Bakker & Demerouti, 2007, 2017; Demerouti et al., 2001). The underlying principle is straightforward – that almost all aspects of the psychosocial work environment can be classed either as a demand or a resource to the worker. Job demands, refers to any social, organisational, physical, or psychological aspects of work that are associated with psychological and/or physiological costs due to the sustained effort (Demerouti et al., 2001). These include heavy workloads, bullying, role conflict, and emotionally demanding work. In contrast, job resources are those social, organisational, physical, or psychological aspects of work that help (i) reduce job demands; (ii) achieve work goals; and/or, (iii) stimulate personal learning and development. Social support, autonomy, and performance feedback are examples of job resources. One of the key strengths of this theory is its simplicity, whereby these demands and resources can vary across sectors, organisations, and even workers.

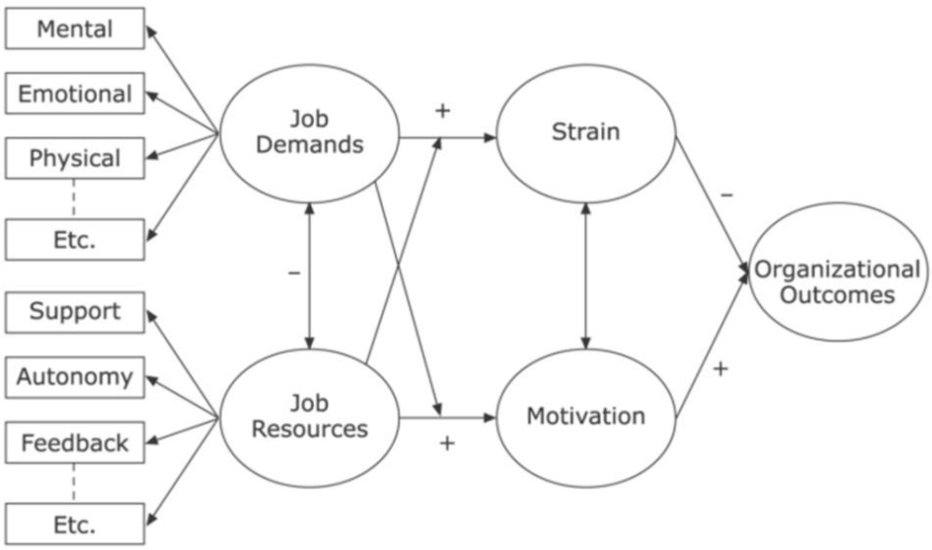

Figure 5.1:The job demands-resources model (Bakker & Demerouti, 2007)

5.2.1 Dual processes

The theory postulates two independent processes (Demerouti et al., 2001) that explain the relationship between psychosocial working conditions (i.e., job demands and resources) and wellbeing (i.e., strain and motivation; See Figure 5.1). While the JD-R model uses the terms strain and motivation, in keeping consistent with the terminology throughout this thesis these are respectively referred to as negative and positive work-related wellbeing.

The *health-impairment process* suggests that demands uniquely predict negative work-related wellbeing, which could manifest as burnout, work-related stress, and health exhaustion amongst other negative wellbeing measures. Here, job demands arouse a stress process that leads to energy depletion (van Emmerik, Bakker, & Euwema, 2009). This requires greater effort that comes at a higher psychological and physical cost to the individual. Indirect degradation may occur here in the form of strategy adjustments (narrowing of attention, increased selectivity, redefinition of task requirements) and fatigue after-effects (risky choices, high levels of subjective fatigue; Hockey, 1993). These compensatory strategies over long periods of time drain an individual's energy, eventually resulting in a breakdown. As such, job demands are not always inherently negative, and may only emerge as a job stressor when a demand necessitates sustained effort that the worker cannot adequately recover from (Meijman & Mulder, 1998; Schaufeli & Bakker, 2004).

In comparison, the *motivational process* focuses on job resources, which uniquely explain positive states within the individual (Demerouti et al., 2001; Schaufeli & Bakker, 2004). In the context of the JD-R model, motivation is an umbrella term representing positive wellbeing, and is typically measured as work engagement, motivation, and job satisfaction. Extrinsically, job resources serve to support the reaching of one's goals; intrinsically, they foster growth, development, and learning. When faced with high demands, an individual could reduce their level of motivation and work engagement with their job to protect their own performance and/or wellbeing (Bakker, Demerouti, de Boer, & Schaufeli, 2003). Job resources also tap into the desire to fulfil basic human needs, such as autonomy, competence, and relatedness, which according to self-determination theory, increases intrinsic motivation and enhances wellbeing (Deci & Ryan, 1985). Having access to more job resources can improve the extent to which the worker feels efficacious, as those with a wider array of resources have more opportunities to learn new behaviours than those whose resources are lacking (Bakker, Hakanen, Demerouti, & Xanthopoulou, 2007; Schaufeli, Bakker, & Van Rhenen, 2009). Therefore, in addition to mitigating the effect of job demands, job resources are essential in their own right.

Testing the dual process proposition. The results from the original JD-R article, set within three different occupational groups, support these two independent processes (Demerouti et al., 2001). Here, job demands (physical workload, time pressure, recipient contact, physical environment, and shift work) uniquely predicted exhaustion, while (dis)engagement was uniquely predicted by job resources (feedback, rewards, job control, participation, job security, supervisor support). Similarly, in a longitudinal three-year study of 2,555 Finnish dentists, Hakanen, Schaufeli and Ahola (2008) found that job demands predicted future depression through burnout; job resources, on the other hand, predicted work engagement and subsequent organisational commitment. Further support for the dual processes is seen in a study of Dutch production workers where workload and reorganisation (i.e., job demands) were the only predictors of absence duration, via burnout (Bakker, Demerouti, de Boer, et al., 2003). Here, the motivational process manifested as job control and participation (i.e., job resources) solely predicting commitment, and in turn, absence frequency.

The JD-R model emphasises the independence of the dual processes, and that evidence of "cross-paths are largely due to suboptimal research designs" (pg. 5; Bakker & Demerouti, 2017). Unfortunately, the evidence does not completely support these statements. For example, the longitudinal study with Finnish dentists also found a weak correlation between job demands and work engagement, and between job resources and burnout (Hakanen, Schaufeli, & Ahola, 2008). Perhaps more importantly, in a meta-analysis of 203 safety-related studies (Nahrgang et al., 2011), job demands (i.e., complexity , risks and hazards) and job resources (i.e., autonomy, knowledge, social support, leadership, safety climate) both predicted work engagement and burnout. A separate meta-analysis found that distinguishing between two types of job demands found work engagement positively predicted by challenge demands and negatively predicted by hindrance demands (LePine, Podsakoff, & LePine, 2005). These findings are congruent with other studies that explicitly test the JD-R model's dual process proposition but find evidence of cross-paths, where job resources also predicted ill-health (Bakker, Demerouti, & Schaufeli, 2003), burnout (Hakanen, Bakker, & Schaufeli, 2006; Schaufeli & Bakker, 2004), and exhaustion (Bakker, Demerouti, & Verbeke, 2004); or that job demands also predicted work engagement (Bakker et al., 2007; Hakanen, Bakker, & Demerouti, 2005). What this suggests is there is strong evidence for separate health-impairment and motivational processes, with job demands the strongest predictor of burnout and job resources consistently the strongest predictor of engagement. However, these processes may not be completely independent and some cross-paths can still occur.

5.2.2 Job resources as a buffer for job demands

As seen in Figure 5.1, the JD-R model proposes that job resources buffer the detrimental effect that demands have on negative wellbeing (Bakker & Demerouti, 2007; Bakker, Demerouti, & Euwema, 2005). This could be achieved by altering the cognitions and perceptions stemming from such stressors; moderating responses that follow the appraisal process; and/or, reducing the health-damaging outcome of such responses (Kahn & Byosiere, 1992). One of the first studies to propose and test this interaction surveyed 1,102 higher education employees (Bakker et al., 2005). Here, 18 out of 32 possible interactions between four job demands (overload, emotional demands, physical demands, home-work interface) and job resources (social support, supervisor relationship, feedback, autonomy) demonstrated that job resources mitigated the negative effect of job demands on exhaustion and cynicism. However, no significant interactions were observed with professional efficacy as the outcome. When this was examined among 230 medical residents (Bakker, ten Brummelhuis, Prins, & van der Heijden, 2011), nearly all hypothesised interactions found high job demands (workload, emotional demands, cognitive demands) and high resources (development opportunities, feedback, supervisory coaching, participation) to relate with lower work-home interference (8 of 12 interactions). However no significant interactions were observed involving autonomy. Altogether, these findings suggest that workers with more job resources are better equipped to cope with high job demands.

Further evidence is seen in a study with home care workers (Xanthopoulou, Bakker, Dollard, et al., 2007), where 66% of the proposed relationships indicated that job resources buffered the negative relationships between job demands and the burnout dimensions. In particular, job resources functioned better as buffers when emotional demands or patient harassment were predictors than when workload or physical demands were predictors. This is attributed to job resources (autonomy, development opportunities, social support, performance feedback) being better suited to mitigate the effect of emotional demands and patient harassment which may be more salient in this environment. In comparison, home care workers having to frequently work on their own means they have little opportunity to address workload or physical demands. This is congruent with the notion that whether job demands and resources play a role depends upon the specific job characteristics that prevail within that organisational context (Bakker & Demerouti, 2007).

Although various job resources could buffer the effect of job demands (van Emmerik et al., 2009), consistent significant interactions can be facilitated when demands and resources match. Moreover, individual differences could also make some job resources’ capacity to mitigate job demands more salient than others (Bakker et al., 2005). It is also plausible that when demands are too high, workers’ ability to fully utilise their available resources is restricted. For example, in one study of home care workers, high job resources levels had a limited effect on alleviating the impact of job demands and instead was strongest when employees had few job demands (Bakker, Demerouti, Taris, Schaufeli, & Schreurs, 2003). Collectively, these findings could explain why not all examples of job resources were found to mitigate the effect of job demands, including those studies where no significant interactions were found at all (Bakker et al., 2004).

5.2.3 High job demands and resources predict positive wellbeing

The JD-R model further postulates that high job demands amplifies the influence that job resources have on positive wellbeing (Bakker & Demerouti, 2014; Hakanen et al., 2005). The explanation for resources' saliency when paired with high demands lies in individuals' propensity to obtain, retain, and protect whatever they value; this could be resources in energetic, material, personal, and social form. Drawing on the conservation of resources theory (Hobfoll, 1989, 2002), it explains that resources beget additional resources. It further argues that when threatened with the possibility of a loss of resources (i.e., through increased demands), then resources provide additional motivational propensity to act (Hobfoll & Shirom, 2001). Consequently, when individuals face low job demands then job resources may be of less relevance or concern (Bakker et al., 2007); however, as job demands increase so too does the importance of job resources.

Examining this hypothesis in a sample of 12,359 employees across 148 Dutch organisations (Bakker, van Veldhoven, & Xanthopoulou, 2010), nearly all predicted interactions between job demands (workload, emotional demands) and resources were found significant in relation to organisational commitment (13 of 16 interactions) and task enjoyment (15 of 16). Job resources here were represented by autonomy, career opportunities, colleague support, leader support, learning opportunities, participation, performance feedback, and skill utilisation. Essentially, the results indicate high demands improved the saliency of resources, allowing engagement to flourish. Bakker and colleagues' (2007) observed similar findings where 14 out of 18 possible interaction effects between pupil misbehaviour and six different job resources (job control, supervisor support, climate, innovativeness, information, and appreciation) were significant in relation to three work engagement outcomes.

5.2.4 The JD-R model and performance

The validity and utility of the JD-R model lies not only in how job demands and resources explain work-related wellbeing, but how these collectively predict performance (Bakker & Demerouti, 2007). Through the health-impairment process, workers who are exhausted or suffering from ill-health lack the capacity or resources to achieve their work goals (Bakker & Demerouti, 2017). Similarly, the motivational process allows workers to tap into energy and enthusiasm, as well as facilitating goal-orientated behaviour. As such positive wellbeing is associated with higher performance, while the opposite is observed with negative wellbeing.

The wellbeing-performance relationship is a well-trodden path in the research literature. This is evident in a meta-analysis of 16 studies (Taris, 2006), where the burnout component of emotional exhaustion related with in-role behaviour (r=-.22), organizational citizenship behaviour (r=-.19), and customer satisfaction (r=-.55). Other systematic reviews in this topic that support the relationship between wellbeing and performance include psychological ill-health and sickness absence (Michie & Williams, 2003); staff engagement and motivation with safety performance (Christian, Bradley, Wallace, & Burke, 2009; Nahrgang et al., 2011); healthcare worker wellbeing and patient care (Hall et al., 2016; Scheepers et al., 2015); and job satisfaction (Judge, Thoresen, Bono, & Patton, 2001) with work-related performance.

Focusing specifically on the dual processes within the JD-R model, Bakker, Demerouti and Verbeke (2004) found job demands predicted exhaustion and subsequent in-role performance, while job resources predicted extra-role performance via disengagement. Supporting the motivational process, a five-day diary study of 42 fast food workers demonstrated that job resources had an effect on personal resources, which in turn influenced work engagement levels and subsequent financial performance (Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2009). Similarly, a longitudinal study observed teachers' autonomy, supervisor support, and developmental opportunities (but not social support), to positively relate to weekly engagement (Bakker & Bal, 2010); which, in turn, was positively related to weekly job performance. Moreover, momentary work engagement was positively related to job resources in the subsequent week.

5.2.5 A multilevel perspective of the JD-R model

Initially conceptualised at the individual-level, there have been questions as to the suitability of the JD-R model at more micro (i.e., within the individual) and macro (i.e., team, organisation) levels (Bakker & Demerouti, 2017; Schaufeli & Taris, 2014). The latter is particularly important given the nested nature of organisational research where employees inherently function in groups; as well as the presence of constructs and outcomes of interest that lie across different hierarchical levels. Where variables operate at different levels, for example job demands as perceived by an individual and performance measured at the team level, researchers traditionally either: (i) aggregate all the individuals' responses within a team to create one team score; or, (ii) use the same team performance score for each member of that specific team (Heck & Thomas, 2015). This violates the compatibility principle which requires all variables within a model to operate at the same level of specificity (Ajzen, 2005).

This is important as one cannot assume that the shared perception of a particular construct at the individual level maintains the same meaning at the team level (Demerouti & Bakker, 2011). This also has implications for the proposed JD-R relationships: would team engagement still mediate the relationship between job resources as postulated by the motivation process? This is vital not only for the purpose of theory building, but practically for organisations to understand whether differences or similarities of constructs (e.g., performance) across levels may require different strategies of managing performance at the individual, team, and organisational levels (Demerouti & Bakker, 2011).

Studies examining the JD-R model from a group or organisational perspective often violate the compatibility principle. This includes studies where individual-level job demands and resources were aggregated to a higher-level to relate with team-financial performance (Xanthopoulou et al., 2009), supervisor-rated team performance (Bakker, van Emmerik, & Van Riet, 2008), and unit-level performance (Harter, Schmidt, & Hayes, 2002). For example, the study that first proposed the JD-R model (Demerouti et al., 2001) aggregated individuals' disengagement and burnout scores to the level of job positions. Although structural equation modelling revealed similar results to the individual-level model, the authors caution against generalisation given that the 21 job positions constituted a very small sample size. Similarly, Salanova et al. (2005) aggregated individual-level training, autonomy, and technology scores to represent organisational-level resources, which in turn was found to predict organisational-level employee self-rated performance through organisational-level engagement and service climate.

Aggregating responses to a team-level does not necessarily mean the team-level version operates in the same manner as its individual-level counterpart. In a longitudinal study of agency workers, unit-level cohesion and social support exacerbated the crossover of perceived job demands from the team to the individual (Westman, Bakker, Roziner, & Sonnentag, 2011). This implies that these unit-level job resources can encourage the spread of burnout within teams and creates an argument that what may be protective at the individual-level may be harmful at the team-level. Another study of 93 nursing departments across seven European countries observed no relationship between department-level teamwork effectiveness and burnout (Montgomery, Spânu, Băban, & Panagopoulou, 2015), although teamwork effectiveness did relate with engagement. Therefore, little is still known about how individual-level constructs function across different levels.

Despite these shortcomings, there is some indication supporting the validity of the JD-R model from a multilevel perspective. In a study involving police officers, both individual-level engagement and burnout were related to their team-level versions (Bakker, van Emmerik, & Euwema, 2006). This implies that not only do team-level constructs maintain a similar meaning to their individual-level counterparts, but that it may be possible for individual-level engagement and burnout to be transferred across different team members. Torrente and colleagues (2012) accounted for the compatibility principle by using the term "my team" instead of "I" on items to provide a referent shift from the individual to the team. Their findings are congruent with the JD-R model, suggesting that team-level social resources positively related to team engagement, which in turn, positively related to supervisor-rated team performance. Similarly, congruence between individual and team-level constructs is evident in a study from Spain (González-Morales, Peiró, Rodríguez, & Bliese, 2012), where teachers' school-level shared perception of burnout was a stronger predictor of burnout one-year later than job demands, resources, and individual-level burnout. Other studies have also shown team-level teamwork (Busch, Deci, & Laackmann, 2013) and social support (Li et al., 2013) to have the same effect on outcomes as their individual-level counterparts.

5.3 Limitations of the JD-R Model

Despite the evolvement of the JD-R model, it is not without its limitations and critique. The lack of consistency in research supporting separate independent processes (i.e., the health-impairment and motivation process), the inconsistent findings of interactions between job demands and resources, and the lack of multilevel understanding of the JD-R have already been reviewed above. In addition, first and foremost, the JD-R model is an open and heuristic model which does not have specific or well-defined job demands, resources, and outcomes (Schaufeli & Taris, 2014). While this flexibility is often viewed as a strength, it has also been criticised as it creates difficulty in generalising the findings from one context with a set of job demands and resources to a different context. In the same way, the JD-R model provides a descriptive framework of how the different constructs are hypothesised to relate to each other. The model does not provide a psychological explanation on why this occurs, and instead relies on other psychological frameworks such as the conservation of resources theory (Hobfoll, 1989; Hobfoll & Shirom, 2001), self-determination theory (Deci & Ryan, 1985), social cognitive theory (Bandura, 2010), and job characteristics theory (Hackman & Oldham, 1980). Therefore, although the JD-R model does well to explain what happens, it is not able to explain why this is the case.

Second, much of the research into the JD-R model has been cross-sectional and/or piecemeal. The growing complexity of the model means it is difficult to capture the entire model in one study. Instead, studies break down the JD-R model into different components, choosing to focus on the dual processes (Demerouti et al., 2001; Lewig, Xanthopoulou, Bakker, Dollard, & Metzer, 2007); the role of job resources to mitigate the negative effect of demands (Bakker et al., 2005; Xanthopoulou, Bakker, Dollard, et al., 2007) or foster work engagement (Bakker, van Veldhoven, et al., 2010); or other related aspects such as job-crafting and personal resources (Petrou, Demerouti, Peeters, Schaufeli, & Hetland, 2012; Vogt, Hakanen, Brauchli, Jenny, & Bauer, 2016). The dominance of cross-sectional studies echoes the organisational research sphere, although there is a slow but growing number of a longitudinal examinations of the JD-R model (Bakker & Bal, 2010; Hakanen et al., 2008; Prieto, Soria, Martínez, & Schaufeli, 2008; Schaufeli, Bakker, & Van Rhenen, 2009). More robust and comprehensive examinations of the JD-R model are therefore needed.

The third limitation centres on the meaning of job demands and resources. There is still uncertainty as to whether these are two separate constructs, or two sides of the same coin (Bakker & Demerouti, 2007; Schaufeli & Taris, 2014). If a job demand is defined as any organisational, social, physical, or psychological aspect of work which requires effort (Demerouti et al., 2001), then if low job resources created effort for the worker would job resources not then become a demand? Furthermore, there is increasing evidence that job demands are not one homogenous construct but can be further separated into challenge and hindrance demands (Crawford, LePine, & Rich, 2010). Meta-analyses reveal that the former is positively related with work engagement while the latter is negatively associated with work engagement (Crawford et al., 2010; LePine et al., 2005). This throws another dimension into the dual process hypothesis of the JD-R model, and may explain the mixed findings between job demands and work engagement within the literature.

5.4 The JD-R Model in Healthcare

One of the criticisms (and strengths) of the JD-R model lies in the range of different job demands and resources that have been used to test this model (Bakker & Demerouti, 2007; Schaufeli & Taris, 2014). Therefore, what constitutes a job demand or resource in one sector may not necessarily have the same impact in another. For example, high job demands and low resources were observed to relate with burnout in a sample of Chinese healthcare workers, but not with blue collar workers (Q. Hu, Schaufeli, & Taris, 2011). This was attributed to that combination being more detrimental amongst healthcare professionals who may be more intrinsically motivated and dedicated to their work than extrinsically driven blue collar workers. Given that this thesis focuses on doctors, a review of the application of the JD-R model within the healthcare context will help to understand its sector validity.

To the knowledge of this student, there have only been two studies to date applying the JD-R framework to doctors (Bakker et al., 2011; Zis, Anagnostopoulos, & Sykioti, 2014). In the first, it was only in eight out of 15 relationships where job resources buffered the negative impact job demands had on the work-home interference levels of Dutch medical residents. Here, job demands were represented by work overload, emotional demands, and cognitive demands; while, job resources were measured with job autonomy, learning and developing, performance feedback, supervisory coaching, and employee participation. Surprisingly, all interactions involving job autonomy were not significant. This is possibly due to residents having a structured training programme where other resources such as support and feedback may be more salient (Prins, Hoekstra-Weebers, et al., 2007). The second study does not explicitly test any of the JD-R model's propositions (Zis et al., 2014). However, their finding that both job demands (emotional demands, intellectual demands, workload, and home-work demands' interface) and resources (autonomy, opportunities for professional development, support from colleagues, and supervisor support) predicted burnout in Greek medical residents is not congruent with the independent health-impairment and motivational processes.

The majority of JD-R studies in healthcare focus on nursing (Bakker & Sanz-Vergel, 2013; Demerouti, Le Blanc, Bakker, Schaufeli, & Hox, 2009; Hansen, Sverke, & Näswall, 2009; Jourdain & Chênevert, 2010; Laschinger, Grau, Ashley, Finegan, & Wilk, 2012; Montgomery et al., 2015; Naruse et al., 2012), although some work has involved dentists (Hakanen et al., 2005, 2008) and mixed-groups of healthcare professionals (Q. Hu et al., 2011; M. C. W. Peeters & Le Blanc, 2001). Only two studies (Jourdain & Chênevert, 2010; Laschinger et al., 2012) supported for the JD-R model's dual process proposition, indicating that job demands predicted burnout while job resources predicted work engagement. However, while the remaining studies demonstrate the presence of cross-effects (Bakker & Sanz-Vergel, 2013; Hakanen et al., 2005; Montgomery et al., 2015; M. C. W. Peeters & Le Blanc, 2001), some indicate that even in the presence of cross-effects, burnout is better predicted by job demands than resources, with the opposite observed for work engagement (Hakanen et al., 2008; Hansen et al., 2009; Q. Hu et al., 2011; Naruse et al., 2012).

Frequently job demands and resources are represented as a latent construct comprising of specific demands and resources. When this is not the case, job demands such as workload, emotional exhaustion, time pressure, and poor-work life balance were consistently among the stronger predictors of outcomes (Bakker et al., 2011; Hakanen et al., 2005; Hansen et al., 2009; Jourdain & Chênevert, 2010; Laschinger et al., 2012; Montgomery et al., 2015; Naruse et al., 2012; Zis et al., 2014). Some studies included psychosocial aspects of work specific to the healthcare sector, for example patient interactions and demands (Demerouti et al., 2009; Hakanen et al., 2005), hostility and aggression (Jourdain & Chênevert, 2010; Laschinger et al., 2012), and being on-call (Naruse et al., 2012). Healthcare-sector related variables such as the number of working hours (Bakker & Sanz-Vergel, 2013) predicted outcomes, although other related variables, including tenure (Bakker & Sanz-Vergel, 2013; Hansen et al., 2009), adherence to the European Working Time Directive (Zis et al., 2014), and specialty-type (Zis et al., 2014) did not.

Congruent with the health-impairment process, three studies (Q. Hu et al., 2011; Jourdain & Chênevert, 2010; Laschinger et al., 2012) found negative wellbeing to mediate the job demands and organisational outcomes (i.e., turnover intention, commitment) relationship, while work engagement mediated job resources' relationship with organisational outcomes. In Laschinger and colleagues' study, emotional exhaustion also mediated the demands and mental health relationship. However, two of these studies (Q. Hu et al., 2011; Laschinger et al., 2012) comprised of partial mediations as demands and resources also had direct relationships with outcomes measured. A fourth study (Demerouti et al., 2009) did not find burnout components to contain mediating properties. Instead job demands predicted burnout at two subsequent time points over eighteen months, although emotional exhaustion only predicted presenteeism between Time 1 and Time 2. Interestingly, Demerouti et al. observed a reciprocal relationship between burnout and presenteeism, suggesting that lack of sufficient recuperation is related with subsequent poor health.

Noticeably lacking from the healthcare sector are studies aiming to validate the JD-R model from a multilevel perspective or by including a measure of quality of care as an outcome. Considering the strong presence of clusters within this sector: wards, departments, hospitals, and professional groups, it is surprising that only one study has considered unit-level variables within the JD-R model. However, Montgomery et al. (2015) did not find department-level teamwork effectiveness in 93 nursing departments to relate with burnout. Equally surprising, considering the main purpose of healthcare systems is to provide safe and good quality of care, none of the studies reviewed in this section use outcome measures which serve as a proxy of care. Instead, where performance outcomes are measured as a consequence of burnout or engagement, more generic organisational outcomes such as turnover intention and commitment are used (Jourdain & Chênevert, 2010; Laschinger et al., 2012).

5.5 Conclusion

The JD-R model provides a useful theoretical framework to test the relationships between psychosocial working conditions of doctors and quality of care. The growing popularity of this model has resulted in increasing support for the main relationships proposed by the model. However, gaps remain in developing a healthcare sector specific understanding of the validity of the JD-R model. The lack of testing for interaction effects between job demands and resources, in addition to the absence of studies focusing on doctors as a profession, provides the rationale for this thesis to fill in these gaps. Equally important is the need to examine whether performance metrics more congruent to the healthcare sector, namely quality-of-care outcomes, are suited to the JD-R model. Moreover, the clustering of doctors within trusts and the regular recording of trust-level outcome data means that the healthcare sector is well positioned to respond to calls for more research to test the JD-R from a multilevel perspective (Bakker & Demerouti, 2017; Demerouti & Bakker, 2011; Schaufeli & Taris, 2014).

As introduced in Chapter One, the second aim of this thesis is to examine the relationship between doctors' psychosocial working conditions and quality of care within the JD-R framework. Therefore, drawing upon the propositions reviewed in this chapter allows for more specific questions to be asked:

i. Do hospital doctors' job demands uniquely predict negative work-related wellbeing; and do job resources uniquely predict positive work-related wellbeing?
ii. Will hospital doctors' job resources moderate the relationship between job demands and negative work-related wellbeing?
iii. Will hospital doctors' job demands moderate the relationship between job resources and positive work-related wellbeing?
iv. Does work-related wellbeing mediate the relationship between hospital doctors' psychosocial working conditions and quality of care provided?
v. Will trust-level demands have the same impact within the JD-R model as that of hospital doctors' job demands?
vi. Will hospital doctors' psychosocial working conditions and work-related wellbeing predict trust-level quality-of-care outcomes?

These form the basis of the subsequent quantitative studies in this thesis, which utilise a secondary data analysis approach. In order to test these assumptions the appropriate datasets are required. Therefore, the subsequent chapters serve to introduce what secondary data is, and establish the validity of the measures that will inform the studies that seek to answer the questions posed above.

Chapter 6 : Secondary Data Analysis and the NHS Staff Survey

This chapter aims to provide an introduction to secondary data analysis and the NHS Staff Survey. It first explains why a secondary data analysis approach was chosen for this thesis. It includes reviewing the strengths and limitations of the NHS Staff Survey and the relevant ethical concerns. The chapter then describes the 2014 version of this dataset. This is followed by a discussion on the utility in creating composite measures, including the importance of validity and reliability in measurement.

6.1 Using Secondary Data in Research

Research surveys of doctors' working conditions and wellbeing typically yield very low return rates (Lee et al., 2013). This is not only due to the challenging working conditions of hospital doctors, but the regularity with which doctors are surveyed about various aspects of their work and patient care. These include annual national surveys such as the NHS Staff Survey (Picker Institute Europe, 2015), the National Training Survey for Doctors-in-Training (General Medical Council, 2016), and the National Consultant Census (Federation of the Royal Colleges of Physicians of the UK, 2016). In addition, doctors are regularly surveyed by their respective Royal Colleges and as part of internal organisational initiatives (Kumpunen et al., 2015; Royal College of Physicians, 2015a; Royal College of Surgeons, 2010). Primary data collection with hospital doctors would have involved an extensive process of obtaining buy-in from NHS trusts or the Royal Colleges. This would likely also have entailed having to obtain NHS ethical approval to collect data within the NHS which can be a long and cumbersome process.

The core surveys involving hospital doctors were reviewed with items preliminarily mapped against popular psychosocial working conditions and wellbeing indicators. The NHS Staff Survey was observed to collect data on working conditions, work-related wellbeing, and quality of care, along with other measures on an annual basis. With the 2014 dataset involving 17,670 doctors from every NHS trust in England, it is unlikely that a larger or more representative sample would have been collected that fit the aims of this thesis. An added advantage of the NHS Staff Survey is its grouping of employees within trusts in England. This facilitates multilevel examination and opens the possibility of linking the dataset with other quality-of-care outcomes collected at the trust-level (Koziol & Athur, 2011; Teoh, 2016). The annual NHS Staff Survey along with regular collection of quality-of-care data also allows the possibility of longitudinal examinations into the relationships between psychosocial working conditions and quality of care. Such a longitudinal analysis was, however, beyond the scope of this thesis.

Despite the advantages of secondary data analysis, there are limitations that need to be acknowledged. Firstly, the original purpose of the data collected differs from the objectives of this study. Consequently, the type of data collected might not be completely congruent with the purpose of this study (Koziol & Athur, 2011). Secondly, lack of clarity surrounding measures used creates concern about its validity and reliability; and although the constructs being examined might be similar to that of this study, the measures used are frequently single/low-item measures whose psychometric properties are not clear (Hair, Black, Babin, & Anderson, 2014). This is typically a result of surveys that emphasise breadth over depth of constructs (Koziol & Athur, 2011). The large sample size from the dataset also creates a problem of too much statistical power which may result in small effect sizes appearing statistically significant (Hair et al., 2014).

Recognising the strengths and limitations of secondary data analysis, and more specifically the NHS Staff Survey, following discussion with thesis supervisors it was decided that the quantitative testing of the model presented in Chapter Five would be conducted through secondary data analyses of the 2014 NHS Staff Survey.

6.1.1 Ethical issues

Secondary data does not involve the recruitment of participants. Hence, it can be argued from an ethical practice to increase the utility and value of existing datasets, and minimise disturbances to potential participants. Regardless, an ethics form was submitted to Birkbeck's Department of Organizational Psychology Ethics Chair. Ethically, secondary data analyses contain some particularly salient issues. These are to ensure that individual participants are not identifiable in the dataset; that reasonable consent is received from participants; and that, the use of the dataset will not cause distress or damage (Morrow, Boddy, & Lamb, 2014; Van den Eynden, Corti, Woollard, & Bishop, 2011). Data protection laws in the United Kingdom mean that agencies which handle and release datasets comply with these specifications (Van den Eynden et al., 2011). In the context of the NHS Staff Survey, its release through the UK Data

Service was only made possible by registering as a postgraduate student with the intention of using this data for thesis work and related publications.

The NHS Staff Survey invitation letters made participants aware that responses would be used by stakeholders within and external to the organisation to improve care for patients and working conditions for staff (Picker Institute Europe, 2015). It also emphasised that all responses would be kept confidential and that individual data released would not be identifiable. As a result, and in compliance with data protection laws, demographic responses were not available when the data was released via the UK Data Service. A request to obtain this data from Picker Institute Europe, who manage the NHS Staff Survey, was denied. Instead, they extended an invitation to conduct analyses on their computers in their offices in Oxford. However, after discussions with thesis supervisors, it was felt that the imposed restrictions on data management and software outweighed the value of this information, and the invitation was declined. Additional data relating to quality of care (e.g., patient satisfaction and hospital mortality) was drawn from NHS agencies that comply with the UK Data Protection Act.

6.1.2 Statistical significance in large datasets

Using datasets with a large sample size, as is the case with the NHS Staff Survey, can result in situations where there is excessive statistical power. This means that even very small, and likely meaningless, effect sizes are observed to be statistically significant (Field, 2014). In the social sciences most constructs are inherently linked although the strength of these relationships varies substantially. Therefore, there is a need to identify which statistically significant relationships are actually practically meaningful. Typically researchers faced with this challenge utilise one or more of the following three options (Hair et al., 2014; T. A. B. Snijders, 2005): reduce the sample size, adopt a more conservative level of significance, and examine effect size.

The first functions by reducing the sample size, often by randomly extracting smaller subsamples of participants in which to repeat analyses (Hair et al., 2014). Reducing the sample size reduces the statistical power available, and any meaningful relationships should replicate across the different subsamples. This approach actually serves to validate results obtained from one dataset within another. This is the approach taken in developing composite measures from the NHS Staff Survey, where factor analyses are carried out in two separate subsamples and then compared. However, the subsequent studies use multilevel modelling where doctors are grouped into trusts and analyses carried out at the individual and trust-level. Although the number of doctors is high (n=14,066), the number of hospital trusts (n=157) is low. Therefore, reducing the number of doctors can result in reducing the number of trusts, which may actually result in too little statistical power for analyses at the trust-level (Hox, 2010).

Reducing the sample is not feasible for multilevel analyses. As such, a more conservative level of significance for individual-level analyses (p<.01) will be used. In addition, significant results will need to have at least a small effect size (Hair et al., 2014). This refers to the magnitude of the relationship between the two variables being measured, without regards for whether it may reflect the true relationship in the population (Field, 2014). Cohen's (1988) distinction between large (.5), medium (.3), and small (.1) effect sizes is widely accepted and will be used to assess effect size in the studies utilising the NHS Staff Survey. It is also worth noting that others (Wolf, 1986) have broken this down further to consider whether an effect size has an educational (>.25) or practical or clinical significance (>.5).

6.2 The NHS Staff Survey

The NHS Staff Survey began in 2003 with the intention to assess staff views about work and wellbeing at their employing NHS trust (NHS England, n.d.). Its purpose was to provide one survey that superseded various existing surveys, including the Department of Health's "10 Core Questions", the Healthcare Commission's "Clinical Governance Review Staff Surveys", and various trusts' own annual staff surveys. The Staff Survey allows NHS organisations to benchmark against other organisations, and the NHS as a whole, on the measures assessed. More specifically, it provides a basis to measure this performance against four pledges set out in the NHS Constitution that NHS employers should provide for their staff (Table 6.1 below; NHS England, n.d.).

Table 6.1: *Four pledges from the NHS Constitution measured in the NHS Staff Survey (NHS England, n.d.)*

	Staff Pledge
1	To provide all staff with clear roles and responsibilities and rewarding jobs for teams and individuals that make a difference to patients, their families, and carers and communities.
2	To provide all staff with personal development, access to appropriate education and training for their jobs, and line management support to enable them to fulfil their potential.
3	To provide support and opportunities for staff to maintain their health, wellbeing, and safety.
4	To engage staff in decisions that affect them and the services they provide, individually, through representative organisations and through local partnership working arrangements. All staff will be empowered to put forward ways to deliver better and safer services for patients and their families.

Aston University was commissioned to develop and then run the NHS Staff Survey; Picker Institute Europe subsequently took over as survey contractors in 2011 (NHS England, n.d.; Picker Institute Europe, 2015). The survey first sat under the auspices of the Healthcare Commission and its replacement, the Care Quality Commission (CQC). From 2011 it was under the charge of the Department of Health, until NHS England took over in 2013. Additionally, the independent NHS Staff Survey Advisory Group exists to provide stakeholder feedback into the survey design and implementation. This group consists of representatives from NHS England, the CQC, the NHS Leadership Academy, NHS trusts, NHS Employers, and trade unions.

All results, questions, and guidance notes since the survey's inception in 2003 are available on the NHS Staff Survey website. Anonymised data for each respondent is made available to researchers via the UK Data Service. The 2014 Staff Survey was the most recent release at the start of the data analysis stage of this thesis.

6.2.1 Sample

It is compulsory for all NHS Trusts to run the Staff Survey (NHS England, 2015b). This includes foundation, acute, specialist, ambulance, mental health, community, and learning disability trusts. NHS Clinical Commissioning Groups, Commissioning Support Units, and Social Enterprises have the option of participating. The survey comprises a random sample of all ten occupational groups: allied health professionals and scientific and technical staff (e.g., occupational therapy, pharmacy, psychotherapy); medical and dental; ambulance; public health; commissioning; registered nurses and midwives; nursing or healthcare assistants; social care; wider healthcare team (e.g., admin & clerical, maintenance/ancillary); and general management. The minimum sample size is dependent on the size of the organisation, with those employing fewer than 600 employees required to run a census and those with more than 3,000 employees sampling 850 employees. Trusts also have the option to increase sample size by conducting an extended sample or a full census. Eligible employees must be directly employed by the organisation on the 1st of September that year. Employees who are on long-term sick leave, sub-contracted, seconded, and student nurses, are not relevant for this survey.

The Staff Survey is administered annually at the end of September (NHS Staff Survey Co-ordination Centre, 2014a). Two subsequent reminders are sent at three-week intervals. The survey is conducted in paper format, with responses returned by post to an external supplier to preserve the confidentiality of the responses (NHS England, 2015b). Since 2014, organisations with at least 65% of their workforce using an active work email address have the option of surveying these employees electronically. Although the survey is repeated annually, individual responses are not tracked; therefore, each survey year provides a cross-sectional dataset.

The 2014 NHS Staff Survey surveyed over 624,000 employees from 287 NHS organisations in England. In total, 255,150 (42%) responses were received. For the purpose of this thesis, the study sample was restricted to the medical and dental occupational group (*n*=17,670), which constituted 6.9% of the total dataset. The NHS Staff Survey does not distinguish between medical and dental staff, and subsequent use of the term "doctor" encompasses both these medical disciplines. The sample was further limited to doctors working in either acute or specialist trusts.

The final sample used in this thesis comprised 14,066 hospital doctors, of which 94.1% worked in acute trusts. In total 157 trusts were represented, including 18 acute specialist trusts. Mean doctors per trust was 89.59 (*SD*=94.76) with a median of 41 doctors, and ranged between 11 and 458 doctors per trust. In terms of tenure, 16.3% had been with their trust for less than a year, and 19.5% had been there for more than 15 years. The rest had tenures between 1-2 years (14.2%), 3-5 years (16.3%), 6-10 years (17.3%), and 11-15 years (14.7%). Due to data protection, demographic information including age and gender was not available with the datasets.

However, past research involving doctors demonstrate age and gender either having no (Guthrie, Tattan, Williams, Black, & Bacliocotti, 1999; Hakanen et al., 2005; Hill, Rolfe, Pearson, & Heathcote, 1998; J. Klein et al., 2011) or limited (von dem Knesebeck et al., 2010) relationships with wellbeing and psychosocial risk exposure.

6.2.2 Content

Since its inception, the NHS Staff Survey has undergone annual changes. Work-life balance, flexible working, childcare arrangements, and turnover intention are examples of measures dropped. In contrast, work engagement, work-related stress, and presenteeism have since been included. A copy of the 2014 Staff Survey is available in Appendix V. The Survey contained 113 questions that cover six general areas, separated into the following categories:

1. Personal development:
    - Contained items on the quality of training and appraisals received
2. The job:
    - Items assessed respondents' experiences of work, including: employee participation, work engagement, job satisfaction, and teamwork
3. The managers:
    - Items related to feedback, communication, and support from immediate and senior managers
4. The organisation:
    - Items referred to the organisation, including whether training is encouraged or if staff would recommend treatment to their friends and family
5. Health, wellbeing, and safety at work:
    - Assessed how the job role affects health, presenteeism, witnessing errors and incidents, and exposure to violence and harassment
6. Background information:
    - Contained demographic questions such as: age, gender, ethnicity, occupational group, and tenure

Many of the measures used in the Staff Survey originate from established measures in the field of occupational psychology (West, 2015). For example, the three items relating to work engagement were drawn from the Utrecht Work Engagement Survey (Admasachew & Dawson, 2010). However, there is no clear guide detailing the exact source of all measures and items used in the survey. In addition, items are adjusted before inclusion in the survey, and between years, based on input from the NHS Staff Survey Advisory Group (NHS England, 2015b).

What is not clear is how to tabulate all Staff Survey items, and how these map against the constructs being measured. In its reporting of results, the Staff Survey maps 29 "Key Findings" against the four Staff Pledges. A document is produced annually for every Staff Survey detailing how each Key Finding is tabulated. It does not, however, explain why this is the case. At times this does not match the face validity of items, such as when the Key Finding on job satisfaction in 2014 required the summing of only 7 out of the 8 job satisfaction items. In addition, 11 Key Findings are based on single item measures, while 35 non-demographic items in the Staff Survey are not included in any of the Key Findings.

Using composite measures here not only allows for better capturing of complex multifaceted latent variables (Hair et al., 2014), but also presents fewer challenges than single item measures when used in multivariate modelling (Kline, 2016). Given the uncertainty with scoring, questions arise around the validity and reliability of the items and constructs. It is therefore imperative to establish the psychometric validity of the measures within the 2014 NHS Staff Survey, especially with regard to creating composite measures.

6.3 Creating Composite Measures

Any form of measurement contains error - that is, the discrepancy between the true score and the obtained score (Field, 2014). This is particularly an issue when using self-reported measures and when assessing latent constructs. Therefore, measures with strong psychometric properties are essential in reducing measurement error and for any meaningful analysis (Hair et al., 2014; Kline, 2016). Too often researchers rely on inferring reliability of measures from other studies, at the expense of assessing validity and reliability within their own studies. The mistaken assumption being that once a measure has been deemed acceptable it is applicable for all studies. Instead, it is the responsibility of the researcher to demonstrate that the psychometric properties of measures used are suitable to the context in which they are used (Kline, 2016).

To understand what makes a measure psychometrically robust it is useful to first briefly review classical test theory (DeVellis, 2006). The theory provides a set of principles that help determine how successful chosen indicators are in recording the underlying latent construct.

Recognising that all measures contain measurement error (Field, 2014), this means that the "true" score comprises of the sum of the observed and the error score. Consequently, a good observed score is one containing a minimal amount of measurement error (DeVellis, 2006). When this occurs, reliability increases, which improves the confidence in the interpretation of the research.

Reliability increases as the number of items within a measure increases, leading to recommendations to reduce measurement error by using composite measures instead of single item measures (DeVellis, 2003; Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012). Composite measures aggregate a number of related items to obtain one score representing the underlying construct (Hair et al., 2014). In addition to providing better coverage of the construct compared to single item measures (Hair et al., 2014; Ironson et al., 1989), using multiple items averages out errors and specificities that are inherent in single items (DeVellis, 2003). Collectively, this is assumed to reduce the level of measurement error. The evidence favouring composite measures over single items is not conclusive, with some studies demonstrating the superiority of composite measures as predictors (Diamantopoulos et al., 2012) and others reporting no real difference (Ironson et al., 1989; Wanous, Reichers, & Hudy, 1997). Nevertheless, single item measures remain difficult to validate.

Single items ignore the unreliability of measurement which is incongruent with structural equation modelling (Petrescu, 2013). The main issues that arise are with identification and convergence. Just or under-identified models mean a unique solution cannot be generated by the analysis, and the specified model cannot be assessed. Although there are techniques to incorporate single items into multivariate modelling statistics (Kline, 2016; Petrescu, 2013), researchers are still encouraged to use multi-item measures where possible (Hair et al., 2014).

Considering the advantages of using composite measures, and the difficulties presented with single item measures, this thesis will use composite measures where possible. The challenge, however, lies in developing composite measures from the 2014 NHS Staff Survey given the limited information provided on scoring. In doing so, it is paramount that the integrity of the developed measures is established by demonstrating its validity and reliability.

6.3.1 Validity

At the heart of understanding validity is construct validity, which asks the fundamental question: are we actually measuring what we claim to be measuring (Kline, 2016)? There is no single definite test to demonstrate construct validity, nor is it possible to do so in a single study. Instead, construct validity is gradually established through the accumulation of information on other aspects of validity (Ironson et al., 1989). It involves two things (Loevinger, 1957): (i) determining the internal model of measure, including domains and subdomains; and (ii) the external aspect, where the measure's relations with other variables are established. These are reliant on different facets of validity, with the former represented by content validity and factor analysis, and the latter through convergent, discriminant, and criterion validity (Goodwin, 1999).

Internal validity. The distinction of internal and external components of construct validity (Loevinger, 1957) presented the first blueprint from which factor analysis is used to understand the validity of a particular measure. As most constructs cannot be observed, factor analysis provides a method in understanding the underlying structure and makeup of a measure (Goodwin, 1999). Composite measures should be unidimensional, meaning that items are strongly related to each other and represent the same construct (Hair et al., 2014). Factor analysis facilitates this process by examining the correlations between large numbers of variables and grouping them into a set of underlying dimensions (i.e., factors).

Factor analysis can be exploratory or confirmatory in nature. These are introduced in greater detail in the next chapter. Briefly, exploratory factor analysis (EFA) seeks to reduce the number of items and reveal what the underlying structure of the variables within the dataset are (Field, 2014; Hair et al., 2014). For the purpose of construct validity, EFA not only indicates how many underlying dimensions exist within the construct, but what items are relevant to these constructs, and how strongly they load onto these (Goodwin, 1999; Nunnally & Bernstein, 1994). In addition, EFA demonstrates how strongly these different dimensions relate to each other – an indicator of convergent/divergent validity. Confirmatory factor analysis allows the researcher to confirm a hypothesised underlying latent variable structure (Byrne, 2012). Poorly fitting models here could indicate incorrect or unsuitable items within the specified internal structure of a construct (Nunnally & Bernstein, 1994). It also allows the determination of one construct's relationship with another. Understanding why items load onto specific dimensions can also be illuminating towards content validity. For example, when items load incorrectly it could indicate items are being misunderstood and do not accurately represent what they are intended to (Nunnally & Bernstein, 1994).

External validity. Assessing a construct against other measures is the basis of examining the external components of a construct's validity (Goodwin, 1999; Hair et al., 2014; Loevinger, 1957). Suitable measures that represent relevant constructs within the original construct's nomological network need to be identified. Convergent validity requires measures to correlate strongly with other measures supposedly representing the same construct. In contrast, discriminant validity is demonstrated when the measure in question correlates lowly, or does not correlate at all, with measures representing unrelated constructs. Similarly, criterion validity examines whether the measure accurately relates to a criterion that it is expected to predict (Kline, 2016). In other words, does the measure explain an appropriate amount of variance in the criterion? The timing of when the outcome variable is measured can vary, and can exist in future (predictive), current (concurrent), or past (retrospective) form. Correlation analyses form the basis for understanding these relationships, although regression and multivariate modelling present more complex alternatives (DeVellis, 2003).

Content validity examines whether the measure's items accurately represent the construct of interest (DeVellis, 2003; Hardesty & Bearden, 2004). This ties closely to face validity, where items appear to measure the construct in question. This not only lends credibility to the measure, but is fundamental to ensuring that construct is accurately captured and presented (Hardesty & Bearden, 2004). Establishing content validity is not reliant on statistical analysis, and is instead based largely on the opinions of subject-matter experts (Field, 2014; Hair et al., 2014). However, face validity does not equal actual validity, as items can appear to be measuring one construct and in reality be measuring something else. This underlines the need for statistical examination of the developed measure, meaning that subjective ratings from experts do not suffice for the purpose of validity (Hardesty & Bearden, 2004).

6.3.2 Reliability

The focus thus far has been on validity. This risks understating the importance of reliability in classical test theory (DeVellis, 2006), which concerns whether an instrument can be interpreted consistently across different situations (Field, 2014). Reliability exists within the measure (i.e., internal reliability) or across different measurement points (Hair et al., 2014).

Coefficient (or Cronbach's) alpha is the most commonly used reliability coefficient (Kline, 2016). It assesses the extent to which items in a measure are consistent, which is known as internal consistency. Coefficients range from zero to one, with higher coefficients indicating stronger consistency. Although scores of about .70 are typically considered acceptable, this cut-off figure for internal reliability originally referred to early stages of research where modest levels of reliability sufficed (Lance, Butts, & Michels, 2006; Nunnally, 1978). Instead, a reliability of *at least* .80 should be used to reduce measurement error (Nunnally, 1978). Aside from internal consistency, test-retest measures the consistency of the measure between two time points while split-half reliability divides a sample in two to separately assess reliability in each sample (Hair et al., 2014). Establishing reliability is fundamental to ensuring that items are consistent in measurement as poor reliability not only reduces statistical power, but also reduces true effect size (Kline, 2016). It is important to emphasise that high reliability does not equal validity, although reliability is an important precursor to validity (DeVellis, 2006).

6.4 Conclusion

The challenges of conducting primary data collection amongst doctors, paired with the suitability of the NHS Staff Survey for the aims of this thesis resulted in the decision to use secondary data analysis for this thesis. However, any analysis based on the 2014 NHS Staff Survey needs to be assured that measures demonstrate validity and reliability. It is beyond the scope of this thesis to fully validate the items of the Staff Survey. However, reasonable steps will be taken to do so. This forms the basis of the next chapter. First, items are selected based on the occupational health psychology literature and an examination of face validity by subject-matter experts. Second, the underlying structures of items are examined through a series of exploratory factor analyses. The relationships between the different constructs are then confirmed through confirmatory factor analysis, which would provide support for convergent and divergent validity. Finally, Cronbach's alphas examine the internal consistency of the composite measures. To cross-validate the findings from this thesis and improve generalisability (Hair et al., 2014), the dataset will be randomly split into two samples with an EFA carried out with one half and a subsequent CFA in the other.

Chapter 7 : Developing Composite Measures from the NHS Staff Survey

The previous chapter introduced the 2014 NHS Staff Survey, and outlined the value in developing composite measures and the importance of establishing the validity and reliability of measures used. The main purpose of this chapter is to build on the principles introduced in the preceding chapter and develop valid and reliable composite measures from the NHS Staff Survey.

To do so, exploratory factor analyses were conducted to establish the underlying dimensions within the 2014 NHS Staff Survey and to reduce the number of items into composite measures. The factors extracted were subsequently confirmed using a confirmatory factor analysis. These were respectively carried out in two separate datasets (*n*=7,033 each). The final section discusses the internal reliability of the newly developed composite measures and the issue of common method bias.

7.1 Introducing Factor Analysis

As previously described, the primary purpose of running exploratory factor analysis (EFA) is to determine the underlying structure of the variables within the dataset (Field, 2014; Hair et al., 2014). This is essential for assessing construct validity. The 2014 NHS Staff Survey consisted of 113 items assessing different aspects of work. To reduce the number of items and to facilitate the creation of composite measures from the NHS Staff Survey, EFA was carried out.

This section outlines the steps taken and decisions considered in the preparation and subsequent analysis. First, the difference between EFA and principle component analysis is reviewed. Then, using the procedures set out by Hair et al. (2014), item selection, sample size, and inter-correlations were considered prior to undertaking the EFAs. Factor extraction and rotation are discussed before the final section presents the results.

7.1.1 Factor analysis or principle component analysis

Two approaches exist for locating underlying dimensions within datasets: common factor analysis and principal component analysis (PCA; Field, 2014; Hair et al., 2014). The difference between the two approaches lies in how explained and unexplained variance is treated. Common factor analysis considers only the common or shared variance, meaning that unique or error variance is not assumed to be of interest to the structure. This focus on shared variance is useful when the purpose of analysis is to identify the latent constructs or dimensions for theory building. In contrast, PCA utilises the total variance, including common, unique, and error variance. PCA is used primarily for data reduction, where the objective is to focus on the least amount of factors to explain the maximum possible total variance within the variables. Within the academic literature there has been substantial disagreement on which approach is superior (Mulaik, 1990; Velicer & Jackson, 2004). In fact, once the number of variables exceeds 30, there is little distinction between common factor analysis and PCA (Gorsuch, 1983). When it comes to creating composite measures, it has been argued that the creation of latent theoretical frameworks means that common factor analysis is more appropriate (Hair et al., 2014; Velicer & Jackson, 2004).

A more important point to consider in the context of this dataset is which approach is more appropriate for non-linear and dichotomous data. Both EFA and PCA are built upon correlational analysis, and therefore assume that included items should be linear and in metric (i.e., interval or ratio) form (Manisera, Van der Kooij, & Dusseldorp, 2010; Meulman, Van der Kooij, & Heiser, 2004). The reality is that Likert scales - despite being ordinal data - are widely used within psychology and the social sciences in place of interval data (Carifio & Perla, 2007). What factor analysis does not handle well is data that is nominal in nature; although some (e.g., Gower, 1966; Jollife, 2002) argue that PCA is able to summarise variation in nominal data. While it is plausible to develop an argument justifying using PCA (and even EFA) with ordinal and nominal data, the advent of non-linear techniques has provided researchers with alternative methods that are congruent with these forms of data (Linting, Meulman, Groenen, & Van der Kooij, 2007; L. K. Muthén & Muthén, 2017). This means that the researcher specifies the level of the data (nominal, ordinal, interval), allowing non-interval data to be quantified and revealing the shape of the relations between them. This can be done using the Mplus statistical software by first specifying items as categorical, and then using an appropriate model estimator to handle this non-linear form of data.

7.1.2 Considerations before factor analysis

The section below discusses some of the points that need to be considered prior to carrying out a factor analysis, mainly the model estimator to be used, items to be selected and included, and the appropriate sample size.

Model estimation. The most common estimator, and the default in many statistical programmes, is maximum likelihood (ML; Flora & Curran, 2004). Its effectiveness, however, is reliant on having an adequately large sample size, data that is normally distributed, and a properly specified model (Browne, 1984). Adaptations of ML have led to a more robust variant estimator (i.e., MLM). The MLM corrects the χ^2 by including a scaling correction using the sample's kurtosis values when violations of normality occur (Asparouhov & Muthén, 2013). As a result, studies have found that even when distributions and sample size vary, MLM remains a good estimator of mean and covariance structures (Curran, West, & Finch, 1996).

MLM still does not address the issue of requiring continuous data. The general consensus is that normally distributed Likert scales with more than four categories are amenable to be treated as continuous data within confirmatory factor analysis (Curran et al., 1996; B. O. Muthén & Kaplan, 1992). However, four variables within the NHS Staff Survey are dichotomous variables with "yes/no" responses. This creates a problem as such variables likely inflate χ^2 values, reduce the strength of parameter estimates due to ceiling and floor effects, and generate pseudo-factors resulting from item extremeness and difficulty (Brown, 2006; DiStefano, 2002; Green, Akey, Fleming, Hershberger, & Marquis, 1997).

Instead, weighted least square (WLS) is an estimator congruent with dichotomous and categorical data as it uses standard estimates drawn from polychoric, polyserial, and tetrachoric correlations (Browne, 1984; Hox, Maas, & Brinkhuis, 2010). To address WLS' vulnerability to small sample sizes and numerous variables (Flora & Curran, 2004; Hox et al., 2010; B. O. Muthén, du Toit, & Spisic, 1997), Mplus provides two robust variants of the WLS: WLSM and WLSMv (L. K. Muthén & Muthén, 2017). These differ in that WLSM uses mean-adjusted χ^2, while the χ^2 in WLSMv is mean and variance-adjusted (Hox et al., 2010). So while estimates and standard errors remain equivalent, the different estimators provide different fit indices. Muthén and Muthén (2010) advocate WLSMv over WLSM in Mplus, as simulation studies have shown WLSM to have higher Type I error rates (Asparouhov & Muthén, 2013; B. O. Muthén et al., 1997) while WLSMv is almost as accurate as normal theory maximum likelihood (Hox et al., 2010). Although the sample size in this study should be sufficiently large to consider using the WLS or WLSM estimators, the NHS Staff Survey includes both skewed and dichotomous data. With this in mind, the WLSMv is selected as the choice estimator as it provides a more conservative and robust approach than the other estimators.

Item selection and inclusion. Two points need to be considered when selecting items for inclusion within EFA. The first relates to the type of variable that should or can be included. As the inclusion of nominal and ordinal data is possible within EFA (L. K. Muthén & Muthén, 2017), traditional concerns about the absence of interval or ratio data are moot. The second point relates to the number of items that should be included within an EFA. The conceptual underpinning of items should inform the selection and inclusion of items as the indiscriminate addition of a large number of items will likely yield poor results (Hair et al., 2014). However, sufficient items are still needed to allow for a reasonable number of items per factor. Of the 113 itemed 2014 NHS Staff Survey, only items relevant to one or more of the five main latent constructs identified in the theoretical model (job demands, job resources, strain, motivation, quality of care) were included. This list of included and excluded items was then discussed with thesis supervisors. To further validate the review of these items, this list was independently reviewed by one subject-matter expert from the University of Nottingham and another from Herriot-Watt University.

Sample size. The reliability of an EFA is reliant on sample size, as correlation values can fluctuate across samples (Field, 2014). Different rules-of-thumb exist to guide researchers: minimum sample size or minimum observations-per-variable ratio. In the former, guidelines vary, including having no less than 50 and preferably more than 100 participants (Hair et al., 2014); having at least 300 participants (Tabachnick & Fidell, 2012); or that fewer than 100 responses was poor, more than 300 good, and more than 1000 excellent (Comrey & Lee, 2012). Others argue an acceptable sample size would have ten or more observations for each variable in the analysis, with some researchers advocating a minimum ratio of 20:1 (Hair et al., 2014). However, Arrindell and van der Ende (1985) demonstrated that observations-per-variable have little impact on stability of factor solutions, concluding that minimum sample size should be used instead. The large sample size of the two samples in this study (n=7,033 each) means all rules-of-thumb for both minimum sample size and observations-per-variable ratio are met.

7.1.3 Considerations during factor analysis

After the factor analysis has been carried out, the following questions need to be considered: how many factors should be extracted? And, how to interpret the factors? These are reviewed in the section below.

Factor extraction. It is possible to obtain as many factors in an EFA as there are variables included (Field, 2014). The process of factor extraction discerns which of these factors are important enough to retain. Traditional factor extraction criterian, including scree plots and the latent root criterion, are suitable for use with categorical EFAs. However, differing results between these two (and other) extraction criterian means researchers are encouraged to use more than one criterian when extracting factors (Field, 2014; Hair et al., 2014; Linting et al., 2007). For example, using a scree plot typically results in one or two extra factors than when using the latent root criterion.

Scree plots are where each eigenvalue (y-axis) is plotted against its corresponding factor (x-axis) in their order of extraction, which helps illustrate the relative importance of each factor. High eigenvalues on the left of the plot progress towards lower eigenvalues on the right: characterised as a curve with a sharp descent that straightens out (Hair et al., 2014). At the point where the shape of the curve changes, or breaks, should indicate the last factor accounting for sufficient variance within the data.

The difficulty in reading scree plots leads to some suggesting the retention of all factors with eigenvalues greater than 1 (Kaiser, 1960). This is known as the latent root criterion and is based on the assumption that an eigenvalue of 1 represents a substantial amount of variance. According to Field (2014), latent root criterion is more accurate when there are less than 30 included variables, or when the sample size exceeds 250. Similarly, others have argued this technique to be most reliant when using between 20 and 50 variables (Hair et al., 2014). Too few variables would result in a conservative number of factors extracted, while too many variables results in too many factors extracted.

Running EFAs in Mplus provides an additional dimension to consider whereby the minimum and maximum number of factors to extract has to be specified prior to analyses. This then leads to a model being specified for each specific set of factors. Unlike SPSS, Mplus provides model fit statistics for each of these specified models as it would in a confirmatory factor analysis (see Section 7.3.2 for more information). As such, final determination of the number of factors to extract is determined based on the collective information from scree plots, eigenvalues, and model fit data.

Factor interpretation and rotation. EFA computes a factor matrix indicating how each variable loads onto each factor. These factor loadings represent the correlation between each variable and the factor (Hair et al., 2014); higher loadings indicate a stronger connection to that particular factor. From a statistical perspective the EFA has completed its purpose to reduce data. However, the researcher needs to decide whether the factor and the items that load onto it present a suitable interpretation of the data used. Should the answer be no, the researcher may then employ a rotational method to obtain a solution that is a simpler and better theoretical fit.

Rotation refers to the turning of the factor axes in order to redistribute variance from earlier factors to later ones, so that a more theoretically meaningful and simpler factor pattern can be achieved (Hair et al., 2014). These rotations can either be orthogonal or oblique (Field, 2014). Prior to rotation all factors are independent, or unrelated to each other. Orthogonal rotation ensures that the factors remain independent, or uncorrelated, during rotation. In contrast, oblique rotation allows factors to be correlated. Because of this flexibility, oblique rotations are more realistic as the underlying theoretical dimensions are assumed to be correlated to each other (Hair et al., 2014). This is important given that nearly all psychological constructs are to some degree related to each other (Field, 2014). Given that this study would theoretically expect factors to correlate to each other, oblique rotation will be used. More specifically, Mplus uses geomin as the default oblique rotator (L. K. Muthén & Muthén, 2017).

After the factors have been rotated, the factor loadings need to be examined for a variable's significance to a factor. Loadings of .40 or higher are considered acceptable (Field, 2014), although a sufficiently powerful sample size (n>350) means >.30 factors loadings have practical significance (Hair et al., 2014). Technically all variables load on all factors to some degree, and it is not uncommon for a variable to substantially load onto more than one factor - an occurrence known as cross-loading (Field, 2014; Hair et al., 2014). It is anticipated that rotation will reduce the likelihood of this occurring, and several options exist should cross-loading occur or a variable not load sufficiently (Hair et al., 2014). First, problematic variables could be ignored, although this would need to be explained and justified. Second, items could be considered for deletion. Before this occurs, the variable's communality value should be examined. This refers to the proportion of variance of the variable that is explained by all the extracted factors (Goldberg & Digman, 1994). Hair et al. (2014) suggests that a communality of less than .50 is of little significance and provides a greater justification for removal. Third, a different rotational method could be used to examine whether more suitable factors could be obtained. In the context of this study, the decision will be framed by the practical, statistical, and theoretical justification and implications.

7.2 Results: Exploratory Factor Analyses

The dataset (*n*=14,066) was first randomly divided into two subsets using the select random cases function in SPSS. Within the first sample subset (*n*=7,033), an EFA using Mplus 8 was run for each of the three different domains (Goldberg & Digman, 1994): psychosocial working conditions, work-related wellbeing, and care outcomes. The ability to concurrently examine nominal and ordinal data means that these variables, where relevant, were added within the same EFA. The second subset was used to test the subsequent CFA.

In discussion with thesis supervisors and external subject-matter experts, a number of items pertaining to health and safety climate were also identified. These were conceptually similar to the concept of psychosocial safety climate (Dollard & Bakker, 2010), which is hypothesised to operate as an antecedent to job demands and resources. However, including these items (and construct), although interesting, is beyond the scope of this thesis. Therefore, these items were not considered in the following EFAs. However, the inclusion of these items into an earlier EFA indicated that it did load as a distinct factor, and may warrant further exploration in future.

7.2.1 Psychosocial working conditions

To create composite measures for working conditions, 24 items from the NHS Staff Survey judged to represent job demands or job resources were included in this analysis. One item ("working excessive hours each week") required a binary yes/no response. Binary responses are unique in that they can be interpreted as both categorical and ordinal data (De Leeuw, 2006; Eastwood et al., 2013). The remaining items comprised 19 that were ordinal in nature, and four items that assessed frequency of exposure to bullying and violent behaviours ("never", "1-2 times", "3-4 times", "6-10 times", "more than 10 times"). All items retained some form of ordinal structure, therefore, the data was congruent for analyses as categorical variables within Mplus (L. K. Muthén & Muthén, 2017).

Sample size and missing data. The sample of 7,033 per study exceeded the rules-of-thumbs on EFA sample sizes (Field, 2014; Hair et al., 2014). Using the variable with the highest number of missing cases, observations-per-variable was 1:284 - exceeding the suggested 1:20 ratio (Hair et al., 2014). Missing observations on each variable ranged from 33 (0.5%) to 219 (3.1%). The WLSMv estimator used provides a robust handling of missing data as it provides consistent estimates here without the need for multiple imputation (Asparouhov & Muthén, 2010c). Missing data in WLSMv is treated with pairwise deletion, which is generally superior to listwise deletion or mean-substitution (Asparouhov & Muthén, 2010c; Peugh & Enders, 2004).

Table 7.1: *Eigenvalues across different factors and corresponding model fit*

	Factor							
	1	2	3	4	5	6	7	8
Eigenvalue	9.21	2.15	1.96	1.62	1.35	1.11	0.82	0.81
RMSEA	.16	.14	.12	.09	.07	.05	†	†
CFI	.82	.88	.92	.96	.98	.99	†	†
TLI	.81	.86	.90	.94	.97	.98	†	†

Note. † model did not converge

Factor extraction. The respective minimum and maximum number of factors to extract for the EFA were set at one and eight respectively. This assumed the possibility of a minimum of one factor, and two additional factors beyond the expected six anticipated in the discussion with subsect matter experts. Table 7.1 lists eigenvalues for each factor and the corresponding model fit if that factor solution would be retained. It only provides model fit data up to a six-factor solution as the model could not converge when a seven or eight-factor model was run. From Table 7.1 it is clear that six factors exceeded the eigenvalue threshold of 1. Examination of the model fit was also congruent with a six-factor solution, as model fit statistics met the markers for good fit (i.e., RMSEA<.05; CFI>.95; TLI>.95; Byrne, 2012; Hu & Bentler, 1998). The inflection point in the scree plot (Figure 7.1) is not clear, and could occur after the second or sixth factor. Considering this information collectively, it was decided to retain six factors.

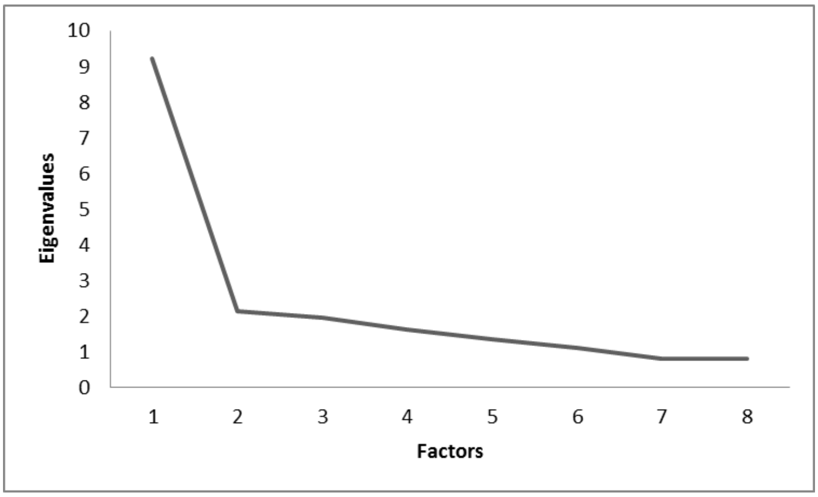

Figure 7.1: Scree plot depicting eigenvalues psychosocial working conditions

Factor loadings. The geomin rotation loaded 21 items onto one of six factors. This therefore highlighted an issue with three items. Three items ("ExH: working extra hours per week"; "DM1: I am unable to meet all the conflicting demands on my time at work"; "being harassed or bullied by managers or colleagues") did not adequately load onto any of the six factors. AG4 (.34), DM1 (.13) and ExH (.23) items all had very low communalities.

The remaining 21 items all had a factor loading of at least .59 (Table 7.2). The first factor consisted of five items pertaining to "manager support", with factor loadings between .78 and .95. This was followed by four items on "role clarity" (loadings: .59 to .84). The third factor referred to "effective team practices" and included three items that had factor loadings between .69 and .87. Factor four comprised two items related to "insufficient work resources", with loadings between .66 and .83. "Workplace aggression" was the fifth factor, and included three items referring to being physically or verbally abused by patients and colleagues (loadings: .60 to .79). The four items that loaded between .65 and .88 onto the last factor related to the degree of "job control" in the workplace.

Summary. EFA revealed six factors with 21 items pertaining to doctors' psychosocial working conditions. Three items - working extra hours per week, experiencing conflicting demands, and being bullied by colleagues were removed due to poor loadings and low communalities. The aggregation of items within the factors resulted in six composite measures: manager support, role clarity, effective team practices, insufficient work resources, workplace aggression, and job control.

Table 7.2: *Factor loadings for oblique rotated solutions with psychosocial working condition items*

Item	1 Manager support	2 Role clarity	3 Effective team practices	4 Insufficient work resources	5 Workplace aggression	6 Job control
My immediate manager can be counted on to help me with a difficult task at work	.95					
My immediate manager encourages those who work for her/him to work as a team	.86					
My immediate manager gives me clear feedback on my work	.86					
My immediate manager is supportive in a personal crisis	.83					
My immediate manager asks for my opinion before making decisions that affect my work	.78					
I always know what my work responsibilities are		.84				
I am trusted to do my job		.76				
I have clear, planned goals and objectives for my job		.60				
I am able to do my job to a standard I am personally pleased with		.59	.34			
Team members have to communicate closely with each other to achieve the team's objectives			.82			
Team members have a set of shared objectives			.87			
Team members often meet to discuss the team's effectiveness			.69			
There are enough staff at this organisation for me to do my job properly				.83		
I have adequate materials, supplies and equipment to do my work				.66		
In the last 12 months how many times have you personally experienced physical violence at work from patients/service users, their relatives or other members of the public?					.68	
In the last 12 months how many times have you personally experienced harassment, bullying or abuse at work from patients/service users, their relatives or other members of the public?					.79	
In the last 12 months how many times have you personally experienced physical violence at work from a manager/team leader or other colleagues					.60	
In the last 12 months how many times have you personally experienced harassment, bullying or abuse at work from a manager/team leader or other colleagues					. 34	
I am involved in deciding on changes introduced that affect my work area/team/department						.88
I am able to make improvements happen in my area of work						.78
I am able to make suggestions to improve the work of my team/department						.80
There are frequent opportunities for me to show initiative in my role						.65

Note: n=7,033. Loadings less than .3 are suppressed.

7.2.3 Work-related wellbeing

Within the job demands-resources model, work-related wellbeing is typically operationalised as burnout and work engagement (Demerouti et al., 2001; Schaufeli & Bakker, 2004), although other related constructs have also been used, including: job satisfaction (Tims, Bakker, & Derks, 2013), psychological strain (Bakker, Boyd, et al., 2010; Dollard & Bakker, 2010), repetitive strain injuries (Bakker, Demerouti, & Schaufeli, 2003), organisational commitment (Bakker, Boyd, et al., 2010; Bakker, van Veldhoven, et al., 2010), and presenteeism (Demerouti et al., 2009).

Pre-analysis checks. In discussion with thesis supervisors and the external subject-matter experts, thirteen items from the NHS Staff Survey were judged relevant to the domain of work-related wellbeing. Eleven of these items recorded ordinal data on a five-point Likert scale. Two items used binary responses (yes/no). As all items contained ordinal structured responses, Mplus was able to treat these as categorical variables (L. K. Muthén & Muthén, 2017).

The item asking respondents if they had gone to work even though they did not feel well enough to perform (i.e., presenteeism) had a substantially higher proportion of missing data (13.9%) as compared to the other thirteen items. Despite the lower number of respondents here, the observation-to-item ratio (432:1) indicated sample size was not an issue in this analysis. The levels of missing data on the other items ranged from 0.4% to 1.9%. The WLSMv estimator was used due to its robustness in handling non-linear and missing data (Asparouhov & Muthén, 2010c; B. O. Muthén et al., 1997).

Factor extraction. The number of factors specified to be extracted was between one and five. This assumed the possibility of a minimum of one factor, and one additional factor beyond the five anticipated in the discussion with subsect matter experts. The eigenvalues and the model fit statistics for each factor solution are presented in Table 7.3. Examining the eigenvalues it appears that a three-factor structure should be retained, although according to model fit statistics it did not have good fit. Instead, the four-factor structure had a better model fit as RMSEA was less than 0.50 while CFI and TLI both exceeded .95 (Byrne, 2012; L. Hu & Bentler, 1998). The scree plot (Figure 7.2) supports the model fit statistics as it indicated an inflection point at both the second and fourth factor. As a result, four factors were retained.

Table 7.3: *Eigenvalues across different factors and corresponding model fit (Work-related wellbeing)*

	Factor							
	1	2	3	4	5	6	7	8
Eigenvalue	6.65	1.32	1.07	0.78	0.69	0.54	0.43	0.39
RMSEA	.18	.12	.06	.045	.02			
CFI	.90	.96	.99	1.00	1.00			
TLI	.88	.95	.99	.99	1.00			

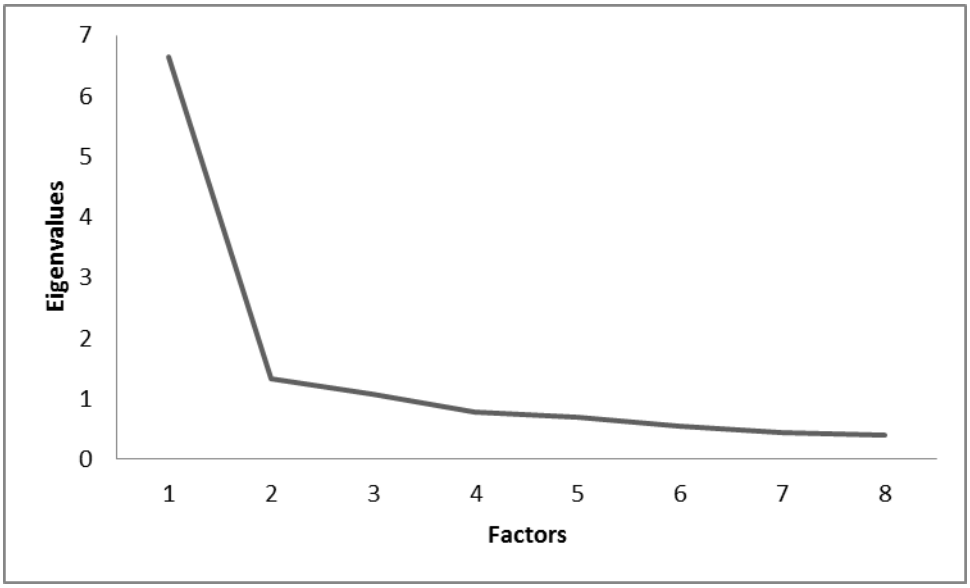

Figure 7.2: Scree plot depicting eigenvalues for work-related wellbeing

Factor loadings. The four-factor geomin rotated loadings revealed that the 13 included items loaded onto one of the four factors (loading >.39; Table 7.4). The first factor was labelled "satisfaction with job content", and consisted of five items with a factor loading between .39 and .94. The second factor had three items that pertained to "satisfaction with support" (loadings: .59 to .97). Three items made up "work engagement", which represented the third factor extracted. Items here loaded from .67 to .95. The final factor referred to as "negative work-related wellbeing", consisted of two items assessing stress and presenteeism with loadings from .53 to .75.

Table 7.4: *Factor loadings for oblique rotated solutions with work-related wellbeing items*

Items	1 Satisfaction with job content	2 Satisfaction with support	3 Work engagement	4 Negative work-related wellbeing
The recognition I get for good work	.94			
The support I get from my immediate manager	.64			
The freedom I have to choose my own method of working	.43	.39		
The extent to which my organisation values my work	.85			
My level of pay	.39			
The support I get from my work colleagues		.59		
The amount of responsibility I am given		.97		
The opportunities I have to use my skills		.79		
I look forward to going to work			.83	
I am enthusiastic about my job			.95	
Time passes quickly when I am working			.67	
In the last three months have you ever come to work despite not feeling well enough to perform your duties?				.53
During the last 12 months have you felt unwell as a result of work-related stress?				.75

Note. n=7,033. Loadings less than .3 are suppressed.

Summary. After an EFA with categorical variables in Mplus, thirteen work-related wellbeing items were found to load onto four factors. Specifically, these factors were: satisfaction with job content, comprising five items; satisfaction with support, had three items; work engagement, comprising three items; and two items loaded onto negative work-related wellbeing.

7.2.3 Quality of care

Items from the NHS Staff Survey pertaining to quality of care were first reviewed with the thesis supervisors and then reviewed by the two external subject-matter experts. Eight possible items were identified.

Pre-analysis checks. Six items were ordinal data on a five-point Likert scale. The remaining two items asked respondents whether they had seen any errors, near misses or incidents that could have hurt patients or staff. These used a binary response of 'yes' and 'no'. Sample size was sufficiently adequate, as the lowest observations-per-variable was 858:1. The proportion of missing data was very low: between 1.1% and 2.6% of data. WLSMv was used as the estimator. Between one and four factors were specified to be extracted. This was based on having one additional factor beyond the three expected factors from the discussion with subsect matter experts.

Table 7.5: *Eigenvalues across different factors and corresponding model fit (Quality of care)*

	Sample	Factor							
		1	2	3	4	5	6	7	8
Eigenvalue	A	3.88	1.52	1.21	0.41	0.35	0.28	0.19	0.16
RMSEA	A	.25	.21	.02	.000				
CFI	A	.88	.95	.99	1.000				
TLI	A	.84	.88	.99	1.000				

Factor extraction. Three types of quality-of-care outcomes were hypothesised, and therefore between one and five factors were specified to be extracted with the EFA. The eigenvalues from these analyses are presented in Table 7.5 above. Using the latent root criterion, where factors with an eigenvalue greater than 1 are retained, support for the proposed three dimension solution was found. This is further supported by the scree plot (Figure 7.3), where the inflection point is clearly seen to occur after the third factor. From Table 7.5, it also evident that the three-factor solution displayed good model fit, as RMSEA was less than .05, while CFI and TLI both exceeded .95 (Byrne, 2012; L. Hu & Bentler, 1998).

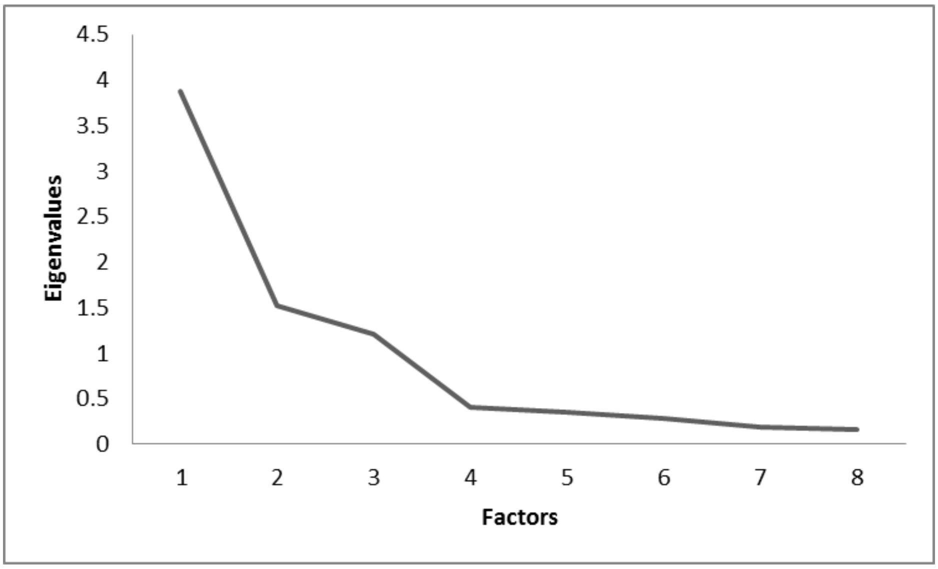

Figure 7.3: Scree plot depicting eigenvalues quality of care

Table 7.6: *Factor loadings for oblique rotated solutions with quality-of-care items*

Items	1 Quality of individual care	2 Number of errors	3 Quality of trust care
I am satisfied with the quality of care I give to patients/service users	.97		
I feel that my role makes a difference to patients/service users	.75		
I am able to deliver the patient care I aspire to	.83		
In the last month have you seen any errors, near misses or incidents that could have hurt staff?		.59	
In the last month have you seen any errors, near misses or incidents that could have hurt patients/service users?		.83	
Care of patients/service users is my organisations top priority			.90
My organisation acts on concerns raised by patients/service users			.91
If a friend or relative needed treatment, I would be happy with the standard of care provided by this organisation			.67

Note. n=7,033. Loadings less than .3 are suppressed.

Factor loadings. After geomin oblique rotation the eight items each adequately loaded (>.59) onto one of the three factors. Factor one ("quality of care provided by individual") comprised of three items, while two items loaded onto the second factor about "number of errors seen". The final factor, "quality of trust care", was made up of three items (Table 7.6).

Summary. EFA was carried out with eight items relating to quality of care. Examination of eigenvalues, model fit, and scree plots resulted in a proposed three-factor solution. The eight items load strongly onto one of three factors: quality of individual care (three items), number of errors seen (two items), and quality of trust care (three items). The first two types were identified within the systematic review in Chapter Four as important care outcomes. The third dimension, contains items which in previous studies have been used as proxies of quality of care (Pinder et al., 2013).

7.3 Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is run when there is some knowledge of the underlying latent variable structure (Byrne, 2012), which in this case is provided by the preceding EFAs. The proposed relationships between underlying factors and observed measures are specified a priori, and then tested statistically as to how well the model fits the data. This link between latent factors and their measured variables, is termed the measurement model (Byrne, 2012). CFAs are often misinterpreted to be rigorously confirmatory, when in reality a less restrictive approach encompassing adjusting and testing alternative models is used (Kline, 2016).

The aim of the analyses in the subsequent chapters is to employ a multilevel perspective. Although this may suggest that a multilevel CFA is more suitable here, this CFA was run at the individual-level. Considering that the EFA carried out in the section above was conducted at the individual-level, it is intended for this CFA to confirm the individual-level model developed above. Although both structural equational and multilevel modelling guidelines suggest carrying out a CFA to confirm factor structures as part of data preparation (Byrne, 2012; Heck & Thomas, 2015), confirming a multilevel structural model is a complex process with a high possibility of the proposed model not converging. Instead, first carrying out an EFA and then a CFA at the individual-levels would allow an initial less complex measurement model to be determined. Subsequently, more specific multilevel CFAs will be conducted prior to the studies in Chapter Nine and Ten. This would then allow for a more focused inclusion of the items and factors used within each study.

The subsequent section is structured around Kline's (2016) six steps in conducting a CFA or structural equation model (SEM): model specification, model identification, measure selection, assessing model fit, model re-specification, and reporting of results. Although these appear in a linear process, in practice CFAs/SEM are likely to be iterative as earlier steps are revisited to address issues discovered later on.

7.3.1 Pre-considerations

Prior to analysing the model three steps need to be considered. The first is to specify the model, then to ensure model identification, before selecting an appropriate estimator.

Model specification. This is the first and most important, step (Kline, 2016). All subsequent steps are based on the model specified here, which is based on the theoretical hypothesis developed and assumes this model to be true. Consideration should also be made towards how the theoretical or research literature could postulate adaptations to the model specified. This is particularly important should the model need to be re-specified later on.

Model identification. Identification pertains to whether or not the data is consistent with a unique set of parameters from the model specified (Byrne, 2012; Kline, 2016). This relates to the variance-covariance matrix of the observed variables and if it can be transposed to the structural parameters of the model. Should a unique estimate for each structural parameter be available, then the model is deemed identifiable and testable. However, should different parameter values define the same model then the model cannot be identified and evaluated empirically, and would need to be re-specified.

Models can be judged to be just-identified, under-identified, or over-identified (Byrne, 2012). *Just-identified* models occur when the number of parameters to be estimated is equal to the number of data variances and covariances. In other words, there is a one-to-one correspondence between the data and the structural parameters. However, this is of little utility as it allows for only one unique solution with no degrees of freedom, meaning it cannot be rejected. Similarly, an *under-identified* model is also of little use as the lack of data points (i.e., number of variance and covariances) in relation to parameters means an infinite number of solutions are possible. Researchers therefore seek to obtain an *over-identified* model, where the number of data points exceeds the number of estimable parameters. The positive degrees of freedom allows for the model to be rejected.

Model estimation. Once a model is identifiable, a suitable estimator needs to be selected. The different model estimators and their variants, including maximum likelihood (ML; Flora & Curran, 2004) and weighted least square (WLS; Browne, 1984), were described in Section 7.1.2 above. WLSMv is advocated for dichotomous data as it uses standard estimates drawn from polychoric, polyserial, and tetrachoric correlations (Browne, 1984; Hox et al., 2010). Therefore, the WLSMv estimator is more conducive to handle non-linear data as it provides a more conservative and robust approach in comparison to the other estimators (Asparouhov & Muthén, 2013; Hox et al., 2010; B. O. Muthén et al., 1997).

7.3.2 Model analysis

The next step examines the suitability of the model. Specifically, this involves assessing the fit of the model, the parameters estimated, and comparing against near-equivalent models (Byrne, 2012; Kline, 2016). A good fitting model is congruent with the sample data. Should this not occur then the measurement model needs to be rejected or re-specified.

Assessing model fit. A number of model fit indices exist to help determine the suitability of the measurement model for the sample data. These can be grouped into three classes: absolute, parsimony, and comparative (Byrne, 2012; Kline, 2016). Fit indices summarise multiple discrepancies into a single measure, meaning it represents the average model fit, or represents only a particular aspect of fit (Steiger, 2007). To prevent researchers from cherry-picking the best fit index, and to obtain a good overview of the model's fit, multiple measures of fit representing each class should be used (Kline, 2016).

Absolute fit indices assess how well the sample data is explained by the specified model (Kline, 2016). This relies entirely on the researcher's model and does not have any other reference point. Chi-Square Test of Model Fit is used to compare the null hypothesis against the alternative hypothesis (Brown, 2006). The null hypothesis predicts that all factor loadings, factor variance and co-variances, and residual variances are valid. Consequently, the probability value accompanying the model's chi-square (χ^2) statistic represents the likelihood that a χ^2 value exceeds the χ^2 when the null hypothesis is true. This means that a lower probability value indicates a poorer fit between the specified model and the perfect fit. As such, the aim is to not reject the null hypothesis, and therefore a *p*-value of >.05 is traditionally seen as desirable (Hair et al., 2014). However, in contemporary research χ^2 is not taken seriously and is sometimes even ignored altogether (Markland, 2007). This is due to its susceptibility to be inflated by sample size; and vulnerability to non-normal patterns of data. This poses a problem for the dataset in this analysis, which contains large sample sizes and non-normal data. Consequently, although χ^2 will still be reported, greater weight will be placed on more robust forms of fit indices that will now be discussed.

The second class are parsimony-adjusted indices, which encompass a correction for model complexity within their formulas and reward parsimonious models (Brown, 2006). The most commonly used index is the Root Mean Square Error of Approximation (RMSEA), which attempts to correct the tendency for χ^2 to reject models due to large sample sizes and observed variables by including both these aspects in its computations (Hair et al., 2014). Low values indicate good fit with zero representing perfect fit. Acceptable values of fit are RMSEA values lower than .05 (Brown, 2006). Hu and Bentler (1998) advocate reporting of RMSEA values as (i) RMSEA is sensitive to model misspecification; (ii) it is possible to generate RMSEA confidence intervals; and (iii) it typically reaches appropriate conclusion on model quality.

In contrast to the first two classes of fit indices, incremental (or comparative) indices compare the specified model against a baseline model. This baseline is the independence model, which assumes no correlation between variables. This has drawn criticism as some researchers argue that the measurement model is being compared against the worst possible model (Miles & Shevlin, 2007), and therefore has little actual significance. Two of the most common indices here are the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI; Byrne, 2012). The former is a normed index, while the latter is non-normed. This means that the CFI ranges between zero and one, with scores higher than .90 considered good fitting (Hair et al., 2014), although others recommend a more conservative >.95 (L. Hu & Bentler, 1998). Although the TLI could exceed a score of 1.0, the same cut-off scores as CFI are used to indicate good fit (Byrne, 2012). CFI and TLI are strongly correlated, and CFI is widely used due to its robustness to sample size and model complexity (Fan, Thompson, & Wang, 1999).

Assessing individual parameters. The indices examined thus far focus on the model as a whole. Should this be acceptable, then individual parameters can be examined for suitability and statistical significance (Byrne, 2012). The first part involves examining the estimated values to ensure that their size and direction are congruent with the underlying theory. Although the effect size should be guided by theory, rules-of-thumb recommend standardised loadings to be at least .5 and ideally .7 or higher (Hair et al., 2014). This demonstrates construct validity by giving confidence that the indicator loads strongly onto the construct. Examination of parameter estimates help identify negative variances or out-of-range-correlations (i.e., Heywood cases) that indicate issues with the model specified. Similarly, excessively large or small standard errors also suggest poor model fit (Byrne, 2012), as they respectively reflect parameters that cannot be defined and determined. Most CFA/SEM statistical packages generate statistical significance values for the parameter estimates by dividing the unstandardised parameter estimate by its standard error. Estimates that are not significant are typically unimportant to the model and are potential candidates for deletion (Byrne, 2012). However, this may also be an indicator of inadequate sample size.

Model comparison. Even though a model is deemed to fit, it does not necessarily mean the model is the best fitting one available (Hair et al., 2014; Kline, 2016). It is plausible that a different model specification could yield a better fitting model. Consequently, where possible, potential alternative models should be specified and compared against the hypothesised one. Fit indices allow for comparison between these different models. One model fit index not yet discussed is Akaike's Information Criterion (AIC), which is used to determine the best fitting model when a series of plausible models are specified (Raykov & Marcoulides, 2000). Unlike CFI and TLI, AIC uses non-nested models, with the model with the lowest value the best fitting. It accounts for model fit (χ^2) factoring in the degrees of freedom and number of estimated parameters.

7.3.3 Measure re-specification and reporting of results

Should the proposed model not be identifiable, or be a poor fit to the data, then there may be a need to revisit the initial model specified for amendments (Kline, 2016). CFA/SEMs provide additional information that aid diagnosis - modification indices (MI). This estimates the amount of χ^2 that would decrease if a specific fixed-to-zero parameter was freely estimated (Hair et al., 2014; Kline, 2016). A larger MI value would consequently provide a better predicted model should the pathway be added. There are no clear guidelines as to what constitutes a large or small MI value (Byrne, 2012; Kline, 2016). While it is possible to free up these pathways to improve model fit, modifications should be few, minor, and theoretically justifiable (Byrne, 2012; Kline, 2016).

It is not uncommon for initial CFA models to be rejected altogether (Kline, 2016). Any subsequent revision must be driven by theoretical logic and not statistical considerations, as the revised model needs to be congruent with the underlying theory and hypothesis. Once an acceptable model is obtained, then the final step is to accurately and completely describe the analysis process and results obtained.

7.4 Results: Confirmatory Factor Analysis

Measurement Model 1. One measurement model comprising of the thirteen composite measures extracted from the preceding EFA section was specified. The second subset of data from the random split (*n*=7,033 each) was used. This meant that the EFA and CFA were carried out in different datasets. The estimator used was WLSMv, to account for ordinal and nominal data. As per the job demands-resources model (Bakker & Demerouti, 2017; Demerouti et al., 2001), it was hypothesised that job demands would be a second-order factor consisting of *insufficient work resources* and *workplace aggression*. *Manager support, role clarity, effective team practices*, and *job control* were part of the second-order factor of job resources. Similarly, *work engagement*, *satisfaction with job content*, and *satisfaction with support* were theorised to form part of a second-order factor called positive work-related wellbeing. In total there were 42 dependent variables, plus an additional 16 continuous latent variables. Mplus was used for analysis.

The specified model was observed to have a positively defined latent variable covariance matrix. Inspection of the latent variable correlation matrix found an issue where one relationship had an *r* value greater than 1: job resources with positive work-related wellbeing (r=1.052). Such a value is theoretically impossible, and may represent an underlying issue with poor validity or highly correlated indicators (Hair et al., 2014). Closer inspection found both job satisfaction sub-measures (satisfaction with job content, satisfaction with support) to also have high correlations on both the second order-factors of motivation (respectively r=.989 & r=859) and job resources (r=.940 & r=.817). This suggests that job satisfaction items load onto other factors, which in itself is not surprising. Job satisfaction as a construct has been criticised for being too broad, not being clearly defined, and tapping into other domains (Briner, 2014; Judge & Klinger, 2008). Consequently, a second measurement was proposed excluding the eight job satisfaction items.

Measurement Model 2. The presence of a latent variable correlation value that exceeded one meant it was not possible to run the original measurement model. The eight items for job satisfaction were removed, leaving eleven measures remaining. Again, second-order factors for job demands and job resources were specified. The removal of job satisfaction meant the second-order factor of positive work-related wellbeing was now redundant. Instead, this was now represented solely by work engagement.

The overall fit of Measurement Model 2 was mixed. RMSEA was .06, which is higher than the recommended <0.05. CFI (.96) and TLI (.95) both exceeded the recommended level of >.95 suggesting good fit (Byrne, 2012; L. Hu & Bentler, 1998). Chi-square (χ^2=14943.59, df=500, p<.001) was significant, although this was not surprising given the large sample size used.

Factor loadings for the individual parameters were statistically significant, with standardised loadings for all items onto their latent variables exceeding the minimum of .5 (Hair et al., 2014). Subsequent examination of modification indices yielded additional information on how much of the χ^2 would be reduced by specifying a relationship where none had been specified (Hair et al., 2014). The modification indices table (Table 7.7) demonstrated that nine of the ten highest changes on χ^2 involved specifying additional pathways for role clarity items. These indices are sensitive to sample size, although the high standardised expected parameter change (StdYX EPC) values provide additional support that specifying a pathway for these items could lead to a better fitting model (Brown, 2006). Although specifying such a relationship is practically possible, there is no theoretical justification for doing so. To link role clarity items and other latent variables would also result in the cross-loading of items (Kline, 2016). At the latent variable level (Table 7.8), nearly all of the highest modification indices in both samples involve role clarity. While a rationale could be made to covary role clarity with other measures of job resources (e.g., manager support), there is no theoretical reason to covary role clarity with outcome measures (e.g., quality of care, work engagement) or second-order factors (e.g., job resources).

Table 7.7: *Highest item level modification indices in Measurement Model 2*

Item	loaded onto	MI	EPC	StdYX EPC
RC4	QC	4798.23	0.77	0.69
RC4	JD	2310.85	-3.87	-0.70
RC4	IR	1841.32	-0.70	-0.57
RC4	ER	1156.82	-0.47	-0.42
RC4	EG	629.00	0.37	0.34
RC2	JR	544.47	-1.03	-0.71
RC2	MS	529.02	-0.27	-0.25
RC4	IL	520.55	-0.40	-0.34
RC4	JR	491.22	1.05	0.72
TM2	JC	290.82	0.20	0.17

Table 7.8: *Highest latent variable level modification indices in Measurement Model 2*

	MI	EPC	StdYX EPC
Role clarity – Quality of care	3335.60	0.22	0.61
Job resources - Role clarity	2401.73	-0.15	-0.55
Role clarity - Manager support	1369.01	-0.12	-0.46
Quality of care - Manager support	987.04	-0.12	-0.21
Job control - Manager support	852.65	0.09	0.26

Measurement Model 2 was not a good fit to the data. Inspection of modification indices tables suggested that relationships could be specified which were neither theoretically sound nor practical. Consequently, the decision was made to remove the four role clarity items from this model.

Measurement Model 3. The removal of role clarity yielded a measurement model consisting of ten measures. Second-order factors of job demands (workplace aggression, insufficient work resources) and job resources (effective team practices, manager support, job control) were specified. The remaining measures were work engagement, negative work-related wellbeing, errors, quality of individual care, and quality of trust care.

As CFAs are reliant on the model specified by the researcher, it is plausible that a good fitting model is not the best possible fitting model (Byrne, 2012; Hair et al., 2014). To reduce the possibility of this occurring, the proposed measurement model was compared against three other models. The first relies solely on first-order factors and discounts the grouping of job demands and job resources. The second acknowledges recent developments in the JD-R literature where demands are not considered homogenous. This meant that only job resources existed as a second-order factor. Third, quality of care exists as a multifaceted concept, with definitions recognising this as encompassing perception of care and errors. Therefore, the third comparison model introduces a second-order factor for quality of care, which includes errors, quality of individual care, and quality of trust care.

Table 7.9: *Model fit for proposed and comparison models*

Model	χ^2	CFI	TLI	RMSEA
Proposed Model	6002.65***	.98	.98	.05
Comparison Model 1 – no second-order factors	5336.12***	.98	.98	.04
Comparison Model 2 – only job resources as a second-order factor	5084.84***	.99	.98	.04
Comparison Model 3 – quality of care as care as second-order factor	7326.37***	.98	.98	.05

Using the WLSMv estimator in Mplus means AIC is not available for model comparisons (L. K. Muthén & Muthén, 2017). Instead, Table 7.9 above compares the proposed model against the three alternative comparison models. The hypothesised measurement model, although good fitting, was not the best fitting model. Rather, both Comparison Models 1 and 2 were better fitting models. Therefore, in the interest of parsimony (Byrne, 2012), Comparison Model 1 without any second-order factors was instead accepted. The acceptance of this model is further supported through examination of individual parameters (Table 7.10). All individual items loaded adequately. Of the thirty items, three exceeded the minimum acceptable standardised loading of .5, with the remaining items meeting the higher threshold of .7 (Hair et al., 2014).

Table 7.10: *Standardised loading*

Latent Variable	Item	Estimate	S.E.	Est./S.E.	P-Value	R^2
Workplace aggression	AG1	0.57	0.03	19.34	***	0.32
	AG2	0.60	0.06	9.43	***	0.36
	AG3	0.87	0.04	21.79	***	0.75

Insufficient work resources	IR2	0.81	0.01	109.84	***	0.65
	IR3	0.80	0.01	102.14	***	0.64
Manager support	MS1	0.93	0.00	359.18	***	0.86
	MS2	0.93	0.00	373.16	***	0.86
	MS3	0.89	0.00	273.99	***	0.79
	MS4	0.88	0.00	250.41	***	0.78
	MS5	0.79	0.01	152.92	***	0.62
Job control	JC1	0.85	0.01	184.23	***	0.71
	JC2	0.91	0.00	278.63	***	0.82
	JC3	0.89	0.00	278.47	***	0.79
	JC4	0.90	0.00	264.23	***	0.81
Effective team practices	TM1	0.87	0.01	138.96	***	0.76
	TM2	0.84	0.01	126.46	***	0.71
	TM3	0.77	0.01	100.88	***	0.60
Work engagement	EG1	0.92	0.00	211.21	***	0.85
	EG2	0.92	0.01	200.80	***	0.84
	EG3	0.72	0.01	85.86	***	0.52
Negative work-related wellbeing	IL1	0.85	0.02	38.38	***	0.72
	IL2	0.55	0.02	29.84	***	0.30
Quality of individual care	QC1	0.89	0.01	181.43	***	0.80
	QC2	0.74	0.01	87.54	***	0.55
	QC3	0.94	0.01	174.27	***	0.88
Quality of organisational care	ORG1	0.88	0.01	193.90	***	0.77
	ORG2	0.86	0.01	179.12	***	0.75
	ORG3	0.86	0.01	156.41	***	0.74
Errors seen	ER1	0.89	0.02	37.43	***	0.80
	ER2	0.81	0.02	38.10	***	0.65

7.5 Reliability

The final step is to establish the reliability of the composite measure developed. This facilitates acceptance that measures are consistent, which is imperative in classical test theory (DeVellis, 2006).

Cronbach's alpha is widely used to assess internal consistency (Sheng & Sheng, 2012). It is only more recently that the suitability of such prevalent use has been queried with regard to non-normal data (Peters, 2014; Sheng & Sheng, 2012; Sijtsma, 2009). In particular, data that is skewed, leptokurtic, or binary in nature is incongruent with Cronbach's alpha, which relies on Pearson correlations that assume multivariate normality. Instead, omega reliability, which does not assume tau-equivalence, has been found to be a better indicator of reliability when assumptions of alpha are violated (Peters, 2014; Revelle, W., & Zinbarg, 2009). When assumptions are not violated, alpha and omega scores are identical. As the three items in the *workplace aggression* measure are positively skewed (i.e., doctors report no or few incidents of aggression), omega reliability is used for this measure. In the same way, *negative work-related wellbeing* and *errors* which used binary 'yes/no' response had internal consistency assessed by a

Kuder-Richardson 20 (KR20) test. The remaining measures were assessed using Cronbach's alpha.

Table 7.11 demonstrates that six of the ten measures exceed the recommended threshold of .8, indicating adequate consistency (Lance et al., 2006). Moreover, none of the measures have values exceeding .95, which could indicate item redundancy. However, low internal consistency is observed on four measures: *insufficient work resources* (Sample A: .72; B: .72), *workplace aggression* (.58; .57), *error* (.56; .58), and *negative work-related wellbeing* (.44; .46). Internal consistency is vulnerable to the number of items within measures, with fewer items associated with lower consistency (Eisinga, Grotenhuis, & Pelzer, 2013). This is potentially an issue for the latter measure, which comprised of only two items. The low number of items means it is not possible to examine how removal of items influences observed scores. Instead, following discussion with thesis supervisors it was decided that such a low internal reliability was not appropriate as a measure. Therefore, although the EFA and CFA suggested that the presenteeism and work-related stress items both loaded on a construct named negative work-related wellbeing, these items would instead be treated as two independent single items rather than having to utilise a measure consisting of two items with low internal consistency. The low reliability on the remaining four measures remains a limitation for the studies that utilise these measures.

Table 7.11: *Internal reliability scores for measures*

Measure	Sub-Sample A	Sub-Sample B
Workplace aggression	0.58	0.57
Insufficient work resources	0.72	0.72
Manager support	0.92	0.93
Job control	0.89	0.90
Effective team practices	0.81	0.81
Work engagement	0.84	0.84
Negative work-related wellbeing	0.44	0.46
Quality of individual care	0.80	0.81
Quality of organisational care	0.82	0.85
Errors seen	0.56	0.58

7.6 Common Method Bias

All the items from the NHS Staff Survey were collected at a single time point using the same self-report method, making it vulnerable to common method bias (Podsakoff et al., 2003). This occurs when the variance does not represent the construct being measured but rather the method used to measure it. It is a key source of measurement error and potentially provides an alternative explanation for any relationships observed, ultimately undermining the validity of the findings made. Podsakoff et al. (2003) outlined a series of procedural and statistical remedies to reduce the possible impact of common method bias. However, in the context of secondary data, there is no control over the data collection process, including: creating temporal, proximal, psychological, or methodological separation of measurement; counterbalancing question order; and improving measure items. Despite this, as presented in Section 6.2.2, the items in this survey are based on established questionnaires that should have considered the impact of phrasing and participant response. Moreover, the survey process emphasised protecting respondent anonymity and reducing evaluation apprehension, reducing the likelihood of participants purposively responding in a socially desirable manner.

Recognising the lack of procedural control over the data collection, two statistical procedures were carried out to examine whether common method bias was an issue within this data source. The first was Harman's single-factor test (Podsakoff et al., 2003). Traditionally, this involves loading all items onto an exploratory factor analysis. Should all the items load onto a single-factor, or if one general factor accounts for the majority of the variance then it is likely that a significant level of common method bias existed. Additional un-rotated EFAs indicate that items did not load onto a single-factor and that even when a single-factor was specified this factor only accounted for 22.22% and 22.38% of the variance in sub-samples A and B respectively. Harman's test can also be carried out using CFA (Podsakoff et al., 2003), where all items are specified onto one factor. Doing so with this dataset indicated a poor fitting model (Sample A: RMSEA=.17, CFI=.75, TLI=.73; Sample B: RMSEA=.17, CFI=.75, TLI=.73), further suggesting that the proposed measurement model in Table 7.9 was better fitting.

The results indicate that common method bias was not likely to be a significant issue. However, Podsakoff and colleagues (2003) argued that Harman's test is an insensitive test as it is unlikely that multiple factors would appear in an EFA. Ideally, a marker variable technique should be used (Lindell & Whitney, 2001). This involves examining whether relationships can be observed between items that should, theoretically, not be correlated with each other. If such relationships did exist then it is likely that they are inflated by common method variance. However, as all items within the NHS Staff Survey pertained to working conditions, wellbeing, organisational functioning, and performance, none of the excluded items could function as a marker variable.

Instead, a single unmeasured latent factor was modelled and controlled for within the CFA (Podsakoff et al., 2003). Although this does not allow for the identification of specific sources of method bias, it controls for systematic variance among items. This yielded a well-fitting model (Sample A: RMSEA=.04, CFI=.98, TLI=.98; Sample B: RMSEA=.04, CFI=.98, TLI=.98) that was similar to the final proposed measurement model. Squaring the unstandardised regression coefficients (Sample A: .41; Sample B: .36) between the items and the latent factor demonstrated that common method variance accounted for 16.4% and 13.2% of the total variance. As an additional step, the differences between the standardised regressions from the models with and without the common latent factor were compared. Only two items (AG3, IL1) in Sample A had a large difference (>.200), suggesting that it may be affected by common method variance (Gaskin, 2017). This difference was not observed in any of the remaining items in either sample. Therefore, although it is not possible to completely rule out common method bias influencing subsequent analyses, the findings from the Harman test and the single common latent factor test suggest common method variance is not likely to be a significant issue. In addition, to address any potential influence of common method bias, where possible additional data sources reflecting demands placed on trusts and quality-of-care outcome measures will be used.

7.7 Conclusion

The reliance of this thesis on the NHS Staff Survey as a secondary data source means there is no possibility of influencing the constructs being assessed and the items used to do so. Therefore, the purpose of this chapter was to develop psychometrically suitable measures from the 2014 NHS Staff Survey, which will in turn inform subsequent analysis of the psychosocial working conditions experienced by hospital doctors in the NHS. Through an exploratory and confirmatory factor analysis respectively carried out in two separate samples of 7,033 hospital doctors each, a measurement model with eleven constructs was identified.

Chapter 8 : Job Demands and Resources as Predictors of Hospital Doctors' Work-related Wellbeing (Study 2)

Chapter Two reviewed a plethora of studies linking doctors' psychosocial working conditions with their work-related wellbeing (e.g., Lee et al., 2013; Scheurer et al., 2009). However, it clearly identified the absence of testing these from a theoretical perspective. In line with the aim of this thesis, this study frames the relationship between these two constructs within the job demands-resources (JD-R) model. More specifically, it aims to examine the direct effect that job demands, job resources, and trust-level demands have on doctors' work-related wellbeing, and whether job resources interact with job or trust-level demands.

This chapter first presents the study hypotheses and methods. This is followed by an introduction into multilevel modelling, including model and data considerations prior to building a multilevel model, and the process of testing multilevel models. The results are then presented before discussing these in the context of this study and the thesis.

8.1 Introduction

Drawing upon the propositions reviewed in Chapter Five where the JD-R model was introduced, this chapter comprises of a study that poses to answer three research questions.

The first asks whether the dual process proposed by the JD-R model results in job demands being a better predictor of negative work-related wellbeing (i.e., strain) than job resources (Bakker & Demerouti, 2007; Demerouti et al., 2001); and if in turn, doctors' job resources will better predict positive work-related wellbeing (i.e., motivation) than job demands. Theoretically the JD-R model postulates that both these processes are independent, although the research evidence provides some indication of cross-effects (Hakanen et al., 2008; Nahrgang et al., 2011). To date there have been limited attempts to frame these relationships in the research involving doctors within the JD-R model (Prins, Hoekstra-Weebers, et al., 2007); this study attempts to do this by testing the following two hypotheses:

> H_1: Hospital doctors' work-related stress and presenteeism will be more strongly predicted by their job demands than their job resources; and
>
> H_2: Hospital doctors' work engagement will be more strongly predicted by their job resources than by their job demands.

The second research question asks what happens to hospital doctors' work-related wellbeing when job demands and resources interact. This draws upon two different propositions within the JD-R model (Bakker & Demerouti, 2017): (i) that job resources buffer the negative influence job demands have on negative work-related wellbeing (e.g., work-related stress, presenteeism); and (ii), that high job resources paired with job demands are associated with work engagement. The evidence for this remains somewhat mixed (Bakker et al., 2007; Xanthopoulou, Bakker, Dollard, et al., 2007), including one study of Dutch medical residents where less than half of potential interactions predicted conflict between home and work (Prins, Hoekstra-Weebers, et al., 2007).

> H_3: Job resources will moderate the relationship between job demands with work-related stress and presenteeism. More specifically, this relationship will be weaker for hospital doctors who experience high levels of job resources than those who experience low-levels of job resources.

The JD-R model proposes that a moderate level of job demands can increase the saliency of job resources' relationship with work engagement (Bakker & Demerouti, 2017). However, one explanation for the inconsistency of evidence supporting this relationship is due to the type of job demand being examined. There is increasing evidence that job demands are not one homogenous construct but rather exist as challenge and hindrance demands (Crawford et al., 2010; LePine et al., 2005); challenge demands stimulate work engagement while hindrance stressors are detrimental to it. The former promotes growth and mastery whilst the latter thwarts it. In terms of the two job demands identified in Chapter Seven, workplace aggression clearly is a hindrance demand. For insufficient work resources the distinction is not clear; it could be perceived as hindering doctors' progress but may also allow the development of mastery. Additional responsibilities and increased workload are both characteristics of challenge demands (Crawford et al., 2010), and are likely to occur when hospital doctors perceive a lack of staffing and material resources. Therefore, it is postulated that insufficient work resources functions as a challenge demand. Consequently, this study distinguishes between workplace aggression and insufficient work resources to predict that:

> H_4: High job resources paired with high insufficient work resources have a stronger relationship with hospital doctors' work engagement than when insufficient work resources are low.

Finally, from a multilevel perspective the nested nature of hospital doctors within trusts means understanding how events at the trust-level impact on individual doctors will further understanding of a systems perspective of workplace wellbeing (Lowe & Chan, 2010). While the JD-R model has been extrapolated to the organisational (or unit)-level, most of this tentative examination has focused on shared-perceptions of burnout or work engagement in relation to performance outcomes (González-Morales et al., 2012; Torrente et al., 2012) and less on demands and resources (Westman et al., 2011). The growing concern on the condition of the NHS has focused attention on trust-level demands.

The final question this study therefore asks is whether trust-level demands have the same impact within the JD-R model as that of doctors' job demands (Schaufeli & Taris, 2014). Past research has shown that mergers, senior leadership support, and communication are among the organisational-level factors that have been found to relate with job satisfaction and work-related stress of healthcare workers in the NHS (Powell et al., 2014). Elsewhere, there is evidence that organisational-level predictors influence the strength of the relationship between job demands and resources with work-related wellbeing (M. K. Tucker, Jimmieson, & Oei, 2013). For example, individuals in groups with strong consensus of their leadership reported weaker relationships between job demands and depression, than groups with a weak consensus towards their leaders (Bliese & Britt, 2001). Therefore, using two common proxies for trust-level demands (Maben, Peccei, Adams, & Robert, 2012; Madsen, Ladelund, & Linneberg, 2014) - the number of emergency admissions and bed occupancy rates, it is predicted that:

> H_5: High trust-level demands would positively predict doctors' work-related stress and presenteeism, but not predict doctors' work engagement.

> H_6: High trust demands will moderate the individual-level relationships between job resources and work engagement. More specifically, this relationship will be stronger when trust-level demands are high than when trust-level demands are low.

8.2 Method

8.2.1 Sample

The study sample is described in depth in Section 6.2.1 and was made up of 14,066 doctors from 157 acute trusts. Mean doctors per trust was 89.59 (*SD*=94.76) with a median of 41 doctors.

8.2.1 Materials

Individual-level measures. Measures of doctors' job demands, job resources, work-related stress, presenteeism, and work engagement were drawn from the 2014 NHS Staff Survey. The process of developing composite measures was described extensively in Chapter Seven. Job demands consisted of two different measures: a two item measure of insufficient work resources and three items measuring workplace aggression. Job resources consisted of effective team practices (three items), manager support (five items), and job control (four items). Work-related stress and presenteeism were each measured by one item while work engagement was a three item measure.

Trust-level demands. Two measures were used as proxies for trust demands. The first was bed occupancy rates from NHS England (2015a). This represented the average overnight bed occupancy rates within the trust between October and December 2014. The second was the number of emergency admissions to the trust between October to December 2014 (NHS England, 2015c). For both variables a higher value represented more demands placed onto the trust.

Control variables. Previous studies have demonstrated doctors' seniority to influence their working conditions and quality-of-care outcomes (Dornan et al., 2010; Prins, Hoekstra-Weebers, et al., 2007; Ryan et al., 2014). Consequently, at the individual-level, doctors' organisational tenure was included as a proxy for seniority. Responses on this item were on a five-point scale representing "less than a year", "1-2 years", "3-5 years", "6-10 years", "11-15 years", and "more than 15 years".

At the trust-level, two variables were controlled. First was whether or not the trust was a specialist trust, as specialist trusts have a more narrow focus and are typically better resourced than their non-specialist counterparts. The second control factor was the size of the trust, which was represented by the number of beds available (NHS England, 2015a). Both these factors have been controlled for in similar studies (Admasachew & Dawson, 2011; Lim, 2014; Powell et al., 2014).

8.3 Analysis

This section introduces multilevel modelling, and reviews model and data considerations prior to building a multilevel model. It then briefly reviews the process of testing multilevel models (Byrne, 2012; Heck & Thomas, 2015; Stride, 2016).

8.3.1 Multilevel modelling

The reality of organisational research is that individual workers are typically nested within teams or departments, which are then situated within organisations, who in turn operate within specific sectors and/or geographical locations (Byrne, 2012; Croon & van Veldhoven, 2007). This nesting of people within groups means that they are frequently exposed to similar environments that with socialisation over time results in them becoming more alike to their in-group, and having less in common with those from other groups (Croon & van Veldhoven, 2007). Traditional statistical tests, including regressions and structural equation modelling, do not account for this nesting. This actually violates these tests' assumption that requires data to be independent of each other (Sjetne, Veenstra, & Stavem, 2007).

When studies avoid a multilevel perspective, they focus only on the lowest (i.e., individual) or the highest (group) level of measurement. This means data from the individual-level has to be aggregated to the group-level, or data from the group-level has to be disaggregated down to the individual-level (Byrne, 2012; Heck & Thomas, 2015). Aggregating data reduces statistical power as inadvertently fewer units of analysis exist at the group-level than at the individual-level (Hox et al., 2010). Also, as aggregation results in a single mean for each group it fails to account for any of the variation from the individual-level (Duncan, Jones, & Moon, 1996) and loses out on capturing individual behaviour and perceptions (Kozlowski & Klein, 2000). In contrast, disaggregating data to the individual-level discounts the role of the team, organisation, or larger culture that the individual operates in (Duncan et al., 1996; Kozlowski & Klein, 2000). Crucially, when a measure at the higher level (e.g., team sales) is disaggregated to the individual-level then everyone in the same group will have identical scores on that measure. This inflates standard error scores, parameter estimates, and significant tests which increases the probability for Type I errors (B. O. Muthén & Satorra, 1995).

It is therefore imperative that variables need to be measured clearly at the appropriate level and that they are matched with variables of a similar level. If not, this violates the compatibility principle which requires all variables within a model to operate at the same level of specificity (Ajzen, 2005). A multilevel perspective counteracts this by allowing interactions between individual and organisational/systems perspectivse (Ryu, 2015). More specifically, it summarises variability at the higher (i.e., between-group) level, as well as within-group variability at the lower individual-level (Byrne, 2012; Heck & Thomas, 2015). The lowest level of measurement, typically the individual, is referred to as Level-1, while the next level up is Level-2 and so forth onwards. Each individual's total score is thereby separated into an individual (or within) component and a group (or between) component. The former examines the individual's deviation from the group average while the latter encompasses the disaggregated group mean.

8.3.2 Model and data consideration

Prior to carrying out the proposed multilevel analysis, the implications of the estimator, sample size, and centering had to be considered. This forms the basis of the section below.

Estimator. Unlike the confirmatory factor analyses in the previous chapter, this study used observed variables rather than latent factors. With such variables, maximum likelihood (ML) estimation handles the uncertainty in observed data. More specifically, a likelihood function derived from the underlying sampling distribution of the outcomes (e.g., normal, binomial) is used to estimate the optimal values for the unknown parameters in a proposed model (e.g., regression parameters, means, variance; Heck & Thomas, 2015). In large samples, such as this study's sample, the ML estimator is robust to departures from normality and missing data, and generates asymptotically consistent and efficient estimates (Hox, 2010). Even when used in less-than-ideal situations (e.g., non-multivariate normality or with small sample sizes), the ML estimator still produces reasonable estimates. This extends to include models with sampling distributions that are not normally distributed, such as binomial, multinomial, Poisson, or negative binomial (L. K. Muthén & Muthén, 2017). In situations where outcomes (i.e., work-related stress and presenteeism) are dichotomous, ML uses other model specification and estimation procedures to take into account the outcomes' underlying probability distribution and scale. A final advantage with the ML estimator is its allowance to compare chi-square goodness-of-fit estimates to compare different specified models (Heck & Thomas, 2015). Therefore, considering these advantages and the presence of dichotomous outcome measures, the ML estimator is most appropriate for the observed variables used in this study's analyses.

Sample size. Multilevel analyses' relationship with data is not as clear cut as it is with single-level analyses, e.g., structural equation modelling. Factors such as the number of Level-2 units, examination of interactions, and estimation of random slopes add complexity to the data (Hox et al., 2010; T. A. B. Snijders, 2005). Moreover, it believed that a large number of published multilevel studies may have fallen victim of Type II errors due to insufficient statistical power (Heck & Thomas, 2015). Considering between-group analyses occurs at Level-2, it is imperative that sample size is not only sufficient at the individual-level but at the group-level as well. Based on simulation studies, multilevel models with small sample sizes at Level-2 (i.e., n<50) are vulnerable to bias estimates of Level-2 standard errors (Maas & Hox, 2005; T. A. B. Snijders, 2005). In this study, the surveying of every acute English NHS Trust (n=157) means this should not be an issue.

Centering. Centering refers to the rescaling of predictor variables so that the intercept can be better interpreted (Heck & Thomas, 2015; Stride, 2016). This provides the expected value of an outcome when the covariate is equal to some designated value of theoretical interest. From a multilevel perspective mean-centering is most commonly used, although other approaches exist too (e.g., median-centering). More specifically, group-mean centering occurs where each observation is measured in terms of its difference from the mean of other observations within its own group. In contrast, grand-mean centering is where the grand-mean is subtracted from individuals' score. For example, group-mean centering focuses on the difference between a participant's score and their team's mean, while in grand-mean centering the difference is in relation to the mean of all participants (Heck & Thomas, 2015).

Both approaches are commonly used in multiple regression analysis; however, some additional consideration is required for multilevel analyses (Enders & Tofighi, 2007). Group-mean centering removes group-level differences from the predictor variable, thereby removing its ability to predict between-group (i.e., Level-2) variance (Stride, 2016). This reduces conflation between the within and between-part variance which improves interpretation. Hence, when the effect of X on Y is hypothesised to occur only at the individual-level (i.e., Level-1), group-mean centering is almost always essential. At between-group level, group-mean centering is redundant as each participant has the same value on the Level-2 predictor (Enders & Tofighi, 2007). Instead, grand-mean centering is typically used. Recognising this, the use of centering in this study means that predictors at the individual-level were group-mean centered while trust-level predictors were grand-mean centered.

8.3.3 Testing the proposed multilevel model

Testing the hypotheses involved building a series of models (Heck & Thomas, 2015; Stride, 2016). Similar to hierarchical regression, the complexity of the analysis increases as additional predictors are included with each subsequent model. For the purpose of this study six models were required. The first was the unconditional model (M0), followed by models with the control variables (M1), individual-level predictors (M2), trust-level predictors (M3), interactions at the individual-level (M4a), and the prediction of individual-level slopes by trust-level predictors (M4b).

The unconditional model. The first model determines the level of variance that is already accounted prior to the inclusion of any predictors (Stride, 2016). It serves to partition the variance of an outcome into its within and between-group components (Heck & Thomas, 2015). The unconditional model also provides the base chi-square statistics which will be subsequently used to assess whether model improvement occurs. For this study, a single model including all three outcome measures: work-related stress, presenteeism, and work engagement, was specified.

If variance at the higher-level (i.e., between-groups) is minor or non-existent it negates the need for multilevel analysis. Intraclass correlation (ICC), represented by the Greek letter ϱ, is used to determine the amount of variance due to between-group variation through the equation $\varrho=\sigma^2_b /(\sigma^2_b + \sigma^2_W)$; where σ^2_b is the between-group variance and σ^2_W the within-group variance. Non-trivial ICC levels indicate that single-level analysis' assumptions of independence are violated (Heck & Thomas, 2015); this results in biased reduced standard errors rate that increases the probably of making a Type I error.

Hox (2002) wrote that ICCs under 0.05, or less than 5% of the grouping variable variance, can be ignored. However, this rule-of-thumb has also been considered redundant. Instead, this should be guided by the size of the Level-2 sample, the outcome being examined, and the practice of other researchers in related fields (Heck & Thomas, 2015). In contrast, others argued that as long as the data presents in clusters then multilevel modelling should be used (Nezlek, 2008; Stride, 2016). Although a particular outcome measure may display little between-group variance, it does not mean that no variation occurred in the relationship between the outcome measure and other measures in the model.

A more robust alternative is to calculate the design effect (*deff*) which accounts for ICC in relation to average cluster size (B. O. Muthén & Satorra, 1995). This is particularly relevant when the number of individuals per cluster is large, as small ICCs can still impact significant testing (Barcikowski, 1981). Design effect can be calculated through the formula: $deff = 1+(s-1)r$, where s is the average cluster size and r is the ICC. *deff* values over 2 suggest that taking clustering into account is necessary.

Adding models. The inclusion of new variables within a model allows the examination of whether there is a decrease in residual variance and improvement in the model (Hox, 2010; Stride, 2016). The order of the models is dependent on the hypotheses made. Upon the inclusion of each model, both within-trust and between-trust residual variance would be examined. The former is expected to decrease when individual-level predictors are included, while the latter decreases when trust-level predictors are included. Model improvement is then assessed through changes in deviance between the loglikelihood of the simpler versus the more complex model (Stride, 2016).

8.4 Results

8.4.1 Descriptive results

Table 8.1 presents the descriptive statistics and correlations for all study variables, which were obtained using SPSS. Work engagement correlated negatively with work-related stress (r=-.60) and presenteeism (r=-.36), while work-related stress and presenteeism correlated positively with each other (r=.24). At both the individual and trust-level, all 28 correlations were significant at p<.01. At the trust-level, the number of emergency admissions correlated with manager support (r=-.27), effective team practices (r=-.24), job control (r=-.32), work engagement (r=.-.21), and insufficient work resources (r=.42). Bed occupancy correlated with effective team practices (r=-.16), workplace aggression (r=.16), and insufficient work resources (r=.23).

Table 8.1: *Descriptive statistics and correlations*

Variable	*N*	*M*	*SD*	1	2	3	4	5	6	7	8
1. Workplace aggression	13700	3.61	1.09	-	.19**	-.10**	-.08**	-.11**	.21**	.18**	-.16**
2. Insufficient work resources	13890	6.09	2.00	.42**	-	-.40**	-.28**	-.41**	.29**	.21**	-.40**
3. Manager support	13563	17.64	4.63	-.21**	-.58**	-	.41**	.57**	-.26**	-.18**	.42**
4. Effective team practices	13581	11.51	2.53	-.25**	-.43**	.50**	-	.47**	-.17**	-.13**	.34**
5. Job control	13832	14.44	3.67	-.29**	-.61**	.65**	.61**	-	-.26**	-.18**	.51**
6. Work-related stress	13807	0.33	0.47	.34**	.56**	-.47**	-.41**	-.38**	-	.29**	-.36**
7. Presenteeism	12138	0.51	0.50	.32**	.26**	-.24**	-.33**	-.26**	.24**	-	-.21**
8. Work engagement	13907	11.82	2.38	-.35**	-.55**	.60**	.40**	.52**	-.60**	-.36**	-
9. Bed occupancy	157	88.03	9.07	.16*	.23**	-.11	-.16*	-.12	.13	.02	-.06
10. Emergency admissions	150	689.32	369.80	.15	.30**	-.27**	-.24**	-.32**	.05	-.07	-.21**

Note. ***p*<.01; **p*<.05. Correlations above the diagonal are individual-level correlations. Correlations below the diagonal are trust-level correlations, with individual-level measures aggregated to the trust-level (*N*=157).

Table 8.2: *Model fit statistics*

	-2LL	Deviance, *df* change	Sig.	Within-trust work-related stress variance	Between-trust work-related stress variance	Within-trust presenteeism variance	Between-trust presenteeism variance	Within-trust work engagement variance	Between-trust work engagement variance
M0. Unconditional model	96380	n/a	n/a	0.22	0.002	0.25	0.002	5.60	0.079
M1. Fixed effects of control	94708	1672, 15	*p*<.001	0.22	0.001	0.25	0.002	5.60	0.058
M2. Fixed effect of individual-level predictors	79769	14938, 15	*p*<.001	0.19	0.001	0.23	0.003	3.76	0.086
M3. Fixed effect of trust-level predictors	78070	4208, 6	*p*<.001	0.19	0.001	0.23	0.002	3.75	0.082
M4a. Moderations at individual-level	75542	20, 18	*p*>.05	0.19	0.001	0.23	0.002	3.75	0.082
M4b. Random effect and cross-level interactions	78062	8, 14	*p*>.05	0.19	0.001	0.23	0.002	3.75	0.082

Table 8.3: *Standardised coefficients for predictors in Models 1-3*

	Work-related Stress			Presenteeism			Work Engagement		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
Tenure (w)	.06***	.03**	.03**	.05***	.03***	.03***	-.07***	-.02**	-.02*
Specialist (b)	-.36*	-.36*	-.32*	-.17	-.16	.20	.41***	.41***	.40**
Beds (b)	.12	.08	.02	-.21	.21	-.71**	-.21	-.10	-.32
Insufficient work resources (w)		.16***	.16***		.12***	.12***		-.19***	-.19***
Workplace aggression (w)		.16***	.16***		.15***	.15***		-.07***	-.07***
Manager support (w)		-.10***	-.10***		-.06***	-.06***		.11***	.11***
Effective team practices (w)		-.01	-.02		-.03**	-.03**		.09***	.09***
Job control (w)		-.10***	-.10***		-.06***	-.06***		.32***	.32***
Bed occupancy (b)			-.24			-.03			.31*
Emergency admissions (b)			.10			.56*			-.50*

Note. ***p<.001, **p<.01, *p<.01; (b) = trust-level predictor; (w) = individual-level predictor.

8.4.2 The unconditional model

Table 8.2 presents the model statistics for the six specified models on the three dependent variables: work-related stress, presenteeism, and work engagement. The unconditional model (M0) indicated low ICC for work-related stress (0.008), presenteeism (0.006), and work engagement (0.014). However, when these were converted to *deff* scores their respective values were 1.79, 1.71, and 2.23, suggesting that clustering could be taken into account.

8.4.3 Fixed effects for control variables

Control variables at the individual (tenure) and trust (trust type, number of beds) level were added to M1. From Table 8.2 it is evident these additions reduced model deviance ($\chi^2(15)=14938$; $p<.001$). However, the control variables had differing effects on the three dependent variables (Table 8.3). Although the relationships that tenure had with work-related stress ($\beta=.06$), presenteeism ($\beta=.05$), and work engagement ($\beta=-.07$) were significant, the effect sizes were negligible. Doctors working at non-specialist trusts were more likely to report experiencing work-related stress ($\beta=-.36$) and less work engagement ($\beta=.41$) than doctors in specialist trusts.

8.4.4 Fixed effects for individual-level predictors

To test H_1 and H_2 the individual-level predictors of insufficient work resources, workplace aggression, job control, manager support, and effective team practices were added to the first model. Table 8.2 shows that model deviance decreased significantly ($\chi^2(15)=14938$; $p<.001$). M2 demonstrated that both job demands positively predicted work-related stress and presenteeism, and negatively predicted work engagement (Table 8.3). The opposite effects were observed when work engagement was the outcome, with the exception of effective team practices on work-related stress which reported no significant relationship ($p>.05$). By examining the standardised coefficients, insufficient work resources and workplace aggression were stronger predictors of work-related stress and presenteeism than any of the three job resources in M2, supporting H_1. Table 8.3 also demonstrates that job control ($\beta=.32$), manager support ($\beta=.11$), and effective team practices ($\beta=.09$) predicted work engagement. Workplace aggression had a negligible relationship with work engagement. However, H_2 was not completely supported as insufficient work resources ($\beta=-.19$) was a stronger predictor of work engagement than effective team practices and manager support.

8.4.5 Fixed effects for between-trust predictors

The addition of number of emergency admissions and bed occupancy rate as trust-level demands in M3 (Table 8.2) resulted in a significant change in model statistics ($\chi^2(6)=4208$; $p<.001$). The number of emergency admissions at the trust-level predicted doctors' presenteeism ($\beta=.56$) and work engagement ($\beta=-.50$), but not work-related stress (Table 8.3). Surprisingly, trusts with more of their beds occupied correlated with higher doctor work engagement ($\beta=.31$). No support was found for H_5.

8.4.6 Moderation between job demands and resources

To test H_3 and H_4, six new interaction terms representing each job demand and resource interaction were added to M4a as predictors. However, Table 8.2 shows that these did not result in M4a having a better fit compared to M3 ($\chi^2(18)=20$; $p>.05$). Consequently, there was no evidence that job resources moderates the effect of job demands on work-related stress, presenteeism, and work engagement as predicted by H_3 and H_4.

The final model (M4b) specified a random slope for the significant relationships from M2 and M3. More specifically, this was between job control, manager support, and effective team practices with presenteeism and work engagement. The number of emergency admissions was modelled to predict the random slopes where presenteeism was the outcome variable, while emergency admission and bed occupancy predicted the random slopes where work engagement was the outcome variable. As seen in Table 8.2, these additions did not result in an improved model fit ($\chi^2(14)=8$; $p>.05$), meaning that H_6 was rejected as no cross-level interactions between trust-level demands and doctor-level job resources were observed.

8.5 Discussion

The present study aimed to test the predictive associations of job demands, job resources, and trust-level demands in relation to three work-related wellbeing measures (work-related stress, presenteeism, work engagement) in a sample of doctors from English hospitals. Support was obtained for the JD-R model's proposition that job demands (workplace aggression, insufficient work resources) predicted work-related stress and presenteeism, while job resources (job control, manager support) predicted work engagement. No interactions were observed between any of the job demands and resources in relation to the three wellbeing measures. From a multilevel perspective, both trust-level demands (bed occupancy rate, number of emergency admissions) had minimal impact on work-related stress and presenteeism. Both trust-level demands did predict work engagement, which was not predicted. Finally, no evidence of cross-level interactions involving job resources and trust-level demands were found. Despite the majority of hypotheses being rejected, this study makes several contributions to understanding hospital doctors' work-related wellbeing and the field of occupational health psychology. The discussion below first considers the validity of JD-R model among hospital doctors before reflecting on the multilevel implications. The implications for hospital doctors' work-related wellbeing as a construct are then considered, followed by the study's limitations and conclusions.

8.5.1 The validity of the JD-R model among doctors

To date the main propositions of the JD-R model have received little attention among samples of doctors; although, it has been tested with nurses (Montgomery et al., 2015), dentists (Hakanen et al., 2005), and care home workers (Xanthopoulou, Bakker, Dollard, et al., 2007). The results provide contemporary support on the dual process component of the JD-R model (Demerouti et al., 2001; Hakanen et al., 2008), where job demands are associated with work-related stress and presenteeism, while work engagement is primarily associated with job resources. Not only does this support the diversity of constructs that reflect doctors' wellbeing, but suggests that interventions to address and manage work-related wellbeing among hospital doctors should target both job demands and resources.

Not one of the hypothesised individual-level interactions between job demands and resources were found. One explanation for this inconsistency lies in the type of job demand being examined. Recognising the difference between challenge and hindrance demands (LePine et al., 2005), the former stimulates work engagement while the latter is detrimental to it. In the context of this study, both workplace aggression and insufficient work resources correlated negatively with work engagement, suggesting that they may in fact impede any form of mastery or growth. This, in turn, means that both job demands are unlikely to foster work engagement even when doctors are presented with appropriate job resources.

The findings for workplace aggression, nevertheless, contradict the research where job resources interacted with similar job demands, including patient harassment in homecare workers (Xanthopoulou, Bakker, Dollard, et al., 2007) and pupil misbehaviour with teachers (Bakker et al., 2007) to predict work engagement. However, the JD-R model suggests that job resources should appropriately match the job demand it interacts with (Bakker & Demerouti, 2017). Physical and verbal abuse has been found to lead to significant emotional distress among healthcare workers, and that support from colleagues or managers are inadequate in addressing the issue (Henderson, 2003; Pellico, Brewer, & Kovner, 2009). Therefore, it may be that the detrimental impact of workplace aggression and insufficient work resources is such that, none of the three job resources tested in this study adequately mitigate job demands' impact on doctors' work-related stress or presenteeism.

8.5.2 A multilevel perspective of the JD-R model

The present study also empirically contributes to the JD-R model and the wider understanding of the work-related predictor-wellbeing relationship by integrating a multilevel perspective to the model. Contrary to expectations, bed occupancy rates did not predict work-related stress or presenteeism, while emergency admissions only predicted presenteeism. Moreover, despite hypothesising that trust-level predictors will not relate with work engagement, this relationship was found. These findings counter the dual process pathways suggested by the JD-R model at the individual-level (Demerouti et al., 2001), and lend support to the notion that the model may operate differently across different levels (Bakker & Demerouti, 2017). This is seen where unit-level job resources exacerbated burnout among agency workers (Westman et al., 2011), or where department-level teamwork effectiveness did not have the anticipated correlation with nurse burnout (Montgomery et al., 2015). This could also explain why neither of the trust-level demands interacted with the relationships that job resources had with the three work-related wellbeing outcomes. This is despite the evidence drawn from individual-level designs predicting its occurrence (Bakker et al., 2006).

Interestingly, the different directions in the relationship between both trust-level predictors with work engagement reinforces the distinction between challenge and hindrance job demands (LePine et al., 2005). It is plausible that bed occupancy rate is an example of a challenge demand; more specifically, as long as spare beds remain then the resources exist to cope with the demand faced (Madsen et al., 2014). Instead, it is only when operating at or exceeding capacity that this could become a hindrance demand. In contrast, emergency admissions are unplanned, non-routine, and often complex, with the potential to interfere with other tasks at hand (Lawrence et al., 2016; Maben et al., 2012). Consequently, a high number of emergency admissions functions as a hindrance demand, with detrimental relationships to work engagement and presenteeism.

These findings also raise questions as to what other trust-level demands and resources impact doctors' work-related wellbeing. For example, senior leadership support and trust restructuring has been found to predict healthcare workers' job satisfaction and health (Lim, 2014; Powell et al., 2014). Alternatively, these factors could operate as moderators; Tucker, Jimmieson, and Oei (2013) found that job control mitigated the negative impact of job demands on anxiety in groups with high collective self-efficacy, but exacerbated this relationship when groups' collective self-efficacy was low.

A further question lies in whether job demands and resources at the individual-level can be aggregated to the trust-level to reflect a shared group experience. For example, Torrente and colleagues (2012) used the term "my team" instead of "I" to provide a referent shift from the individual to the team. Their findings were congruent with the JD-R model, where high team-level social resources predicted high team engagement and better team performance. Other studies have moved the referent to nursing department-level teamwork (Montgomery et al., 2015) and social workers' team resources (Busch et al., 2013), with mixed effects on wellbeing. Although the items in the present study did not allow job demands and resources to be aggregated to the trust-level, future research should consider whether such aggregation makes conceptual sense to further a multilevel understanding of the JD-R model (Bakker & Demerouti, 2017).

8.5.3 Expanding hospital doctors' work-related wellbeing beyond burnout

In addition the testing the validity of the JD-R model, by including work-related stress, presenteeism, and work engagement, this study expands the research into doctors' work-related wellbeing beyond the commonly used burnout (Prins, Gazendam-Donofrio, et al., 2007). All three measures are important in their own right: work-related stress as a traditional strain measure of affect; presenteeism, as a behavioural indicator (Admasachew & Dawson, 2011); and work engagement, representing a positive motivational state (Schaufeli & Bakker, 2004). The results support the proposition that negative (i.e., presenteeism, work-related stress) and positive (i.e., work engagement) work-related wellbeing are separate health constructs. Although work engagement negatively correlated with work-related stress and presenteeism, the effect size suggests that these are not conceptual opposites. Therefore, to obtain a comprehensive picture of the state of hospital doctors' work-related wellbeing, it is imperative that researchers and practitioners do not focus solely on burnout. It is equally vital to further understand how these different constructs relate to each other, what their antecedents are, and how they impact on patient care.

8.5.4 Limitations

A number of limitations exist within the present study. The first is the sample's heterogeneity that does not distinguish between the specialty and level (e.g., registrar, consultant) of the doctor. The hospital doctors surveyed likely have varied job roles that differentially impacts on their wellbeing. Second, the lack of demographic data means it was not possible to control for these factors. However, previous research involving doctors demonstrate age and gender to have no (Hakanen et al., 2005) or limited (Bernburg et al., 2016; Prins, Gazendam-Donofrio, et al., 2007) effect on wellbeing or psychosocial exposure. Third, the cross-sectional design means causality cannot be determined. The fourth limitation to consider is the single item measures used for presenteeism and work-related stress. These are vulnerable to measurement error, and potentially fail to accurately capture the broad and complex constructs that they represent (Heck & Thomas, 2015). Nevertheless, the multiple publications that demonstrate the relationships involving both these items collectively support their construct validity (Admasachew & Dawson, 2011; Powell et al., 2014). Similarly, the fifth limitation centers on the low internal reliability scores (see Section 7.5) for insufficient work resources and workplace aggression. This unreliability increases measurement error and likely reduces the effect sizes for correlations involving these measures (Lance et al., 2006). Sixth, despite the checks for common method bias in Section 7.6 it is not possible that common method variance may still have some effect on the relationships observed. Finally, measures were treated as observed variables and not as latent factors which would have accounted for measurement error (Kline, 2016). Although multilevel structural equation modelling (MSEM) has become more common (Preacher, Zyphur, & Zhang, 2010), initial attempts to utilise latent variables resulted in an interaction model that was too computationally heavy and unable to converge. Appendix VI presents the results of the model of the fixed effects of individual and trust-level predictors onto work-related wellbeing from a MSEM design. While the results are generally similar, larger effect sizes are noted for workplace aggression which suggests that measurement error may have influenced the results.

8.6 Conclusion

8.6.1 Study conclusion

This study highlights the JD-R model as a useful framework in which to understand the predictive association between job demands, job resources, and trust-level demands on the work-related wellbeing of hospital doctors. Despite the validity of the model's main propositions receiving mixed success, the results demonstrate that different independent pathways predict negative work-related wellbeing (work-related stress, presenteeism) and work engagement in hospital doctors. Crucially, the proposed interactions between job demands and job resources were not observed. This emphasises the need to better understand the differences between challenge and hindrance demands, as well as how to better match job demands with job resources when trying to mitigate the detrimental impact of job demands. Finally, the study integrates a multilevel perspective into the JD-R model, demonstrating that trust-level demands do influence work engagement in hospital doctors. Collectively, these findings highlight the complexity of work-related antecedents to hospital doctors' work-related wellbeing. Hence, it is likely that any successful health intervention will have to target the appropriate antecedent pathway and recognise the role of trust-level factors when trying to manage hospital doctors' work-related wellbeing.

8.6.2 Implications for thesis

Reflecting on the thesis' main research question, which is whether hospital doctors' psychosocial work environment influence quality of patient care, the findings from this chapter presents the initial evidence linking doctors' psychosocial work environment to their work-related wellbeing. It also provides some support at the individual-level for the relevance of the JD-R model to a sample of hospital-based doctors. What this study's model did not do is include quality of patient care. Therefore, the subsequent studies expand this initial model to test whether the relationship that job demands and resources have with work-related wellbeing can be extended to influence the quality of care being offered to patients. Furthermore, the next chapters utilise the more robust multilevel structural equation modelling to account for measurement error.

Chapter 9 : The Mediating Role of Doctors' Work-related Wellbeing (Study 3)

Chapter Eight demonstrated that separate pathways existed between the psychosocial working conditions of doctors and their work-related wellbeing. More specifically, job demands (workplace aggression, insufficient work resources) predicted presenteeism and work-related stress, while job resources (manager support, job control) predicted work engagement. This chapter extends the previous analyses to examine the mediating role of doctors' work-related wellbeing, in the relationship between psychosocial working conditions and self-reported quality of care. The latter is measured through the number of errors seen in doctors' work environment, the quality of care doctors feel they themselves provide (i.e., quality of individual care), and the quality of care that doctors feel their trust provides (i.e., quality of trust care). The chapter first reviews the study hypotheses before introducing the concept of statistical mediation from both a single and multilevel perspective. It subsequently outlines the study methodology and then presents and discusses the results found.

9.1 Introduction

Quality of care as a construct was reviewed in Section 3.1. To summarise, the Department of Health (2008) defines quality of care in the NHS as comprising three aspects: patient safety, clinical excellence, and the experience of patients. Chapter Three highlights the dearth of studies examining the antecedents of quality of care provided by doctors. This includes both psychosocial working conditions and work-related wellbeing. In terms of the former, the systematic review in Chapter Four did find 21 studies testing a relationship between doctors' psychosocial working conditions with quality of care, although these mostly lacked a clear theoretical framework, and failed to acknowledge the organisational context and predictors that these relationships were set in. This chapter aims to further examine the job demands-resources model (JD-R) amongst doctors in English hospitals by extending the relationships between job demands and resources with work-related wellbeing observed in Chapter Eight, to examine their effect on doctor-rated quality of care.

Job demands are hypothesised to associate with lower quality of care due to overloaded doctors wasting energy and time coping with their conditions (Jex, 1998) as well as ignoring important contextual cues and information (S. Cohen, 1980). In contrast, job resources should positively associate with quality of care as job resources mitigate the negative effect of job demands, provide opportunities to cope with challenging situations, and obtain support and resources to achieve work goals (Bakker & Demerouti, 2017; Deci & Ryan, 1985; Shirom et al., 2006). Therefore, this study hypothesises that:

> H_1: Insufficient work resources and workplace aggression (i.e., job demands) will positively predict the numbers of errors seen, and negatively predict both the quality of individual and trust care.

> H_2: Manager support and job control (i.e., job resources) will negatively predict the numbers of errors seen, and positively predict both quality of individual and trust care.

As discussed in Chapter Three, the evidence suggests a relationship between doctors' work-related wellbeing and quality of care (Hall et al., 2016; Scheepers et al., 2015). However, much of this research is drawn from multidisciplinary and nursing samples, with few studies focusing on doctors. Drawing on the information reviewed, it is hypothesised that:

> H_3: Doctors' work-related stress and presenteeism will predict quality of care. More specifically they will positively predict the numbers of errors seen, and negatively predict the quality of individual and trust care.

> H_4: Doctors' work engagement will negatively predict the numbers of errors seen, and positively predict both quality of individual and trust care.

A multilevel perspective was introduced in Chapter Eight, demonstrating that trust-level predictors (number of emergency admissions and the bed occupancy rates), are associated with hospital doctors' presenteeism and work engagement. This system-level perspective (Lowe & Chan, 2010) is imperative in understanding the complexity of quality-of-care antecedents. It is anticipated, congruent with the JD-R model, that trust-level demands should function in a similar manner to job demands (Schaufeli & Taris, 2014). Therefore, it is predicted that:

> H_5: The number of emergency admissions and bed occupancy rates will negatively predict quality of individual and trust care, and positively predict the number of errors seen.

According to the JD-R model's dual pathways, job demands and resources' impact on performance will be mediated by work-related wellbeing. While support of this is evident in the wider literature (Bakker & Bal, 2010; Xanthopoulou et al., 2009), the evidence among doctors is not only minimal but also mixed (Krämer et al., 2016; Loerbroks et al., 2016; Weigl et al., 2015). Drawing on the JD-R model's theoretical postulations and the research supporting it, it is hypothesised that:

> H_6: Doctors' work-related stress and presenteeism will mediate the relationship between job demands (insufficient work resources, workplace aggression) with quality of care.

> H_7: Doctors' work engagement will mediate the relationship between job resources (manager support, job control) with quality of care.

Assuming that these mediation postulations extend to trust-level demands, work-related wellbeing should also function as a mediator towards quality of care. However, as Chapter Eight only found emergency admissions to predict presenteeism and work engagement, while bed occupancy predicted work engagement, the following mediations are predicted:

> H_8: The number of emergency admissions' relationship with quality of care will be mediated by work engagement and presenteeism.

> H_9: Bed occupancy rates' relationship with quality of care will be mediated by work engagement.

Finally, contrary to the JD-R model, Chapter Eight observed a relationship between doctors' perceived insufficient work resources with work engagement. It is therefore plausible that work engagement functions as a mediator between this job demand and quality of care:

> H_{10}: Doctors' work engagement will also mediate the relationship between insufficient work resources and quality of care.

9.2 What is Mediation?

To test the mediation hypotheses specified above it is essential to understand what statistical mediation refers to. Mediation is a form of third variable testing that aims to better explain the mechanisms between two variables and is a pivotal form of analysis in psychology research (Hair et al., 2014). The mediator functions as an outcome of the predicting variable, and in turns impacts upon the outcome variable (Baron & Kenny, 1986). In contrast, a related but different form of third variable testing is moderation; which, refers to a variable that alters the strength or direction between a predictor and an outcome (Baron & Kenny, 1986). As outlined in Section 9.1 above, it is predicted that poor psychosocial working conditions would be associated with poor quality of care. This direct relationship is known as the *direct effect* (Baron & Kenny, 1986; Hair et al., 2014). However, Hypotheses 6 to 9 postulate that job demands, job resources, and trust-level demands all impact onto hospital doctors' work-related wellbeing, which in turn influences the quality of care being provided. Therefore, work-related wellbeing is arguably a mediator that explains how psychosocial working conditions influences quality of care, with this relationship known as the *indirect effect* (Baron & Kenny, 1986; Hair et al., 2014). This section provides a more in-depth review into mediation and how it is tested from a single and multilevel perspective.

9.2.1 Baron and Kenny's (1986) principles of mediation

Although the concept of mediation has long existed in psychology, it was Baron and Kenny's seminal article in 1986 that provided a framework in which to understand and test for mediation. This resulted in a proliferation of studies in the social sciences utilising mediation models (Kenny, Kashy, & Bolger, 1998). Typically referred to as the traditional mediation analysis, this approach proposes a four causal-step process displayed in Figure 9.1 below. Each pathway is represented by a label (i.e., *a, b, c, c'*) that should be met for full mediation to occur.

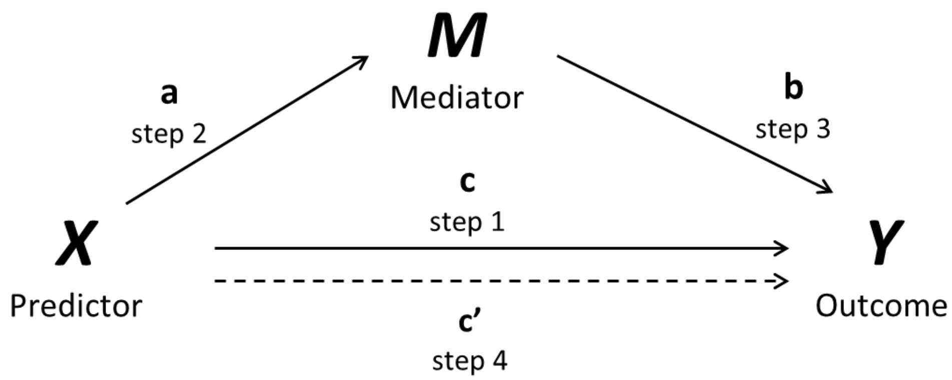

Figure 9.1: Baron and Kenny's (1986) four causal-step mediation approach

The four-causal steps are typically tested via a series of multiple regressions (Kenny et al., 1998). In the first step (path *c*), a direct significant relationship needs to be established between the predictor variable ('*X*') and the outcome variable ('*Y*'). This step is pivotal in determining whether there even exists an effect that could be mediated. In the second step, the mediator ('*M*') is regressed onto the predictor ('*X*') to establish path '*a*'. Step three determines whether a relationship exists between the mediator and the outcome variable (path *b*) by regressing the latter onto the former. Finally, full mediation (path *c′*) is accepted if the effect of *X* on *Y* is zero when *M* is controlled for. Should the fourth step not be satisfied then the first three steps collectively indicate partial mediation. From a theoretical perspective, relationships in psychology are subject to multiple confounding variables; therefore, a more realistic expectation is for mediators to significantly reduce the effect of path *c* rather than expecting full mediation to occur (Baron & Kenny, 1986).

Despite the impact of Baron and Kenny's mediation approach, this approach is not without criticism (A. F. Hayes, 2009; MacKinnon, Lockwood, & Williams, 2004; Zhao, Lynch, & Chen, 2010). Crucially, these four steps do not actually provide a measure for indirect effect *ab*, thereby relying solely on the presence of these four steps to establish mediation. However, Baron and Kenny (1986) themselves write that the fourth step is redundant unless full mediation is postulated. Furthermore, some have queried whether the direct relationship between the predictor and outcome (Step One) is even needed (Kenny, 2016; Kenny et al., 1998). This is evident where the total effect is affected by two groups of participants that vary in their relationship between *X* and *Y* (A. F. Hayes, 2009). For example, this can occur when a relationship in males and females are of a similar strength but operate in opposite directions, thereby cancelling out each other. Hence, necessitating this step reduces the available statistical power to detect mediation (MacKinnon et al., 2004; Zhao et al., 2010). An additional limitation lies in the need for observations to be independent of each other (Kenny et al., 1998). However, when participants are clustered into groups (e.g., schools, hospitals) this violates this assumption of independence. This is a problem in multilevel designs, and is discussed in the section below.

9.2.2 Multilevel mediation

The clustering of doctors within trusts means there is no independence of observations. Consequently, traditional mediation analyses are not suitable as these would lead to downwardly biased standard errors (Preacher et al., 2010). Unlike single-level mediation, multilevel mediation models vary as to whether the variable is located at Level-1 or 2, leading to a number of possible models. Kenny, Kashy, and Bolger (1998) first distinguished between upper and lower-level mediation (Figure 9.2); the former refers to a Level-2 *M* mediating a Level-2 *X* onto a Level-1 *Y* (i.e., a 2-2-1 design), while the latter encompasses either a Level-2 (2-1-1) or Level-1 (1-1-1) *X* with a Level-1 *M* and *Y*. More complex designs, particularly those involving

Level-2 *M*s and *Y*s (e.g., 1-1-2, 2-1-2, 1-2-2), have since been proposed (Bauer, Preacher, & Gil, 2006; Preacher et al., 2010).

Advances in software and modelling have allowed researchers to progress from using multilevel modelling (with observed variables) to Multilevel Structural Equation Modelling (MSEM) with mediation analyses (Preacher et al., 2010). In MSEM, the variance of a variable is separated into its between and its within component (Asparouhov & Muthén, 2006), variables at Level-1 typically only contain the within-variance component, while variables at Level-2 contain the between-variance component (Bauer et al., 2006). Should a variable contain both a within and between-variance component, the within-component will not only be uncorrelated with its own between-variance component but also with all other between-variance components in the model. In the same way, the between-variance component will not be correlated to any of the within-variance components in the model.

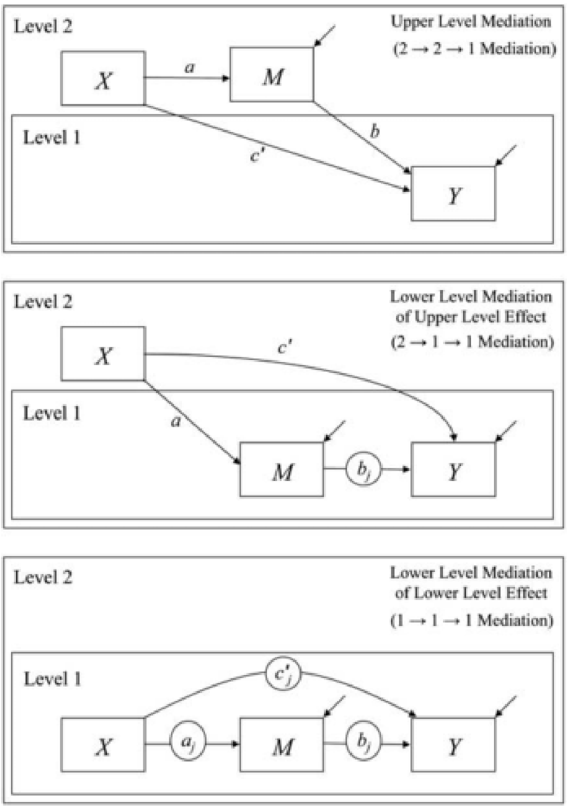

Figure 9.2: Upper and lower level mediation in a two-level model (from Bauer et al., 2006)

The separation of within and between-variance components means that relationships can be independently modelled at both levels. This allows for multilevel designs where any, or even all, of the constructs can be modelled at Level-1 or Level-2 (Preacher, Zhang, & Zyphur, 2011;

Preacher et al., 2010). However, because the within-variance components are separate from the between-variance components, it is not possible for either to affect the other. As such, relationships at the lower-level can only examine how the within-variance components of variables relate with each other. This also applies with relationships at the higher-level that can only utilise between-variance components. Consequently, the term *cross-level mediation* can be misleading as in the case of a 2-1-1 design, this actually refers to between-variance of the Level-2 predictor impacting upon the between-variance component of the Level-1 mediator, and not the within-variance or total-variance of the Level-1 mediator (Zhang, Zyphur, & Preacher, 2009). This extends to the 1-1 part of the design, where the between-variance components are related between the mediator and the outcome. This is in order to avoid any conflation between within and between-variance components. Therefore, the mediation only really occurs at the between-group level. As long as at least one of the constructs in the mediation design, regardless if it is the predictor, mediator, or outcome, is modelled at Level-2, then the entire relationship can only occur at the higher-level.

Advantages of MSEM. At least three issues exist indicating MSEM is better suited to mediation than multilevel modelling (Preacher et al., 2010). The first pertains to the MSEM's utilisation of latent variables rather than observed variables in multilevel modelling, which accounts for measurement error (Kline, 2016). Measurement error increases reliability within models and the ability to conduct more meaningful analysis (Hair et al., 2014; Kline, 2016). By integrating latent variables into multilevel modelling, abstract constructs are hypothesised to be represented by observed measures (Goodwin, 1999; Heck & Thomas, 2015). Second, because group standings of all variables at Level-1 are considered latent, sampling error is corrected. In comparison, the group standings of a Level-1 X variable is represented at Level-2 through group means (Raudenbush & Bryk, 2002). Finally, traditional multilevel mediation models constrain the within and between-effect to be equal across two Level-1 variables, such as between M and Y in a 2-1-1 model (Preacher et al., 2011). However, by separating variance at both levels, the potential problems of conflated between and within-level effects are avoided in multilevel modelling. This results in more precise estimations of indirect effects (Preacher et al., 2010). Therefore, considering the superiority of MSEM over multilevel modelling with regards to mediation, the former will be used to test the proposed mediation hypotheses.

Estimator. The default estimators used in SEM, discussed in Section 7.1.2, require variables to be normally distributed and to contain sufficient range (Browne, 1984; Curran et al., 1996). Dichotomous items contain insufficient responses as their response range is restricted and their error distribution is non-random (Heck & Thomas, 2015). Similarly, ordinal data is often assumed as continuous (Carifio & Perla, 2007) when in reality Likert-scales provide skewed distributions as responses bunch up on one end of the scale (Heck & Thomas, 2015). This too biases the models parameters, standard errors, and fit indices (B. O. Muthén & Kaplan, 1992). As the dataset is the same as that from Chapter Seven, this creates a problem as all the items to be included in this analysis are either dichotomous or Likert-based. Therefore, as was the case with the CFAs carried out earlier, WLSMv will be used as the model estimator (Asparouhov & Muthén, 2013; Hox et al., 2010; L. K. Muthén & Muthén, 2017). Not only is WLSMv able to handle dichotomous and ordinal variables but it provides a more conservative and robust approach in comparison to the other estimators.

Missing data. Despite the robustness of WLSMv, this estimator is vulnerable to missing data during multilevel modelling. This is as missing data biases the correlational estimates, which in turn biases the structural parameters estimates (Asparouhov & Muthén, 2010a). To address this, simulation studies have revealed that multiple imputation followed by the WLSMv estimator is the most straightforward and robust manner in dealing with missing data in two-level models (Asparouhov & Muthén, 2010a). It is important to emphasise that the WLSMv when used in confirmatory factor analyses is robust in relation to missing data (Asparouhov & Muthén, 2010c).

Therefore, to address missing data in the dataset, multiple imputation should be used. This is considerably more robust than other techniques that would likely yield bias outcomes, such as mean substitution, regression-based imputation, and listwise deletion (Heck & Thomas, 2015). For example, mean substitution reduces variance while listwise deletion can inflate standard error scores. Multiple imputation in Mplus identifies missing data patterns before imputing plausible variables based on an EM algorithm (Peugh & Enders, 2004). This comprises of an iterative two-step process, where missing data is imputed before the covariance matrix and mean vector are estimated repeatedly until there are trivial differences in the covariance of adjacent iterations. From these imputed values, complete datasets can then be analysed with mean estimates and standard errors. Mplus allows a Bayesian estimator to be used to impute missing categorical and ordinal data (Asparouhov & Muthén, 2010b). Crucially, simulation studies reveal that five imputed datasets using Bayesian imputation with WLSMv obtained robust results and that it performed as well as ML and Bayesian estimators (Asparouhov & Muthén, 2010a, 2010b, 2010c).

Bootstrapping in MSEM. The limitations of Kenny et al.'s (1998) traditional mediation approach has led to some advocating the Sobel Test (Sobel, 1982) (i.e., the *delta* method), which involves significance testing. However, this too has its weaknesses as it only works well in large samples and when sampling distribution is normal (A. F. Hayes, 2009; Tofighi & Thoemmes, 2014). Instead, resampling methods such as bootstrapping have become increasingly dominant (Field, 2014; A. F. Hayes, 2009). A non-parametric method, bootstrapping works by resampling with replacements over many times (e.g., 5000 times). The indirect effect is calculated from each sample creating a sampling distribution. This allows for the correction of bias as the mean of the bootstrapped distribution will not equal the indirect effect (Kenny, 2016). Consequently, a confidence interval can be estimated. If a zero occurs within the interval then the indirect effect is not significant. Confidence intervals add value to the bootstrapping approach as its increases power, provides more accurate confidence interval estimates suitable for hypotheses testing, and does not assume a normal sampling distribution (Field, 2014; A. F. Hayes, 2009; Kenny, 2016).

Despite the advances with mediation analyses at both the individual and multilevel, this has not extended into multilevel bootstrapping. In trying to estimate confidence intervals for multilevel indirect effects, Preacher et al. (2012) described three methods: the distribution of product method, parametric bootstrapping, and non-parametric bootstrapping. However, only parametric bootstrapping has successfully been used in the multilevel context, and will hence be the only method to be described here. Essentially the multilevel parametric bootstrap is equivalent to the Monte Carlo bootstrap method at the single-level (MacKinnon et al., 2004). Only one parameter estimate is assumed to be normally distributed, with no assumption made as to the distribution of the indirect effect (which usually is not normally distributed; Bauer et al., 2006). In simulation studies, Pituch et al. (2006) found this method to perform as well as bias-corrected bootstrapping, which is difficult to run with multilevel designs. Using a web-based utility by Selig and Preacher (2008), Preacher et al. (2010) demonstrated that it was possible to obtain multilevel bootstrapping confidence intervals without using raw data, and that yielded asymmetric confidence intervals faithful to the skewed sampling distributions of indirect effects.

Effect size. The traditional mediation approach (Baron & Kenny, 1986) focuses on whether the mediation has a full or partial effect. This presents little utility to researchers who may view the mediating effect as occurring as a continuum. However, attempts to generate a standardised indirect effect metric has to date been challenging (Preacher & Kelley, 2011). For example, early attempts have focused on calculating the effect size as the proportion of the indirect effect (*ab*) to the total effect (*c*) (Sobel, 1982). However, this has been criticised for providing misleading estimates on practical significance, neglecting the possibility of multiple mediators, and that negative values and values exceeding one are possible (A. F. Hayes, 2009; Preacher & Kelley, 2011). Consequently, Kappa-squared (κ^2) was proposed (Preacher & Kelley, 2011). This represented the ratio of the observed indirect effect relative to the maximum possible indirect effect given the scales of the variables involved. While increasingly popular, Wen and Fan (2015) identified that parts of the formulas used to calculate κ^2 are generally not true, which results in κ^2 not actually being monotonic - making the effect size neither understandable or explainable. Considering the limitations in this area, Shrout and Bolger (2002) recommended using Cohen's (1988) suggestions for effect sizes to distinguish between large (.5), medium (.3), and small (.1) effect sizes. However, because the indirect effect is the result of multiplying two effect sizes, the effect sizes should in fact be squared (Kenny, 2016). Consequently, .25 represents a large effect size, .09 a medium, and .01 a small effect size.

9.2.3 MSEM and the study hypotheses

The hypotheses in Section 9.1 are all congruent with the MSEM approach. Separating the variance-components of each construct into its within and between-component makes it possible to examine the direct and mediating relationships of both Level-1 (i.e., individual-level) and Level-2 (i.e., trust-level) variables. The former include the constructs that represent job demands (insufficient work resources, workplace aggression), job resources (manager support, job control), work-related wellbeing (work engagement, work-related stress, presenteeism), and quality of care (quality of individual care, quality of trust care, errors seen). The within-variance components of all these relationships are examined. The inclusion of number of emergency admissions and bed occupancy rates are both trust-level variables. Hence, these two variables can only be examined with the between-variance components of the three work-related wellbeing and three quality-of-care measures.

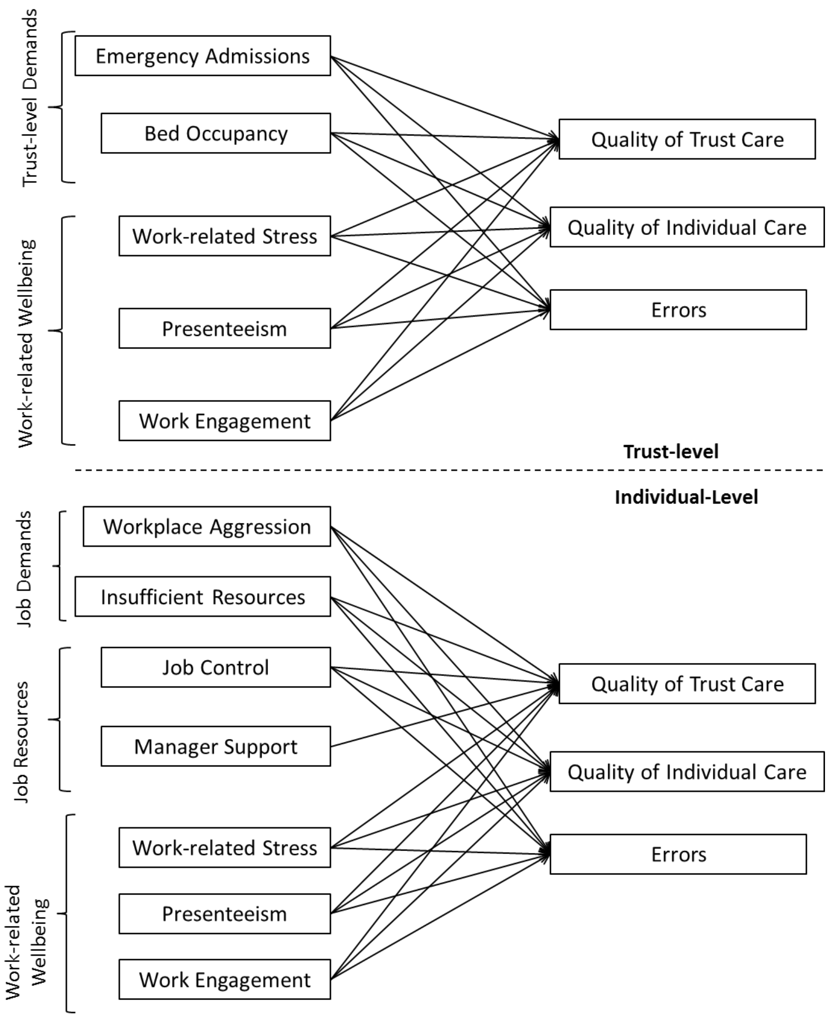

Figure 9.3: Direct relationships between working conditions and quality of care

Figure 9.3 presents the direct relationships hypothesised at the individual-level that job demands (H_1), job resources (H_2), work-related stress and presenteeism (H_3), and work engagement (H_4) have on quality of care. At the trust-level, the same figure illustrates the direct relationships that the trust-level demands (H_5) have on quality of care. H_6, H_7, and H_{10} (Figure 9.4) postulates that work-related wellbeing mediates the relationships between job demands and resources with quality of care. All the variables here occur at the individual-level, and therefore represent a 1-1-1 MSEM design. H_8 and H_9 utilise trust-level demands, with the mediator and outcome both consisting of individual-level measures. However, it is important to note that in this 2-1-1 design it is only the between-trust variance components of the mediating and outcome variables that are being used.

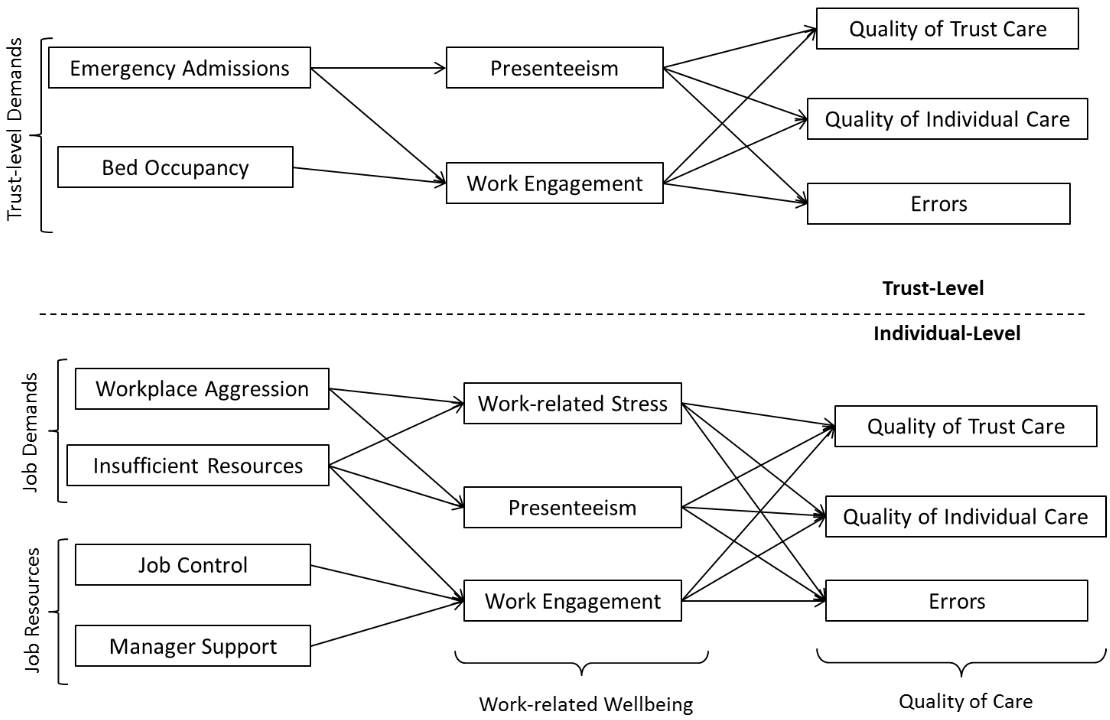

Figure 9.4: Work-related wellbeing as mediators between working conditions and quality of care

9.3 Method

9.3.1 Sample

The study sample was described in depth in Section 6.2.1. It consisted of 14,066 doctors from 157 acute trusts. Mean doctors per trust was 89.59 (*SD*=94.76) with a median of 41 doctors.

9.3.2 Materials

Measures used in Chapter Eight. At the individual-level, the measures for job demands (insufficient work resources, workplace aggression), job resources (job control, manager support) and work-related wellbeing (work-related stress, presenteeism, work engagement), were all used and described in the preceding study (Chapter Eight). They were first introduced during the creation of composite scales (Chapter Seven). Again, tenure was used as a control variable. The trust-level variables – trust type, bed occupancy rates, and number of emergency admissions were also described in Chapter Eight.

Measures dropped from Chapter Eight. Two measures from the preceding study were dropped in this study: effective team practices (a job resource) and number of hospital beds (a trust-level control). This was due to the non-impact both had on work-related wellbeing measures in that study. Therefore, in the interest of parsimony, and to avoid Type II errors by reducing statistical power (Becker, 2005), both measures were excluded from this study.

Quality-of-care measures. Three new measures that represented quality of care were included in this study. Quality of trust care measured how well doctors felt that the overall quality of care their employing trust was providing. Quality of individual care represented how well doctors rated the quality of care they themselves were able to provide. Both these measures used three items rated on a five-point Likert scale. Errors seen was measured with two items asking whether doctors had in the previous month witnessed any errors, near misses, or incidents that could hurt staff or patients/service users. Each item was responded with a dichotomous "yes" or "no".

9.3.3 Analysis

To test the study hypotheses, a series of models were built from the bottom up (Heck & Thomas, 2015; Stride, 2016). This was similar to the process used in the study in Chapter Eight (see Section 8.3.2 for more detail). After the multilevel factor structure was confirmed, the first step determined the unconditional model. The second model added the control variables, with subsequent models testing the direct impact that individual-level (Model 3) and trust-level (Model 4) variables had on quality of care. At the individual-level, seven measures were added to represent job demands (H_1), job resources (H_2), and work-related wellbeing (H_3). At the trust-level, number of emergency admissions and bed occupancy rates were included as demands (H_4). The fifth model examined the indirect effects between job demands and resources with quality of care (H_6, H_7, H_{10}). In line with Preacher et al.'s (2010) recommendation, Muthén and Asparouhov's (2008) MSEM approach was applied to mediation analyses. As all variables existed at the individual-level, this represented a 1-1-1 mediation design where the within-variance components were the focus of the analyses. The sixth and final model tested whether work-related wellbeing mediated the relationship between trust-level demands with quality of care (H_8, H_9). As trust-level demands operate at the trust-level, while work-related wellbeing and quality of care were at the individual-level, this represented a 2-1-1 mediation design. However, because Level-2 variables were being used, to prevent variance conflation only the between-trust variance components on all variables were used in this model (Zhang et al., 2009).

These analyses were carried out in Mplus 8 (L. K. Muthén & Muthén, 2017). Weighted least square (WLSMv) was used as the model estimator as not only is it able to handle dichotomous and ordinal variables but it provides a more conservative and robust approach in comparison to other estimators (Asparouhov & Muthén, 2013; Hox et al., 2010). Five datasets were generated using a Bayesian estimator to impute missing categorical and ordinal data (Asparouhov & Muthén, 2010b). This was to address potential biased estimators from the WLSMv estimator due to missing data (Asparouhov & Muthén, 2010a). As chi-square difference testing is not usable with the WLSMv estimator (L. K. Muthén & Muthén, 2017) and the chi-square difference test '*difftest*' not available for multilevel analyses, Wald chi-square test of parameter equalities was used instead to compare models. This functions by testing the null hypotheses that a set of parameters are equal to a set value. In order to obtain bootstrapped confidence intervals, Selig and Preacher's (2008) programme simulated the sampling distribution of the indirect effects. The number of bootstrapped samples was set at 20,000 at 95% confidence intervals. Trust-level predictors were grand-mean centered.

9.4 Results

9.4.1 Descriptive results

Table 9.1 provides the correlation values at the individual and trust-level for the study measures. At the individual-level, the three quality-of-care measures all correlated significantly with workplace aggression, insufficient work resources, manager support, job control, work-related stress, presenteeism, and work engagement ($p<.001$). More specifically, quality of trust and individual care both correlated positively with job resources and work engagement, and negatively with job demands, presenteeism, and work-related stress. The converse relationships were observed for errors seen.

The same patterns of relationships were observed for quality of trust care, quality of individual care, and errors seen at the trust-level (Table 9.1). In terms of trust-level demands with quality of care, bed occupancy rates only correlated with quality of trust care (r=-.23). Mean number of weekly emergency admissions correlated with quality of trust (r=-.37) and individual (r=-.30) care, and errors seen (r=.18).

Table 9.1: *Within-trust and between-trust correlations*

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Tenure	1	-.05***	.15***	-.16***	-.03**	.06***	.05***	-.07***	-.21***	-.03***	.04***
2. Aggression	-.11	1	.19***	-.10***	-.11***	.21***	.18***	-.16***	-.15***	-.15***	.24***
3. Insufficient work resources	-.16	.42***	1	-.40***	-.41***	.29***	.21***	-.40***	-.51***	-.43***	.27***
4. Manager support	-.11	-.21**	-.58***	1	.57***	-.26***	-.18***	.42***	.52***	.31***	-.15***
5. Job control	-.05	-.29***	-.61***	.65***	1	-.26***	-.18***	.51***	.54***	.38***	-.11***
6. Work-related stress	.03	.34***	.56***	-.47***	-.38***	1	.29***	-.36***	-.25***	-.23***	.19***
7. Presenteeism	.01	.32***	.26***	-.24**	-.26***	.24**	1	-.21***	-.19***	-.12***	.13***
8. Work engagement	-.10	-.35***	-.53***	.60***	.52***	-.60***	-.36***	1	.49***	.44***	-.17***
9. Quality of trust care	-.02	-.37***	-.76***	.63***	.68***	-.52***	-.27***	.54***	1	.45***	-.23***
10. Quality of individual care	.03	-.39***	-.68***	.46***	.51***	-.44***	-.16*	.55***	.58***	1	-.20***
11. Errors seen	.04	.38***	.51***	-.36***	-.23**	.46***	.17*	-.37***	-.46***	-.55***	1
12. Bed occupancy rate	-.08	.16*	.23**	-.11	-.12	.13	.02	-.06	-.23***	-.12	.08
13. Emergency admissions	-.03	.15	.30***	-.27***	-.32***	.05	-.07	-.22**	-.37***	-.30***	.18*

Note. ***p<.001; **p<.01; *p<.05. Correlations above the diagonal are individual-level correlations. Correlations below the diagonal are trust-level correlations, with individual-level measures aggregated to the trust-level (N=157).

9.4.2 Confirmatory factor analysis

In the previous study (Chapter Eight), effective team practices did not predict work-related stress, presenteeism, or work engagement. As a result their items were excluded from this study while three new quality-of-care constructs (trust care, individual care, error) were included. Furthermore, the testing of presenteeism, work engagement, the three quality-of-care constructs as outcomes of trust-level predictors meant that a trust-level model of these constructs had to also be confirmed. Collectively, these necessitated a new CFA.

Table 9.2: *Standardised loadings for within-trust level items*

Latent construct	Item	Estimate	S.E.	Est./S.E.	*p*-Value
Workplace aggression	AG1	0.54	0.021	26.21	***
	AG2	0.60	0.037	16.18	***
	AG3	0.89	0.028	32.40	***
Insufficient work resources	IR2	0.80	0.005	165.07	***
	IR3	0.81	0.005	171.93	***
Manager support	MS1	0.92	0.002	564.38	***
	MS2	0.93	0.002	515.94	***
	MS3	0.89	0.002	466.35	***
	MS4	0.88	0.002	408.90	***
	MS5	0.79	0.003	240.16	***
Job control	JC1	0.84	0.003	305.31	***
	JC2	0.91	0.002	446.34	***
	JC3	0.89	0.002	428.95	***
	JC4	0.90	0.002	474.66	***
Work engagement	EG1	0.92	0.003	306.06	***
	EG2	0.92	0.002	377.84	***
	EG3	0.71	0.005	151.00	***
Quality of individual care	QC1	0.89	0.003	270.31	***
	QC2	0.70	0.004	162.57	***
	QC3	0.92	0.003	303.99	***
Quality of trust care	ORG1	0.88	0.003	270.31	***
	ORG2	0.86	0.004	162.57	***
	ORG3	0.85	0.003	303.99	***
Errors seen	ER1	0.87	0.017	51.63	***
	ER2	0.81	0.016	50.15	***

Note. ****p*<.001

The overall fit of the measurement model was good. Although RMSEA (.06) was higher than the recommended <.05, CFI (.96) and TLI (.95) were congruent with the levels suggesting good fit (Byrne, 2012; L. Hu & Bentler, 1998). Chi-square (χ^2=26234.55; *df*=390; *p*<.001) was significant. A review of individual parameters in Table 9.2 indicates that all items, with the exception of two, exceeded the recommended threshold of .7 standardised loadings; these two items still surpassed the minimum acceptable standardised loading of .5 (Hair et al., 2014). At the trust-level, Table 9.3 presents that all items had strong loadings onto their respective constructs.

Table 9.3: *Standardised loadings for between-trust level items*

Between-level latent variable	Item	Estimate	S.E.	Est./S.E.	p-Value
Work engagement	EG1	1.00	0.001	999.00	999
	EG2	0.98	0.034	28.90	***
	EG3	0.99	0.069	14.32	***
Quality of individual care	QC1	1.00	0.001	999.00	999
	QC2	0.97	0.072	13.96	***
	QC3	0.99	0.032	31.83	***
Quality of trust care	ORG1	1.00	0.001	999.00	999
	ORG2	0.99	0.025	40.52	***
	ORG3	0.80	0.049	16.35	***
Errors seen	ER1	1.00	0.001	999.00	999
	ER2	0.84	0.121	6.94	***

Note. ****p*<.001

9.4.3 The unconditional model

The specified unconditional model involved all three outcome variables. From their ICC values it was not clear whether multilevel modelling was needed: quality of trust care (0.08), quality of individual care (0.03), and errors seen (0.02). However, their respective *deff* scores exceeded the recommended value of 2 (7.62, 3.44, and 2.32), indicating that clustering should be taken into account (B. O. Muthén & Satorra, 1995). The fit of the unconditional model was good (χ^2=1947.01; *df*=34; RMSEA=.06; CFI=.99; TLI=.98).

9.4.4 The addition of control variables onto quality of care

Two control variables (tenure, trust status) were added to the basic model. This resulted in significant change in chi-square based on the Wald Test (χ^2=81.22; *df*=2; *p*<.001). The model fit was good (χ^2=574.97; *df*=47; RMSEA=.03; CFI=.99; TLI=.99). As seen in Table 9.4, tenure was a predictor of quality of trust care (β=-.12), but did not predict errors seen or quality of individual care. Trust status had no impact on any of the three quality-of-care measures.

9.4.5 Direct effects of individual-level variables onto quality of care

The third model, with the inclusion of job demands, job resources, and work-related wellbeing obtained a good model fit (χ^2=5463.72; *df*=367; RMSEA=.03; CFI=.99; TLI=.99). The Wald Test indicated that these additions made a significant improvement to the model (χ^2=4740.71; *df*=9; *p*<.001). As Table 9.4 presents, work engagement was predicted by: insufficient work resources (β=.39) and job control (β=.32). Both work-related stress and presenteeism were predicted by insufficient work resources (β=.51, β=.26) and workplace aggression (β=.24, β=.26). Support was found for H_1, H_2, and H_4 as all three quality-of-care outcomes were predicted by insufficient work resources, workplace aggression, job control, and work engagement. No support was found for H_3 as work-related stress and presenteeism did not have the expected relationship with any of the quality-of-care outcomes. Although according to Table 9.4 tenure significantly predicated five outcomes, the effect sizes were all less than .10, which is the cut-off threshold for a small effect size (Cohen, 1988). The issue of sample size and power was reviewed in Section 6.1.2.

9.4.6 Direct effects of trust-level variables onto quality of care

Two trust-level predictors (mean weekly emergency admissions, bed occupancy rates) were added as part of Model 4. This resulted in an improvement on model fit (χ^2=4644.83; *df*=382; RMSEA=.03; CFI=.99; TLI=.99) with the Wald Test also being significant (χ^2=19.09; *df*=3; *p*<.001). Mixed support was found for H_5. Only two relationships involving trust-level predictors were significant where the mean number of emergency weekly admissions predicted quality of individual care (β=-.28) and work engagement (β=-.36). Bed occupancy rate did not predict any of the five examined outcome constructs.

Table 9.4: *Standardised and unstandardised coefficients for direct effects onto quality of care*

Predictor	Work-related stress		Presenteeism		Work engagement	
	β	b	β	b	β	b
Tenure (w)	.07*** (.05, .09)	0.06*** (0.04, 0.07)	.05*** (.03, .07)	0.03*** (0.02, 0.05)	-.07*** (-.09, -.05)	-0.12*** (-0.12, -0.07)
Insufficient work resources (w)	.51*** (.48, .53)	0.58*** (0.54, 0.63)	.26*** (.23, .29)	0.25*** (0.22, 0.29)	-.39*** (-.41, -.37)	-0.82*** (-0.89, -0.75)
Workplace aggression (w)	.24*** (.21, .28)	0.46*** (0.39, 0.53)	.26*** (.23, .30)	0.42*** (0.36, 0.49)		
Job control (w)					.32*** (.30, .34)	0.49*** (0.46, 0.53)
Manager support (w)					.06*** (.04, .08)	0.07*** (0.04, 0.09)
Specialist (b)			.29 (-.07, .66)	0.15 (-0.04, 0.33)	.30 (-.11, .71)	0.35 (-0.01, 0.84)
Bed occupancy rate (b)					.30 (-.02, .62)	1.09 (-0.13, 2.32)
Emergency admissions (b)			-.02 (-.31, .27)	0.00 (-0.01, 0.01)	-.36** (-.63, -.08)	-0.01* (-0.01, -0.01)

Predictor	Trust care		Individual care		Errors Seen	
	β	b	β	b	β	b
Tenure (w)	-.12*** (-.13, -.10)	-0.13*** (-0.15, -0.11)	-.01 (-.03, .01)	-0.02 (-0.04, 0.01)	.06*** (.04, .08)	0.06*** (0.04, 0.09)
Insufficient work resources (w)	-.54*** (-.57, -.49)	-0.95*** (-1.01, -0.80)	-.52*** (-.56, -.47)	-0.90*** (-0.99, -0.80)	.37*** (.32, .43)	0.60*** (0.47, 0.74)
Workplace aggression (w)	-.04*** (-.06, -.01)	0.10*** (-0.17, -0.020)	-.08*** (-.11, -.05)	-0.22*** (-0.31, -0.14)	.39*** (.35, .44)	1.06*** (0.83, 1.28)
Job control (w)	.21*** (.19, .24)	0.26*** (0.22, 0.31)	.11*** (.09, .14)	0.14*** (0.11, 0.17)	.16*** (.13, .19)	0.19*** (0.14, 0.24)
Manager support (w)	.18*** (.16, .21)	0.16*** (0.14, 0.18)	-.05** (-.07, -.03)	-0.04** (-0.06, -0.03)	-.03* (-.06, -.01)	-0.03 (-0.05, -0.01)
Work engagement (w)	.08*** (.06, .11)	0.07*** (0.04, 0.09)	.26*** (.23, .28)	0.21*** (0.19, 0.23)	-.02 (-.05, .02)	-0.01 (-0.04, 0.01)
Work-related stress (w)	.19*** (.15, .22)	0.27*** (0.22, 0.33)	.10*** (.05, .14)	0.15*** (0.08, 0.21)	-.02 (-.06, .02)	-0.03 (-0.09, 0.03)
Presenteeism (w)	.04*** (.01, .07)	0.07*** (0.02, 0.12)	.10*** (.07, .13)	0.18*** (0.13, 0.23)	-.07*** (-.11, -.03)	-0.11*** (-0.18, -0.05)
Specialist (b)	.26 (-.10, .62)	0.61 (-0.26, 1.48)	-.12 (-.48, .23)	-0.17 (-0.64, 0.31)	-.35 (-.75, .04)	-0.35 (-0.75, 0.05)
Bed occupancy rate (b)	-.10 (-.34, .14)	-0.75 (-2.52, 1.02)	-.09 (-.39, .20)	-0.40 (-1.68, 0.87)	-.09 (-.42, .25)	-0.27 (-1.32, 0.78)
Emergency admissions (b)	-.16 (-.38, .06)	-0.01 (-0.01, 0.01)	-.28* (-.53, -.03)	-0.01* (-0.01, 0.01)	.10 (-.19, .38)	0.01 (-0.01, 0.01)
Work engagement (b)	.60*** (.37, .83)	1.21*** (0.61, 1.82)	.68*** (.47, .90)	0.80*** (0.49, 1.11)	-.33 (-.70, .04)	-0.29 (-0.60, 0.03)
Presenteeism (b)	-.36** (-.63, -.10)	-1.71* (-3.11, -0.32)	-.09 (-.41, .24)	-0.23 (-1.12, 0.66)	.28 (-.13, .68)	0.56 (-0.26, 1.37)

Note. β = standardised beta coefficients; b = unstandardised beta coefficients; **p*<.05; ***p*<.01; ****p*<.001; (w)=within-trust level; (b)=between-trust level; parentheses represents 95% confidence interval.

9.4.7 Indirect effects between job demands and resources with quality of care

The inclusion of twenty one mediation pathways into the specified model resulted in a better fitting model (χ^2=4637.56; *df*=381; RMSEA=.03; CFI= .99; TLI=.99). Wald Test again was significant (χ^2=4943.98; *df*=16; *p*<.001). Sixteen of the mediated pathways were significant (Table 9.5), of which three had a medium effect size (β>.09) and twelve had a small effect size (β>.01). These provided support for H_6, H_7, and H_{10}. Presenteeism mediated all six relationships between job demands (insufficient work resources, workplace aggression) with quality of trust care, individual care, and errors seen. None of the relationships between either job demands or job resources with error were mediated by work-related stress or work engagement.

Table 9.5: *Standardised and unstandardised coefficients for indirect effects between within-trust job demands and resources with quality of care*

Predictor - Quality-of-care relationship	Unstandardised	Standardised
Mediator: Work-related stress		
Insufficient work resources - Trust care	0.16*** (0.12, 0.20)	0.13
Insufficient work resources - Individual care	0.09*** (0.04, 0.13)	0.06
Insufficient work resources - Error	-0.02 (-0.05, 0.02)	-0.05
Workplace aggression - Trust care	0.13*** (0.10, 0.15)	0.05
Workplace aggression - Individual care	0.07*** (0.04, 0.10)	0.03
Workplace aggression - Error	-0.01 (-0.04, 0.02)	-0.02
Mediator: Presenteeism		
Insufficient work resources - Trust care	0.02** (0.01, 0.03)	0.01
Insufficient work resources - Individual care	0.05*** (0.03, 0.06)	0.04
Insufficient work resources - Errors seen	-0.03*** (-0.05, -0.01)	-0.08
Workplace aggression - Trust care	0.03** (0.01, 0.05)	0.01
Workplace aggression - Individual care	0.08*** (0.05, 0.10)	0.03
Workplace aggression - Errors seen	-0.05*** (-0.08, -0.02)	-0.07
Mediator: Work engagement		
Insufficient work resources - Trust care	-0.05*** (-0.07, -0.04)	-0.04
Insufficient work resources - Individual care	-0.17*** (-0.19, -0.16)	-0.13
Insufficient work resources - Error	0.01 (-0.01, 0.03)	0.03
Job control - Trust care	0.03*** (0.02, 0.04)	0.05
Job control - Individual care	0.10*** (0.09, 0.12)	0.14
Job control - Errors seen	-0.01 (-0.02, 0.01)	-0.03
Manager support - Trust care	0.01*** (0.01, 0.02)	0.01
Manager support - Individual care	0.01*** (0.01, 0.02)	0.02
Manager support- Errors seen	-0.01 (-0.01, 0.01)	-0.01

Note. ****p*<.001; ***p*<.01; parentheses represents 95% confidence interval

9.4.8 Indirect effects between trust-level demands with quality of care

Nine pathways were specified on whether presenteeism and work engagement mediated the relationships between trust-level demands and the three quality-of-care outcomes. This resulted in a good fitting model (χ^2=4644.83; *df*=382; RMSEA=.03; CFI= .99; TLI= .99) with Wald Test also indicating a significant change (χ^2=41.70; *df*=9; *p*<.001). Only two significant mediated pathways with small effect sizes were reported (Table 9.6), namely: that work engagement mediated the relationships between the number of emergency admissions with both quality of trust (β=-.06) and individual-level (β=-.03) care. Therefore, there was little support for H_8 and H_9.

Table 9.6: *Standardised and unstandardised coefficients for indirect effects between trust-level demands and resources with quality of care*

Predictor - Quality-of-care relationship	Unstandardised	Standardised
Mediator: Presenteeism		
Emergency admissions - Trust care	0.01 (-0.01, 0.01)	.01
Emergency admissions - Individual care	0.01 (0.00, 0.01)	.01
Emergency admissions - Errors seen	0.01 (-0.01, 0.01)	.02
Mediator: Work Engagement		
Emergency admissions - Trust care	-0.01** (-0.01, -0.01)	-.06
Emergency admissions - Individual care	-0.01** (-0.01, -0.01)	-.03
Emergency admissions – Errors seen	0.01 (0.01, 0.01)	.05
Beds occupancy - Trust care	1.32 (-0.40, 3.04)	.05
Beds occupancy - Individual care	0.88 (-0.18, 1.93)	.03
Beds occupancy - Errors seen	-0.31 (-0.80, 0.18)	-.04

Note. ***p*<.01

9.5 Study Discussion

The present study aimed to test the predictive associations of job demands, job resources, trust-level demands, and work-related wellbeing to three self-reported quality-of-care measures (trust care, individual care, errors seen) in a sample of doctors from English hospitals. It further examined whether work-related wellbeing functioned as a mediator in these relationships. As expected, insufficient work resources, workplace aggression, and job control all predicted the three outcome variables. However, work-related stress only predicted quality of trust and individual care, while presenteeism only predicted quality of individual care. Work engagement and number of emergency admissions (a trust-level measure) only predicted quality of individual care. In terms of mediation, presenteeism mediated all six hypothesised relationships between job demands and quality of care. Work engagement only mediated the relationships that insufficient work resources, job resources, and the number of emergency admissions had with quality of trust and individual care; it was not a mediator when errors seen was the outcome measure. Although support for the hypotheses was mixed, this study makes several contributions towards this thesis and the wider field.

9.5.1 Direct and indirect relationships at the individual-level

The findings provide support that job demands (i.e., insufficient work resources, workplace aggression) are associated with poorer levels of quality of care by the trust and the individual, as well as more errors seen. This could function directly, where overloaded doctors' performance is impaired as they waste energy and time coping with their conditions (Jex, 1998) or by ignoring important contextual cues and information (S. Cohen, 1980). Job demands also appear to influence quality of care indirectly, via levels of presenteeism and work-related stress. This is congruent with the JD-R model's (Bakker & Demerouti, 2017; Demerouti et al., 2001) health-impairment process. More specifically, insufficient work resources and workplace aggression were associated with increased presenteeism and work-related stress, which in turn was associated with lower quality of trust and individual care. Job demands likely arouse a stress process that leads to energy depletion (van Emmerik et al., 2009). This may force strategy adjustments among doctors, including narrowing of attention, increased selectivity, and redefinition of task requirements (Hockey, 1993). These compensatory strategies over long periods of time drain an individual's energy, eventually leading to increased levels of ill-health. In turn, doctors then lack the capacity to achieve work goals (Bakker & Demerouti, 2017), resulting in poorer quality of care being provided. These findings align with the existing literature, although the use of presenteeism and work-related stress complements the existing research that has mainly focused on burnout (Q. Hu et al., 2011; Lewig et al., 2007) and depression (Loerbroks et al., 2016) as mediators.

For job resources, although job control had a direct relationship with all three quality-of-care variables, manager support only had a relationship with quality of trust care. It may be that the rotations by some hospital doctors, particularly those in the early stages of their career means they do not have the opportunity to build adequate relationships with their managers (McGowan et al., 2013). Moreover, the line management structure for doctors within hospitals may not be clear, and could refer to the medical director, senior consultants, administrative managers, or clinical supervisors (Kilminster & Jolly, 2000; Orman & Thornton, 2010). Regardless of the source of management, line managers may also be restricted by organisational policies and systems factors in their ability to influence the workload or resources made available to their staff (Zadow & Dollard, 2015). Therefore, manager support here may refer to the wider organisation, explaining why the outcome pertaining to care being provided by the trust was the only quality-of-care measure to relate with manager support. In comparison, errors seen and quality of individual care are more clearly linked to the doctor's own behaviours and local work environment. It is here where job control reflects the doctor's ability to manage their own work environment and would be expected to relate with doctors' ability to provide good quality care.

Despite this, the mediation analyses indicated that both job resources do impact quality of care by the trust and the individual, when mediated by work engagement. This suggests that while job control utilises both a direct and indirect effect to influence quality of care, manager support mainly does so indirectly. From a methodological perspective, this is congruent with early criticism of the four step mediation process that necessitated a direct relationship between the predictor and outcome (A. F. Hayes, 2009; MacKinnon et al., 2004). These findings support the JD-R model's (Bakker & Demerouti, 2017) motivation process, where job resources leads to better wellbeing through its ability to mitigate the negative effect of job demands, provide opportunities to cope with challenging situations, and obtain support and resources to achieve work goals (Bakker & Demerouti, 2017; Deci & Ryan, 1985; Shirom et al., 2006). Better work-related wellbeing, in turn, is associated with better quality of care.

Closer examination of the effect sizes show that most of these are considered small (J. Cohen, 1988). These small effect sizes are perhaps not surprising given the direct effects that job demands and resources had with quality of care, as well as the possibility of numerous other mediators that could explain these relationships. Example of these mediators from the systematic review (Chapter Four) include burnout, job satisfaction, and depressive symptoms (An et al., 2013; Loerbroks et al., 2016; Shirom et al., 2006; Weigl et al., 2015). Moreover, although beyond the role of this thesis, the personal resources (e.g., self-efficacy, optimism, and resilience) of doctors also likely mediate these relationships (Xanthopoulou et al., 2009; Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2007). While this appears to counter the full mediation that has been reported to occur within individual studies testing the JD-R model (Bakker, van Emmerik, et al., 2008; Schaufeli & Bakker, 2004), these findings are congruent with the small indirect effect sizes observed in meta-analytical reviews of the health-impairment and motivational process (LePine et al., 2005; Nahrgang et al., 2011). It is worth noting that three medium effect sizes were observed, two of which involved work engagement as a mediator and quality of individual care as the outcome. This suggests that quality of individual care, which out of the three outcome measures is the one which the individual has most influence over, may be more strongly influenced by the motivational process (i.e., via work engagement) than by the health-impairment process (i.e., via work-related stress or presenteeism). It also emphasises the importance of moving away from only focusing on negative work-related wellbeing and to consider positive manifestations of wellbeing as well (Scheepers et al., 2015).

Although the predicted relationships between job demands and resources with errors seen were observed, the only significant indirect effect involving this outcome was when presenteeism was a mediator. Extensive literature discusses the diversity of error definitions and how they are interpreted and reported by participants (Probst & Estrada, 2010; Rosenman et al., 2006). Not only are safety errors low occurring events, but they typically experience underreporting as workers fear reprisal (Probst & Estrada, 2010). Consequently, self-reported errors often obtain low frequency data that do not accurately reflect reality, and skew data that can be detrimental for analyses (Christian et al., 2009). The two items used here attempt to provide clarity through a definition provided to responders, while at the same time not attempting to assign any form of responsibility to any member of staff. However, some researchers (Probst & Estrada, 2010; Raleigh et al., 2009) have also noted that increased reporting of errors actually is a reflection of a mature safety culture. It could be argued that high self-reported errors actually are a good thing. For example, previous studies using the same error items from this study, albeit with an earlier dataset, observed that lower infection rates (M. West, Dawson, Admasachew, & Topakas, 2011) and better patient experience (Raleigh et al., 2009) were actually associated with more errors reported. Collectively, these inconsistencies may undermine the relationships involving errors seen, and explain why the relationships were not as consistent as they were with the other quality-of-care indicators.

9.5.2 Direct and indirect relationships at the trust-level

At the between-trust level, results indicated that as the number of emergency admissions increases, work engagement decreases, which in turn is associated with a reduction in doctor-rated quality of care by the individual and by the trust. It may be that an increase in emergency admissions places additional workload on doctors, not allowing them to utilise the job resources available to them, or to pursue their work in a meaningful and purposeful way. This hinders experiencing work engagement which restricts the provision of good quality care. The absence of an indirect effect involving errors seen as an outcome could be attributed to the issues with the measurement of it, as discussed in the previous section. Despite these findings, two further observations exist: that bed occupancy rates was not in any way a predictor on any of the quality-of-care measures; and that presenteeism was not a significant mediator. While it would be fair to conclude that the JD-R model may operate differently across different levels (Bakker & Demerouti, 2017), a more useful question to reflect is why this may be.

High demands on hospitals is associated with increased mortality rates (Boden et al., 2016; Madsen et al., 2014) and hospital infections (Kaier, Mutters, & Frank, 2012). This has been attributed to increased backlog of cases, overspill into other wards and departments, crowding, the postponement and cancellation of procedures, and the increased likelihood of staff taking shortcuts, all of which may mean patients not receiving the appropriate care when needed and patient safety being compromised (College of Emergency Medicine, 2014; Madsen et al., 2014; Schilling, Campbell, Englesbe, & Davis, 2010). The empirical evidence has not been completely consistent, as some studies have indicated no, or even opposite, relationships between hospital demands and quality of care (Kaier et al., 2012; Volpe, de Miranda Magalhães, & Rocha, 2013). Quality of care at the trust-level is extremely complex with numerous factors antecedent to it. Bed occupancy, as well as admission rates, could be proxies for, or confounded by, other factors that could be a more important predictor of quality of care (Boden et al., 2016; Kaier et al., 2012; Volpe et al., 2013). These include staff-patient ratios, staff safety behaviours, the presence of specialist wards, physical distribution of beds and wards, and quality of staff training. Focusing specifically on bed occupancy, it has been argued that as long as occupancy rates were within a designated range then it would not impact upon care outcomes (Borg, 2003); or, that operating near 100% bed occupancy can be an indication of efficiency and productivity (Madsen et al., 2014).

All the studies reviewed above examined quality-of-care outcomes through hospital-level quality-of-care statistics; however this study utilised three measures self-reported by doctors and measured at the individual-level. This assumes the existence of a relationship between self-reported measures of quality of care and hospital statistics, when there have been few attempts to test for this. Where this has been done the evidence has been inconclusive. For example, although nurse-reports of quality of care predicted lower death risks (Aiken, Clarke, Sloane, Lake, & Cheney, 2008), Howell et al. (2015) found that the number of self-reported patient safety incidents in England to not associate with hospital mortality and patient satisfaction. It may be that doctors' perception of the quality of care being provided is influenced by other individual factors, such as their own working conditions or work-related wellbeing. It is also likely that the hospital quality-of-care statistics are influenced by a myriad of organisational and individual-level factors that extend beyond just the doctors' psychosocial working conditions and work-related wellbeing (Krämer et al., 2016). Therefore, while this study provides some support that work-related wellbeing could mediate this relationship, numerous other factors exist that could also mediate this relationship. This serves to elucidate how complex the relationships between trust-level demands and quality of care are.

9.5.3 Limitations

There are a number of limitations that need to be acknowledged. Most of these limitations apply from the previous study in Chapter Eight and are reviewed in Section 8.5.4: the heterogeneity of doctors; single item measures; low internal reliability for workplace aggression, insufficient work resources, and number of errors seen; possible common method bias; and a cross-sectional design. In addition, the issue with using self-reported errors has already been discussed above (Probst & Estrada, 2010; Rosenman et al., 2006).

9.6 Conclusion

9.6.1 Study conclusion

The results in this chapter indicate that most hypotheses at the within-trust level were supported, while those at the between-trust level were not. At the within-trust level, the results are largely consistent with the JD-R model (Bakker & Demerouti, 2017; Demerouti et al., 2001), indicating that job demands and resources have both a direct and indirect effect on quality of care, with job demands operating through a health-impairment process (via work-related stress and presenteeism) and job resources through a motivational process (via work engagement). This demonstrates not only that hospital doctors' psychosocial work environment has an impact on their work-related wellbeing, but has implications for patient care as well. At the between-trust level, a high number of emergency admissions were associated with lower quality of care being provided, mainly through the mediating role of work engagement. The absence of findings involving errors or bed occupancy rates highlights the complexity in predicting quality of care at the trust-level.

9.6.2 Implications for thesis

This study provides the initial support that doctors' job demands and resources does relate with quality of care, and that their work-related wellbeing functions as a mediator. It also reinforces the utility of the JD-R model (Bakker & Demerouti, 2017; Demerouti et al., 2001) in explaining the mechanisms between them. However, the study raises a question on the suitability of doctors' self-reported quality of care. Therefore, the next study intends to replicate a similar framework as this study, but using quality-of-care statistics at the trust-level instead. Specifically, this will address issues related to common method variance and self-reporting, while at the same time linking with data that generate substantial public and political interest.

Chapter 10 : Trust Quality-of-Care Outcomes (Study 4)

Chapter Nine provided support for the job demands-resources (JD-R) model (Bakker & Demerouti, 2017; Demerouti et al., 2001) by demonstrating that doctors' psychosocial work environment has an impact on their work-related wellbeing, and towards their own ratings of patient care. This current chapter builds on Chapter Nine by replacing quality-of-care outcomes as rated by doctors, with trust outcome data collected within the NHS. More specifically, these are: the summary-level hospital mortality indicator, the number of patient safety incidents, and patient satisfaction with their doctors. The introduction describes the advantages of organisational data, before reintroducing the three core aspects of quality of care. The chapter then presents the study aims and hypotheses, followed by the study methodology, results, and discussion.

10.1 Introduction

Most of the hypotheses from Chapter Nine were supported, indicating that doctors' job demands and resources and their work-related wellbeing have an impact on three different self-rated quality-of-care outcomes. Although Section 7.6 demonstrated that common method bias was not likely to substantially influence this data, utilising quality-of-care outcomes from a different source renders any argument for common method bias moot (Podsakoff et al., 2003). Moreover, utilising trust quality-of-care outcomes potentially provides greater significance than doctor self-reported outcomes; this is as from a theoretical perspective, significantly less attention is paid towards linking the experience of healthcare staff at work with organisational-level quality-of-care outcomes (Pinder et al., 2013; Welp et al., 2015). Moreover, linking the NHS Staff Survey to other data collected by NHS agencies increases the utility of the dataset, and therefore its value and return-on-investment (Downs, Gilbert, Hayes, Hotopf, & Ford, 2017; NHS England, 2014). It further validates the dataset by potentially affirming anticipated relationships, and also allows it to operate as a proxy variable in future. As such, within the United Kingdom there are initiatives within the public sector to integrate and link public datasets; this would not only extract better financial value from them but also inform evidence-based policy making (Medical Research Council, 2014; Nuffield Council on Bioethics, 2015; Welcome Trust, 2014).

The Department of Health (2008; 2010) defines quality of care as consisting of three core aspects: clinical excellence, patient safety, and the experience of patients (see review in Section

3.1). In order to align the trust quality-of-care outcome measures used in this study with this definition, one proxy measure was selected for each of these three aspects. The first core aspect of quality in the NHS, clinical excellence, is defined as preventing premature deaths, enhancing quality of life, and assisting recovery (Department of Health, 2008, 2010). Hospital mortality was chosen as a proxy for this aspect; not only does this reflect clinical excellence but is widely used as a performance indicator (Howell et al., 2015; Topakas et al., 2010a, 2010c; Welp et al., 2015). The second core aspect: patient safety, aims to provide a safe care environment without avoidable harm (Department of Health, 2008, 2010). Here, there is increasing impetus in recording and studying patient safety incidents within the NHS (NHS National Reporting and Learning System, 2015); although, to date the evidence as to whether this measure is an appropriate reflection of quality of care is mixed (Howell et al., 2015). The final aspect of quality of care is patient satisfaction. This is despite the criticism as to its validity and reliability as a measure of quality of care (Coyle & Williams, 1999; Crow et al., 2002; Salisbury et al., 2010), including the absence of findings involving patient-rated outcomes in the systematic review in Chapter Four.

The evidence so far (reviewed in Section 3.3) indicates that few studies have tested the psychosocial working conditions and work-related wellbeing antecedents of mortality rates and patient experience, with none examining these antecedents with patent safety incidents as an outcome. Where these have been examined, the evidence has been mixed. For example, although Welp et al. (2015) found doctor and nurses' burnout to predict mortality in Swiss intensive care units, no significant findings were observed when work engagement, work-related stress, presenteeism, and general health were tested as predictors of mortality rates within NHS hospitals (Powell et al., 2014; Topakas et al., 2010a, 2010c). In terms of patient satisfaction, the Boorman Review (2009) reported that trusts with better wellbeing had on average higher rates of patient satisfaction. Similarly, work engagement and job satisfaction from the 2007 (Dawson, 2009), 2009 (Topakas et al., 2010a) and 2011 (Powell et al., 2014) NHS Staff Surveys positively predicted patient satisfaction. However, the same studies found work-related stress and presenteeism to not function as a predictor of patient satisfaction (Powell et al., 2014; Topakas et al., 2010c).

10.1.1 Study aim and hypotheses

The current study extends the study from Chapter Nine. It tests whether doctors' job demands and resources, and work-related wellbeing (i.e., work engagement, work-related stress, presenteeism), are associated with trust quality-of-care outcomes as postulated by the JD-R model (Demerouti et al., 2001). It directly responds to calls for more objective outcome measures to be used in validating the JD-R model (Demerouti & Bakker, 2011; Schaufeli & Bakker, 2004). As highlighted above, there has been a lack of studies testing patient safety incidents as an outcome measure, and mixed findings involving mortality rates and patient satisfaction. Despite this, it is anticipated that as per the JD-R model, better psychosocial working conditions and work-related wellbeing will be associated with better quality outcomes at the trust-level (Figure 10.1). Moreover, work-related wellbeing should mediate the relationship between job demands and resources with trust quality-of-care outcomes (Figure 10.2). More specifically, it is hypothesised that among hospital doctors:

> H_1: Insufficient work resources and workplace aggression (i.e., job demands) will positively predict hospital mortality and patient safety incidents, and negatively predict patient satisfaction.
>
> H_2: Manager support and job control (i.e., job resources) will negatively predict hospital mortality and patient safety incidents, and positively predict patient satisfaction.
>
> H_3: Work-related stress and presenteeism will positively predict hospital mortality and patient safety incidents, and negatively predict patient satisfaction.
>
> H_4: Work engagement will negatively predict hospital mortality and patient safety incidents, and positively predict patient satisfaction.
>
> H_5: Doctors' work-related stress and presenteeism will mediate the relationship between job demands (i.e., insufficient work resources, workplace aggression) with trust quality-of-care outcomes.
>
> H_6: Doctors' work engagement will mediate the relationship between job resources (i.e., manager support and job control) with trust quality-of-care outcomes.

Insufficient work resources predicted work engagement in the studies in Chapters Eight and Nine. Moreover, work engagement mediated the relationship between insufficient work resources with individual and trust quality of care. It is postulated that this relationship should exist when trust quality-of-care outcomes are used. Therefore:

H_7: Hospital doctors' work engagement will also mediate the relationship between insufficient work resources and trust quality-of-care outcomes.

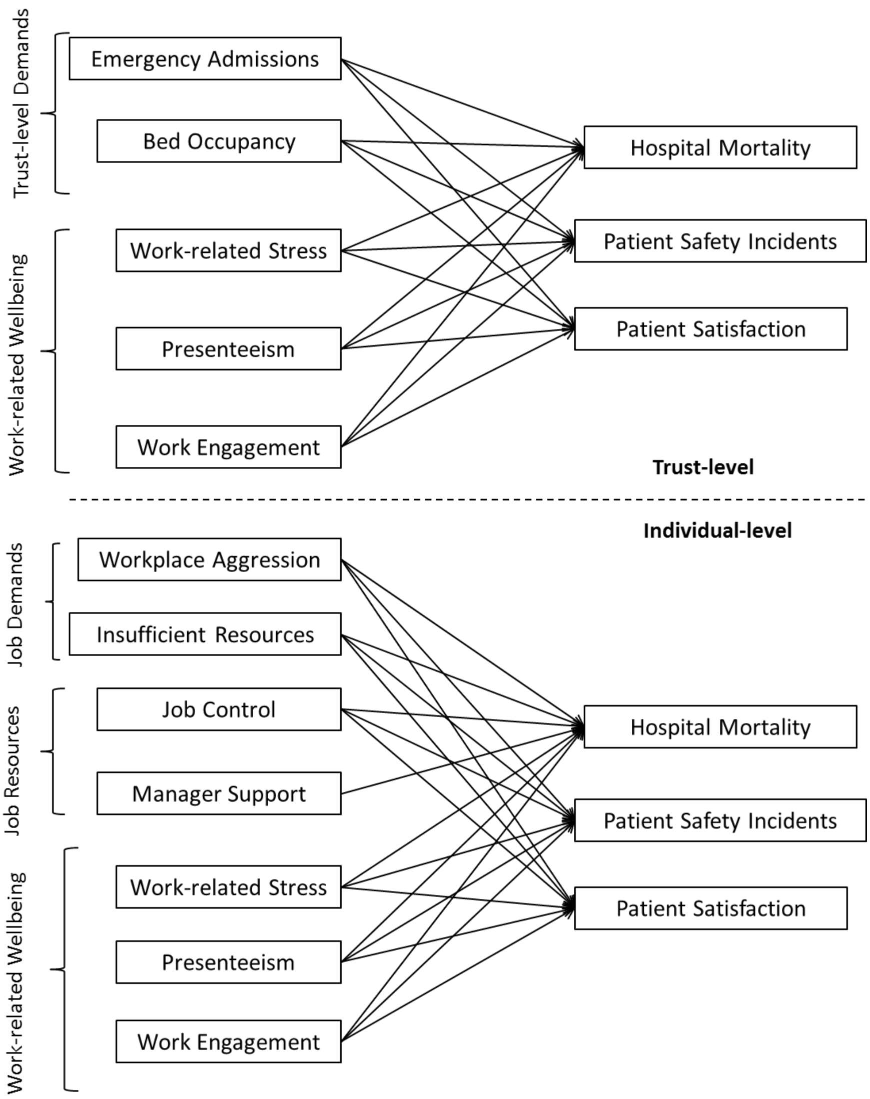

Figure 10.1: Direct relationships between working conditions and trust quality-of-care outcomes

The two preceding studies in this thesis demonstrated the importance of a systems perspective (Lowe & Chan, 2010), whereby trust-level predictors had an impact on doctors' work engagement and presenteeism. However, it had a limited impact on doctor-rated quality-of-care outcomes, and only one indirect effect was observed. This was where work engagement mediated the relationship between the number of emergency admissions and quality of individual care. Given that past research has indicated increased bed occupancy (Boden et al., 2016; Madsen et al., 2014) and number of hospital admissions (Aiken et al., 2008) are associated with increased hospital mortality rates, it is anticipated that both trust-level predictors would influence the three trust quality-of-care outcomes (Figure 10.1). Consequently, it is predicted that in line with the JD-R model, that trust-level demands should function in a similar manner to doctors' job demands (Schaufeli & Taris, 2014). Work-related wellbeing should also function as a mediator towards quality of care (Figure 10.2). However, Chapter Eight only found emergency admissions to predict presenteeism and work engagement, while bed occupancy predicted work engagement. Therefore, the following hypotheses are predicted:

> H8: The number of emergency admissions and bed occupancy rates will positively predict hospital mortality and patient safety incidents, and negatively predict patient satisfaction.
>
> H9: Emergency admissions' relationship with trust quality-of-care outcomes will be mediated by work engagement and presenteeism.
>
> H10: Bed occupancy rates' relationship with trust quality-of-care outcomes will be mediated by work engagement.

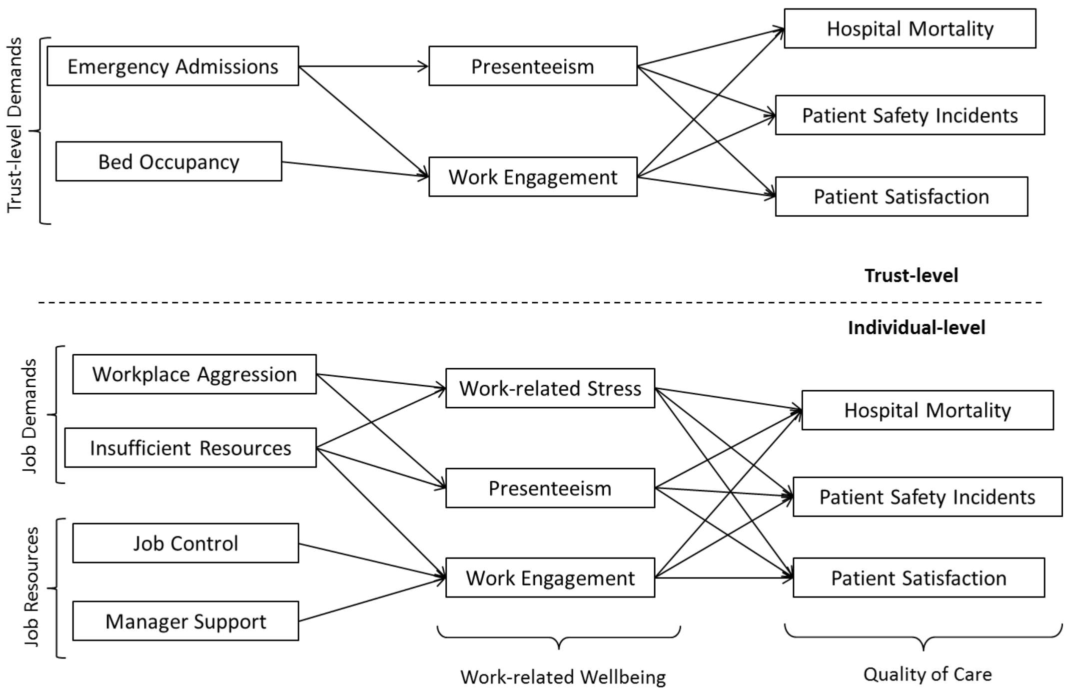

Figure 10.2: Work-related wellbeing as mediators between working conditions and trust quality-of-care outcomes

10.2 Method

10.2.1 Materials

The measures for job demands (insufficient work resources, workplace aggression), job resources (job control, manager support), and work-related wellbeing (work-related stress, presenteeism, work engagement), were all used and described in the preceding two studies, and during the creation of composite measures (Chapter Seven). The two trust-level measures (bed occupancy rates, number of emergency admissions) were previously described in Chapters Eight and Nine. As this study sample consisted only of acute trusts (see Section 10.2.2), the trust type control variable was redundant and excluded.

Quality-of-care measures. Three new measures that represented trust quality of care were included in this study. *Summary hospital-level mortality indicator* (SHMI) represented the ratio between the number of hospital patient deaths and the expected number of deaths based on the average in England, taking into consideration patient characteristics (NHS Digital, 2017). This includes deaths in hospital and within 30 days of discharge. The SHMI reflected the period between October 2014 and September 2015.

Patient satisfaction with doctors was available from the National Inpatient Survey 2015 (Care Quality Commission, 2016). This is an annual survey of inpatients aged 16 and above that spent at least one night in hospital in June 2015. Patients in psychiatric or maternity units were not surveyed. In total 149 trusts were included, yielding 83,116 replies (a 47% response rate). Patients responded to 72 questions on a ten-point scale, on 11 different areas relating to their stay, including: hospital and the ward; nurses; operation and procedures; waiting to get a bed on a ward; and overall experience. To more accurately reflect patients' experience with their doctors, only the three items relating to doctors were used for analyses (e.g., "when you had important questions to ask a doctor, did you get answers that you could understand?"). A higher score represented a more positive experience with doctors.

Patient Safety Incidents (PSI) were obtained from the NHS National Reporting and Learning System (NHS National Reporting and Learning System, 2015), which collects data on PSIs in England and Wales. A PSI is any "unintended or unexpected incident which could have, or did, lead to harm for one or more patients receiving NHS-funded healthcare" (pg. 1). These are reported locally, and are coded according to the incident type and the degree of harm. For the purpose of this analysis only the total number of incidents was used. The data here encompassed a six-month period from October 1st 2014 to March 31st 2015. The mean number of PSIs per month was 751.30 (*SD*=376.71). However, the variance of this measure exceeded the maximum variance allowed for analysis in Mplus; as such, it was divided by a 400 to reduce the scaling of the measure (L. K. Muthén & Muthén, 2017).

10.2.2 Sample

SHMI data was only available for acute trusts, and not acute specialist trusts. Similarly, not all acute specialist trusts participated in the 2015 National Inpatient Survey – resulting in missing data for six of the 18 acute specialist trusts. Consequently, the study sample was restricted to doctors (*n*=13,239) based at acute trusts (*n*=139). The mean doctors per trust was 82.55 (*SD*=48.75) with a median of 81 doctors.

10.2.3 Analysis

This section reviews three points pertaining to analysis. First is recognising the restricted sample size at the trust-level and the implications this has on analysis. Consequently, the second point covers the Monte Carle analyses carried out to assess the power of the proposed models. The third point covers the actual analysis procedures carried out in Mplus.

Trust-level sample size. The presence of the three trust-level quality-of-care outcomes meant that all the hypotheses specified functioned with between-trust variance components (Preacher et al., 2010). This represented a 1-1-2 mediation design for the hypotheses involving insufficient work resources, workplace aggression, job control, and manager support. Where trust-level demands were used, this was a 2-1-2 mediation design. As power at this level is weaker than at the individual-level (Neal & Griffin, 2006), this raises the issue whether a between-trust sample size of 139 is sufficient for the proposed analyses. When the group sample size is low, parameter estimates can be biased (McNeish, 2017). However, simulation studies have demonstrated that groups of at least 30 (Maas & Hox, 2005), 40 (Meuleman & Billiet, 2009), 50 (Hox et al., 2010), or 100 (Hox & Maas, 2001) should suffice for multilevel structural equation modelling.

For example, Meuleman and Billiet (2009) simulated analyses where an observed measure regressed onto a latent factor with four items at both the individual and between-group level. While they found that 40 groups would suffice when large effect sizes were anticipated, at least 100 were needed to detect small effects. They further state that more complex designs would require even more groups. Similarly, other rules-of-thumb used are reliant on balanced group sizes, robust estimators, and ICC values between .10 and .30 (Hox & Maas, 2001; McNeish, 2017; Preacher et al., 2010). McNeish (2017) reviewed 70 papers that used multilevel structural equation modelling and found that 90% of them did not have adequate sample size at the group level. Considering the practical difficulties and cost in increasing the number of groups, some studies have attributed their lack of significant findings at the between-group level to the lack of power (Niks, Gevers, De Jonge, & Houtman, 2016; Petrou et al., 2012). The simulations described above also only reflect direct effects, and to date there is little understanding of how group sample size affects mediation analyses within multilevel structural equation modelling (Preacher et al., 2010). In the first such simulation, McNeish (2017) observed that a group sample size of 100 could suffice for mediation when effect sizes were small, although this was based on a three variable design without any other predictor or control variable.

Monte Carlo. Recognising the issues highlighted in the previous paragraphs, Monte Carlo analysis was carried out in Mplus to test the power of individual parameters and guide sample size suitability (L. K. Muthén & Muthén, 2002). A series of duplications using population parameter estimates which averages the parameter values and standard errors was run, allowing the calculation of the extent to which the proposed population model is covered. The test model focused on the work engagement pathway, with three predictors (job control, manager support, insufficient work resources) and trust quality-of-care outcomes (SHMI, PSI, patient satisfaction). Model estimation was repeated 500 times using population parameter values drawn from this proposed model. Sample size was set as 13,344 with 139 groups of 96 participants. Although all 15 parameters at the between-trust level were within the acceptable 95% coverage range, nearly all of the parameter (14/15) and standard errors (12/15) indicated bias above the recommended 10% (L. K. Muthén & Muthén, 2002).

As it was not possible to increase the number of trusts within the model, the only alternative was to simplify the model by investigating multiple, smaller, concurrent models. A second Monte Carlo simulation was conducted using a simpler model. This simulated a model involving job control, work engagement, and the three quality-of-care outcomes (SHMI, PSI, patient satisfaction). Again all parameter estimates at the between-trust level were within the recommended 0.91 and 0.98 range. However, this time only two of the nine parameter estimates, and none of the standard errors, were biased by more than 10%. Despite this, power actually remained under Cohen's (1980) recommended level of .80 on seven of the nine parameter estimates (range: 0.37 to 0.63).

Analyses. The findings from the Monte Carlo analyses suggest that simpler models would be more appropriate to test the proposed between-trust level hypotheses. Consequently, after confirming the factor structure of the measures used within this study, seven separate models were proposed to test the individual pathways between job demands and resources with doctors' work-related wellbeing and the three trust quality-of-care outcomes. More specifically these seven models tested the following antecedents: (1) job control and work engagement; (2) manager support and work engagement; (3) insufficient work resources and work engagement; (4) insufficient work resources and work-related stress; (5) insufficient work resources and presenteeism; (6) workplace aggression and work-related stress; and (7) workplace aggression and presenteeism. After the direct effects were tested, this was followed by mediation analyses within each of the seven models. To test the trust-level demands an eighth model was tested involving both trust-level predictors (number of emergency admissions, bed occupancy rates), two work-related wellbeing measures (work engagement, presenteeism), and all three quality-of-care outcomes. The two work-related wellbeing measures were then tested as mediators.

Mplus 8 was used to conduct all analyses (L. K. Muthén & Muthén, 2017). In line with the previous two studies the WLSMv estimator was used. This was due to its ability to handle both ordinal and dichotomous data, and that it is a more conservative and robust approach compared to other estimators (Asparouhov & Muthén, 2013; Hox et al., 2010). However, the WLSMv estimator is vulnerable to produce biased estimates when dealing with missing data within multilevel modelling; as such, five datasets were imputed to replace missing data as per guidelines by Asparouhov and Muthén (2010). Wald chi-square test of parameter equalities were used to compare models. Trust-level predictors were grand-mean centred. As the number of patient safety incidents was expected to strongly correlate with the size of the trust, the number of hospital beds within the trust was included as a control variable for this outcome measure. Finally, bootstrapping (set at 20,000 at 95% confidence intervals) using Selig and Preacher's (2008) programme simulated the sampling distribution of the indirect effects.

10.3 Results

10.3.1 Descriptive results

Table 10.1 provides the *N*, means and standard deviations for the measures used in this study. It also presents the correlation matrix at both the individual and trust-level. At the individual-level all correlations were significant, with workplace aggression and insufficient work resources negatively correlated with work engagement, and positively with work-related stress and presenteeism. In contrast, manager support and job control both correlated negatively with work-related stress and presenteeism, and positively with work engagement. The same patterns of correlations for these measures were observed at the trust-level. However, only four significant correlations were found involving the trust-level measures (Table 10.1). More specifically, the number of emergency admissions correlated with insufficient work resources (r=.20), manager support (r=-.17), job control (r=-.24), and work engagement (r=-.17); while patient satisfaction negatively correlated with insufficient work resources (r=-.19).

10.3.2 Confirmatory factor analysis

The 1-1-2 mediation design utilises the between-trust variance components. However, none of the four predictor measures at the individual-level (i.e., insufficient work resources, workplace aggression, job control, manager support) has so far been modelled at the between-trust level. Consequently, a new multilevel CFA was conducted to confirm the factor structure of this study's measures.

The proposed CFA had a good fit as all fit indices surpassed recommended thresholds (RMSEA=.03; CFI=.99; TLI=.99; Byrne, 2012; Hu & Bentler, 1998). Chi-square (χ^2=3091.47; df=218; p<.001) was significant. At the individual-level all items, with the exception of AG1, surpassed the recommended threshold of .7 for standardised loadings. However, AG1 (estimate=.49) was just under the minimum acceptable standardised loading of .5 (Hair et al., 2014). At the between-trust level, all three workplace aggression items did not meet the minimum loadings. Therefore, it was decided to remove the measure altogether from this study.

A revised CFA without the workplace aggression items was conducted. It met all indicators of good model fit. RMSEA was .05, while CFI (.98) and TLI (.98) both exceeded the recommended .95. Chi-square was again found to be significant (χ^2=4500.21; df=142; p<.001). As seen in the tables below, all the items loaded strongly onto their constructs at both the individual (Table 10.2) and between-trust level (Table 10.3).

Table 10.1: *Descriptive statistics and correlations*

Measure	*N*	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Workplace aggression	12900	3.61	1.09	-	.19**	-.10**	-.11**	.21**	.18**	-.15**				
2. Insufficient work resources	13075	6.09	2.00	.30**	-	-.40**	-.40**	.29**	.21**	-.40**				
3. Manager support	12770	17.64	4.63	-.14	-.55**	-	.57**	-.25**	-.18**	.42**				
4. Job control	13019	14.44	3.67	-.18*	-.57**	.63**	-	-.25**	-.18**	.50**				
5. Work-related stress	12998	0.33	0.47	.29**	.54**	-.45**	-.36**	-	.29**	-.36**				
6. Presenteeism	11423	0.51	0.50	.26**	.26**	-.25**	-.24**	-.21*	-	-.21**				
7. Work engagement	13091	11.82	2.38	-.26**	-.54**	.59**	.48**	-.58**	-.33**	-				
8. Bed occupancy	139	89.88	5.62	.14	.07	.08	.08	-.04	.08	.10	-			
9. Emergency admissions	139	730.58	340.24	.09	.20*	-.17*	-.24**	.01	-.03	-.17*	.07	-		
10. SHMI	139	1.00	0.09	.02	-.15	-.01	-.02	-.10	.04	.10	.08	.02	-	
11. Patient safety incidents	139	73.8	847.94	.06	.06	-.05	-.02	.04	.03	-.02	.03	.02	-.07	-
12. Patient satisfaction	139	8.56	0.22	-.14	-.19*	.05	.08	.06	-.13	-.07	-.23	-.07	-.10	.01

Note. ***p*<.01; **p*<.05. Correlations above the diagonal are individual-level correlations. Correlations below the diagonal are trust-level correlations, with individual-level measures aggregated to the trust-level (*N*=139).

Table 10.2: *Standardised loadings for within-trust level items*

Latent Construct	Item	Estimate	S.E.	Est./S.E.	P-Value
Insufficient work	IR2	0.79	0.007	117.52	***
	IR3	0.78	0.007	119.36	***
Manager support	MS1	0.91	0.003	352.18	***
	MS2	0.91	0.002	397.54	***
	MS3	0.88	0.002	354.74	***
	MS4	0.87	0.003	317.15	***
	MS5	0.78	0.005	170.23	***
Job control	JC1	0.83	0.004	235.68	***
	JC2	0.90	0.002	385.04	***
	JC3	0.89	0.003	353.04	***
	JC4	0.88	0.003	313.17	***
Work engagement	EG1	0.91	0.003	310.86	***
	EG2	0.92	0.004	245.65	***
	EG3	0.70	0.005	149.67	***

Note. ****p*<.001

Table 10.3: *Standardised loadings for between-trust level items*

Latent Construct	Item	Estimate	S.E.	Est./S.E.	P-Value
Insufficient work resources	IR2	0.79	0.073	12.83	***
	IR3	1.00	0	999.00	999
Manager support	MS1	1.02	0.018	57.03	***
	MS2	0.97	0.025	38.50	***
	MS3	0.98	0.024	42.98	***
	MS4	0.99	0.02	49.72	***
	MS5	1.00	0	999.00	999
Job control	JC1	0.96	0.03	32.80	***
	JC2	0.99	0.023	44.43	***
	JC3	0.96	0.023	41.66	***
	JC4	1.00	0	999.00	999
Work engagement	EG1	0.86	0.067	13.92	***
	EG2	0.87	0.069	13.54	***
	EG3	1.00	0	999.00	999

Note. ****p*<.001

10.3.3 Direct effects of job demands and resources onto quality of care

Five separate models were analysed to examine the direct effect of job demands and resources and work-related wellbeing had on trust care outcomes. Each model had one job demand or resource, and one work-related wellbeing measure (Table 10.4). The same table also demonstrates that all five models demonstrated good fit, evident from their RMSEA, CFI, and TLI indices.

Table 10.4: *Fit indices for the direct effects onto trust quality-of-care outcomes*

Model	Measure	RMSEA	CFI	TLI	χ^2
1	Job control - Work engagement	0.06	0.99	0.97	3180.31***
2	Manager support - Work engagement	0.03	0.99	0.99	622.08***
3	Insufficient work resources - Work engagement	0.03	0.99	0.99	267.73***
4	Insufficient work resources - Work-related stress	0.01	0.99	0.99	32.09***
5	Insufficient work resources - Presenteeism	0.01	0.99	0.99	30.40***

Note. ***p<.001.

At the within-trust level the only relationships were between job demands and resources, and work-related wellbeing; however, quality-of-care outcomes only existed at the between-trust level. As seen in Table 10.5, work engagement was significantly predicted by all three psychosocial working conditions: job control (β=.57), manager support (β=.46), and insufficient work resources (β=-.53). Insufficient work resources also positively predicted doctors' work-related stress (β=.41) and presenteeism (β=.26).

Table 10.6 displays the direct effects involving trust quality-of-care outcomes. The number of beds per trust positively predicted the number of patient safety incidents in each of the five models. Levels of job control reported by doctors only predicted better patient satisfaction (β=.37), but not hospital mortality or the number of patient safety incidents. The relationship between insufficient work resources and trust quality-of-care outcomes was each tested three times. Here, insufficient work resources negatively predicted patient satisfaction twice (β=-.54; β=-.78) and patient safety incidents once (β=.13). Although no significant relationships were observed here, examination of the 95% confidence intervals suggests that a lack of power may be a contributing factor. Insufficient work resources did not predict hospital mortality rates. Some support was available for H_1, but not for H_2 and H_3. Manager support, presenteeism, and work-related stress did not predict any of the three trust outcome measures. H_4 was rejected as work engagement not only failed to predict hospital mortality and patient safety incidents, it surprisingly predicted patient satisfaction in the opposite direction than predicted (β=-.34; β=-.37; β=-.46).

Table 10.5: *Standardised and unstandardised coefficients for direct effects of job demands and resources onto work-related wellbeing*

Measure		Work engagement	Work-related stress	Presenteeism
Within-trust				
Job control	β	.57*** (.56, .59)		
	b	0.81*** (0.77, 0.85)		
Manager support	β	.46*** (.45, .48)		
	b	0.47*** (0.44, 0.46)		
Insufficient work resources	β	-.53*** (-.54, -.52)	.41*** (.38, .43)	.26** (.24, .28)
	b	-0.92*** (-0.98, -0.86)	0.44*** (0.40, 0.47)	0.25*** (0.21, 0.29)
Between-trust				
Job control	β	.64*** (.43, .85)		
	b	0.77*** (0.40, 1.13)		
Manager support	β	.69*** (.50, .87)		
	b	0.54*** (0.31, 0.76)		
Insufficient work resources	β	-.64*** (-.83, -.46)	.77*** (.59, .96)	.33** (.07, .58)
	b	-0.76*** (-1.08, -0.44)	0.36*** (0.21, 0.51)	0.14* (0.02, 0.27)

Note. ****p*<.001; ***p*<.01; **p*<.05. β = standardised beta coefficients; b = unstandardised beta coefficients; parentheses represent 95% confidence intervals.

Table 10.6: *Standardised and unstandardised coefficients for between-trust direct effects onto trust quality-of-care outcomes*

		SHMI		Patient Safety Incidents		Patient Satisfaction	
Model	Measure	*β*	b	*β*	b	*β*	b
1	Job control	-.23 (-.55, .09)	-0.09 (-0.22, 0.04)	.03 (-.17, .27)	0.14 (-0.69, 0.97)	.37* (.02, .72)	0.36* (0.02, 0.72)
	Work engagement	.18 (-.16, .53)	0.06 (-0.06, 0.18)	-.06 (-.27, .15)	-0.21 (-0.94, 0.53)	-.34* (-.68, -.01)	-0.28 (-0.57, 0.01)
2	Manager support	-.16 (-.53, .22)	-0.04 (-0.13, 0.84)	.10 (-.13, .32)	0.26 (-0.33, 0.84)	.38 (-.03, .80)	0.24 (-0.03, 0.51)
	Work engagement	.14 (-.25, .54)	0.05 (-0.08, 0.49)	-.11 (-.36, .14)	-0.36 (-1.20, 0.49)	-.37* (-.74, -.01)	-0.30 (-0.61, 0.02)
3	Insufficient work resources	-.20 (-.55, .15)	-0.07 (-0.19, 0.05)	.10 (-.09, .29)	0.39 (-0.34, 1.11)	-.54*** (-.84, -.23)	-0.48** (-0.77, -0.19)
	Work engagement	-.07 (-.44, .30)	-0.02 (-0.13, 0.09)	.03 (-.20, .26)	0.10 (-0.63, 0.82)	-.46** (-.77, -.14)	-0.35** (-0.61, -0.08)
4	Insufficient work resources	-.17 (-.75, .41)	-0.17 (-0.75, 0.42)	.26 (-.09, .60)	0.98 (-0.32, 2.08)	-.78* (-.1.43, -.14)	-0.72 (-1.32, -0.15)
	Work-related stress	-.02 (-.58, .55)	-0.01 (-0.46, 0.43)	-.22 (-.64, .20)	-1.81 (-5.37, 1.74)	.68 (-.04, 1.39)	1.33 (-0.20, 2.87)
5	Insufficient work resources	-.25 (-.52, .03)	-0.09 (-0.18, 0.01)	.13* (.02, .24)	0.47* (0.05, 0.89)	-.21 (-.42, .01)	-0.18 (-0.37, 0.01)
	Presenteeism	.18 (-.22, .58)	0.15 (-0.18, 0.48)	-.13* (-.28, .02)	-1.14 (-2.44, 0.17)	-.16 (-.48, .16)	-0.33 (-1.01, 0.35)

Note. ****p*<.001; ***p*<.01; **p*<.05. *β* = standardised beta coefficients; b = unstandardised beta coefficients; parentheses represents 95% confidence intervals; SHMI: Summary Hospital Mortality Indicator.

10.3.4 Indirect effects for job demands and resources

Each model subsequently had three mediation analyses added to it, involving a job demand or resource, a work-related wellbeing measure, and the three trust quality-of-care outcomes. In total, fifteen indirect effects were calculated; only one was observed to be significant (Table 10.7). Here, work engagement mediated the relationship between insufficient work resources and patient satisfaction (β=.59). This meant that H_5, H_6, and H_7 were rejected. None of the five models actually demonstrated significant change of chi-square on the Wald Test, which is partially reflected by the model fit indices being identical to those reported on in Table 10.4. It is worth noting that at the 90% confidence interval level, work engagement mediated the relationship between job control and manager support with patient experience with doctors; and work-related stress also mediated the relationship between insufficient demands and patient satisfaction.

Table 10.7: *Standardised and unstandardised coefficients for indirect effects between within-trust job demands and resources with quality of care*

Model	Predictor - Quality-of-care relationship	Unstandardised	Standardised
	Mediator: Work engagement		
1	Job control - SHMI	0.05 (-0.04, 0.14)	.44
	Job control - PSI	-0.16 (-0.72, 0.40)	.01
	Job control - PSat	-0.22† (-0.47, 0.04)	-.82
	Mediator: Work engagement		
2	Manager support - SHMI	0.02 (-0.05, 0.09)	.27
	Manager support - PSI	-0.19 (-0.66, 0.28)	.00
	Manager support- PSat	-0.16† (-0.34, 0.02)	-.72
	Mediator: Work engagement		
3	Insufficient work resources - SHMI	0.02 (-0.07, 0.10)	.10
	Insufficient work resources - PSI	-0.08 (-0.63, 0.48)	.01
	Insufficient work resources - PSat	0.26* (0.03, 0.49)	.59
	Mediator: Work-related stress		
4	Insufficient work resources - SHMI	-0.01 (-0.17, 0.16)	-.03
	Insufficient work resources - PSI	-0.66 (-1.94, 0.63)	.01
	Insufficient work resources - PSat	0.48† (-0.10, 1.07)	1.10
	Mediator: Presenteeism		
5	Insufficient work resources - SHMI	0.02 (-0.03, 0.07)	.12
	Insufficient work resources - PSI	-0.16 (-0.39, 0.07)	.01
	Insufficient work resources - PSat	-0.05 (-0.15, 0.05)	-.10

Note. *p<.05; †p<.10; β = standardised beta coefficients; b = unstandardised beta coefficients; parentheses represent 95% confidence intervals; SHMI: Summary Hospital Mortality Indicator; PSI: Patient Safety Incidents; PSat: Patient Satisfaction

10.3.5 Direct effects of trust demands onto trust quality of care

At the between trust-level, a new model was tested to investigate the effects of bed occupancy and emergency admissions on doctors' work-related wellbeing (work engagement, presenteeism) and the three trust quality-of-care outcomes. The model demonstrated good fit as RMSEA (.05), CFI (.99), and TLI (.97) all met the threshold for good model fit. Although chi-square was significant (χ^2=1632.90; df=24, p<.001), this was expected due to its susceptibility to large sample sizes (Markland, 2007).

Table 10.8 presents the direct effects that trust demands had on work-related wellbeing. Here, high emergency admissions predicted low work engagement (β=-.49) and high presenteeism (β=.60) amongst hospital doctors. Work engagement was also positively predicted by bed occupancy (β=.23). The predictors of trust quality-of-care outcomes are presented in Table 10.9, little support was found for H_8 as bed occupancy rate only predicted patient satisfaction (β=-.17), while the number of emergency admissions predicted hospital mortality (β=.41). The control variable of number of beds within the trust positively predicted the number of patient safety incidents (β=.81).

Table 10.8: *Standardised and unstandardised coefficients for direct effects of trust demands onto work-related wellbeing*

Measure		Work engagement	Presenteeism
Bed occupancy	β	.23* (.02, .44)	
	b	1.02 (-0.33, 2.37)	
Emergency admissions	β	-.49* (-.88, -.09)	.60** (.21, .99)
	b	-0.01 (-0.01, 0.01)	0.01* (0.01, 0.01)

Note. ***p<.001; **p<.01; *p<.05. β = standardised beta coefficients; b = unstandardised beta coefficients; parentheses represent 95% confidence intervals

Table 10.9: *Standardised and unstandardised coefficients for direct effects of trust demands onto trust quality-of-care outcomes*

	SHMI		Patient Safety Indicators		Patient Satisfaction	
Measure	*β*	b	*β*	b	*β*	b
Bed occupancy	.01 (-.19, .20)	0.01 (-0.32, 0.33)	.04 (-.06, .13)	0.61 (-0.93, 2.15)	-.17* (-.35, -.01)	-0.70* (-1.41, -0.02)
Emergency admissions	.41* (.03, .79)	0.01 (0.01, 0.01)	.10 (-.14, .34)	0.01 (-0.01, 0.01)	-.14 (-.57, .29)	-0.01 (-0.01, 0.01)
Work engagement	.10 (-.17, .37)	0.04 (-0.07, 0.15)	-.05 (-.21, .11)	-0.19 (-0.80, 0.42)	-.15 (-.37, .07)	-0.14 (-0.35, 0.07)
Presenteeism	.03 (-.39, .45)	0.01 (-0.30, 0.37)	-.15 (-.33, -.04)	-1.17 (-2.60, 0.27)	-.17 (-.56, .23)	-0.35 (-1.08, 0.37)

Note. ***p<.001; **p<.01; *p<.05. β = standardised beta coefficients; b = unstandardised beta coefficients; parentheses represent 95% confidence intervals; SHMI: Summary Hospital Mortality Indicator.

10.3.6 Indirect effects between trust demands and trust quality of care

In total nine mediation pathways were specified involving the two trust demands (bed occupancy, emergency admissions), two work-related wellbeing measures (work engagement, presenteeism), and three trust quality-of-care outcomes (SHMI, patient satisfaction, PSI). Model fit statistics was identical to the previous model (RMSEA=.05; CFI=.99; TLI=.97) and therefore retained good fit. However, the Wald Test revealed no significant change on chi-square of the indirect effect model, indicating that none of the nine indirect effects were significant. This meant H_9 and H_{10} were rejected.

10.4 Study Discussion

The results demonstrate that the five predictors (insufficient work resources, job control, manager support, number of emergency admissions, bed occupancy rates) had little impact on the three trust quality-of-care outcomes. Only four out of 15 possible relationships were found. Patient satisfaction with their doctor was influenced by doctors' level of perceived insufficient work resources and job control, as well as trust bed occupancy rates. Insufficient work resources also negatively predicted patient safety incidents. By extension, the lack of findings between the predictors and quality-of-care outcomes meant that only one significant mediation was observed. Despite this, the hypothesised relationships between the five predictors and the three work-related wellbeing measures were still observed. These findings are surprising. Three possible explanations exist as to why the anticipated relationships for quality of care were not observed: distal predictors and outcomes, the validity of trust outcome measures, and statistical power. These are discussed in further detail below.

10.4.1 Distal predictors and outcomes

The first possible explanation to consider is that the three trust quality-of-care outcomes were too distal to accurately reflect changes within the predictors. As the distance between variables grows further apart, there is an increase in the number of competing causes, links in the causal chain, and other random factors that influence the relationship between the two variables of interest (Shrout & Bolger, 2002). The outcome measures - hospital mortality, patient satisfaction, and patient safety incidents - all represent organisational outcomes that are distal to the individual doctors' perception of their work environment and work-related wellbeing. They are influenced by numerous other factors beyond the experiences of one occupational group, including for example: patient characteristics, the experiences of other professional groups, staffing, senior leadership, quality of services, local demographics, and political factors (Powell et al., 2014; Taris, 2006). Consequently, it has been argued that hospital mortality is too blunt a measure to represent quality of care (Bottle et al., 2011) and that it is not sensitive enough to relate with staff wellbeing (Powell et al., 2014). This could also explain the mixed findings in the literature involving antecedents to mortality. In the studies where healthcare staff wellbeing preceded mortality (e.g., Welp et al., 2015), the outcome measure was at the unit level; in contrast, those that did not detect such a relationship examined mortality at the hospital level (e.g., Powell et al., 2014; Topakas et al., 2010a, 2010b).

Despite the JD-R literature advocating further research involving objective and distal outcomes to validate the theory (Schaufeli & Bakker, 2004), to date most of this research has not focused on organisational outcomes. While relationships have been established with objective measures such as sickness absence records (Schaufeli, Bakker, & Van Rhenen, 2009) and team outputs (Costa, Passos, & Bakker, 2015), these relationships have smaller effect sizes than with more proximal measures. Even when Xanthopoulou et al. (2009) demonstrated that fast food workers' work engagement had a relationship with same and next day financial performance, this outcome measure represented shift financial performance that involved the performance of four employees. This could still be considered a proximal outcome compared to the trust outcome measures here that reflect the collective performance of hundreds, if not thousands, of healthcare workers within hospitals. Therefore, although theoretically it is likely that doctors' job demands and resources and work-related wellbeing would influence trust performance, in practice this distal outcome would likely present as a small effect size.

This does not, however, explain why the only relationships involving trust-level demands involved the number of emergency admissions and hospital mortality, and bed occupancy with patient satisfaction. It is also plausible that these relationships at the trust-level are confounded by numerous other factors, including staff-patient ratios, staff safety behaviours, the presence of specialist wards, physical distribution of beds and wards, local deprivation, and national policies (Boden et al., 2016; HSCIC, 2015b; Kaier et al., 2012; Volpe et al., 2013). It has been postulated that bed occupancy rates within a designated range should not impact on care outcomes as sufficient resources exist to address the demands faced (Borg, 2003). This may explain the absence of a relationship between bed occupancy with both hospital mortality and patient safety incidents. In terms of patient satisfaction, research involving emergency room crowding has shown patient satisfaction to be inversely related with crowding (Hillier, Parry, Shannon, & Stack, 2009; Pines et al., 2008). This is attributed to longer wait times to be seen, treated, and admitted. Therefore, while bed occupancy has limited impact on clinical outcomes it could impair patents' experience of their care.

As anticipated, the number of emergency admissions positively predicted hospital mortality. This could be explained by the higher workload that doctors face when the number of emergency admissions increase. In turn this can lead to a backlog of cases, the postponement and cancellation of procedures, overspill into other wards and departments, and the increased likelihood of staff taking shortcuts, all of which may compromise the quality of care being delivered (College of Emergency Medicine, 2014; Madsen et al., 2014; Schilling et al., 2010). The absence of relationships between the number emergency admissions with patient safety incidents and patient satisfaction may be a function of the validity of these outcome measures, and/or issues with statistical power. Both of these are reviewed in the sections below.

10.4.2 Validity of trust outcome measures

Although care was taken to select the most appropriate trust quality-of-care outcomes, in reality there are concerns about all three measures' validity. This may have biased any analyses involving them. The voluntary reporting of patient safety incidents is subjected the same issues of safety error reporting that was discussed in Section 9.5.1. Not only are these typically infrequent events, but they are often underreported as staff fear reprisal (Probst & Estrada, 2010). Alternatively, rather than being an indicator of poor safety, high error reporting has also been observed to reflect a mature safety culture (Raleigh et al., 2009). For example, Howell et al. (2015) found that staff safety initiatives and confidentiality around error reporting had moderate and positive correlations with patient safety incident reporting, suggesting that high number of incidents could be a reflection of better safety culture. To date this is the only known study to have used NHS patient safety incidents as an outcome measure to wellbeing or psychosocial working conditions. However, patient safety incident rates in Howell et al.'s study did not correlate with patient satisfaction or mortality rates; consequently, the researchers there concluded that using these incidents as a measure of patient safety or trust quality is likely to be inaccurate.

Hospital mortality data is routinely collected and therefore can be susceptible to mistakes (Howell et al., 2015). Even within its calculation there is disagreement as to how deaths are coded and what factors should be included or excluded. For example, it has been argued that not-for-resuscitation and palliative care deaths should be exempt from calculations when these were the main admissions reasons (Bottle et al., 2011), or that deprivation should be included as a control factor (HSCIC, 2015b). Similarly, the political and performance implications of hospital mortality means that these can be vulnerable to adjustments that present more favourable standards (Bottle et al., 2011). Moreover, as the mortality indicator used here is a comparative indicator, changes in patient outcomes in some trusts would impact mortality scores at other trusts as well (Boden et al., 2016). These limitations mean that while some believe that hospital mortality is a useful indicator of quality of care being delivered (Howell et al., 2015; HSCIC, 2015b), others find it to be neither a helpful or informative metric (Hogan et al., 2015).

Similarly, the issue of whether patient satisfaction is a valid proxy for quality of care was reviewed not only in Section 3.1.1 but was also a key discussion point in the earlier systematic review (Section 4.5.3). The concern with patient satisfaction measures lies in the difficulty conceptualising what this represents, and its poor links with other more objective quality measures (Crow et al., 2002; Salisbury et al., 2010). Despite these concerns, patient satisfaction with their doctors actually was significantly correlated with doctors' levels of job control and perception of insufficient demands, as well as trust bed occupancy rate. While these findings are expected, surprisingly doctors' work engagement was negatively associated with patient satisfaction. This counters the positive relationships where patient satisfaction was the outcome measure (Powell et al., 2014; Topakas et al., 2010a). The rationale for this finding is not clear, particularly as the previous study also reported a positive relationship between doctors' work engagement and self-rated quality of care. There is no explanation for this theoretically, and is something that warrants future examination.

Ultimately, these findings collectively raise questions about the trust quality-of-care outcomes used, and in particular the use of hospital mortality and patient safety incident data. While some studies support these outcome measures as proxies for quality of care, others do not. What this suggests is that the validity of these measures is reliant on the context of the study. Therefore, rather than outright accepting or rejecting the suitability of these measures, a more appropriate intervention should consider when and where is it right to use these outcome measures.

10.4.3 Statistical power

The third explanation with regards to the lack of hypotheses supported revolves around the possible lack of statistical power. This was identified as a potential issue in Section 10.2.3 above due to there only being 139 acute trusts in the country. Recognising the distal trust outcome measures, this would have required strong power to be able to find a significant relationship. Although a simpler model was specified based on the Monte Carlo analyses, the analyses also revealed the recommended 80% level for high power (S. Cohen, 1980) was not achieved. In fact, power level was only between 35% and 55% for hospital mortality, 10% and 60% for patient safety incidents, and 60% and 80% for patient satisfaction. This difference in power levels may also explain why most significant relationships were reported in relation to patient satisfaction. Similarly, although only one significant indirect effect was found, if a 90% confidence interval (rather than 95%) was used it would have presented three additional significant indirect effects. This links in with similar issues involving transnational research which faces restrictions by the number of countries in the world (Meuleman & Billiet, 2009). Post-hoc modelling involving a single predictor, mediator and outcome revealed little change to the findings observed here, although a few additional significant observations were observed. This implies that in addition to developing less complex multilevel models, researchers should reflect on the implications that distal variables may have on the effect size and in turn, the statistical power that is needed.

10.4.4 Limitations

The three discussion points above could be construed as limitations to be addressed in future research: distal predictors and outcomes, concerns about the validity of trust outcome measures, and the lack of statistical power. These are in addition to the limitations pertaining to the heterogeneity of doctors, low internal reliability for some measures, and the use of single item measures that remain from the studies in Chapter Eight and Nine (see Section 8.5.4). One further limitation that warrants acknowledgement is the timing of when and for how long outcome measures were recorded for. The 2014 NHS Staff Survey was a cross-sectional survey collected between October and December 2014. Appropriate trust outcome measures therefore had to either be from within this period of time, or after it. However, the three outcome measures represented different time durations. SHMI reflected the entire year between October 2014 and September 2015 (NHS Digital, 2017). Patient safety incidents were recorded over a six-month period from October 2014 to March 2015 (NHS National Reporting and Learning System, 2015). In contrast, the 2015 National Inpatient Survey was limited to patients who spent at least one night in June 2015 (Care Quality Commission, 2016). These different timings, both in terms of duration and time of the year mean that the data could be vulnerable to seasonal changes. For example, the seasonal demands on the NHS over the winter months place additional pressures on the system and the staff within it (Boden et al., 2016). These winter months map squarely onto the period in which patient safety incidents were recorded for this study, while in the case of mortality, the demands of these winter months are balanced out across the remaining months in the year. However, both the national inpatient and staff surveys are outside these demanding winter months. As such, it is plausible that the data from these latter surveys would change had they covered the same annual period as hospital mortality, which in turn may have yielded very different results altogether.

10.5 Conclusion

The findings from this study did not support the propositions from the JD-R model (Bakker & Demerouti, 2017; Demerouti et al., 2001). The findings indicate that as expected, insufficient work resources, manager support, job control, bed occupancy rates, and number of emergency admissions predicted the doctors' work-related wellbeing (work-related stress, presenteeism & work engagement). However, most of these did not predict quality of care at the trust-level, which was measured through the number patient safety incidents, hospital mortality, and patient satisfaction. Not only was the absence of these findings incongruent with the JD-R model, but the lack of findings involving indirect effects meant there was no support for the model's health-impairment and motivational processes. However, these shortcomings may not be due to the JD-R per se, but rather due to the study design. The trust outcome measures used has been the subject of past concerns about their validity as a proxy of quality of care, while their presence as distal outcomes means effect sizes were always likely to be small. This makes the restricted sample size at the trust-level even more important. It is unlikely that any one of these factors alone explain the lack of support for the study hypotheses. Instead, vulnerabilities on any one aspect accentuate the shortcomings in the other aspects. These implications of these findings for the thesis aims, research, practice, and policy, are reflected on in the next, and final, chapter.

Chapter 11 : Discussion

This final chapter brings together the previous ten chapters, including the four studies within this thesis. It relates these back to the original aims and research questions posed but does not repeat the individual discussion sections from each study. Instead, it serves to: (i) reiterate the thesis aims; (ii) provide a brief overview of the systematic review and its implications for this thesis; (iii) discuss the relevance of the job demands-resources (JD-R) model to the context of hospital doctors in England; (iv) consider the key implications for policy, practice, the NHS Staff Survey, and secondary data research; and (v) reflect on the key limitations and possible future research directions.

The aim of this thesis was threefold. The first was to enhance the understanding of the relationship between psychosocial working conditions of hospital doctors in England and quality of care. Second, to examine this relationship within a theoretical framework, namely the JD-R model (Bakker & Demerouti, 2007; Demerouti et al., 2001). Finally, the third aim was to link existing sources of data collected within the healthcare sector. This included both trust-level data that were operationalised as organisational demands and quality-of-care outcomes.

11.1 The Systematic Review and Research Questions

To address the first thesis aim, the systematic review and meta-analysis in Chapter Four explored what psychosocial working conditions faced by doctors have been examined in the existing literature, and what impact these have on different types of quality-of-care outcomes. From the 21 studies found, it was evident that this relationship was not clear, and that complex differential effects existed. Although most of the studies examined showed that aspects of job demands and resources predicted quality of care, these pertain mainly to clinical excellence and patient safety, and not patient experience. This highlighted three key limitations within the existing literature.

First, there have been few attempts to frame the relationship between doctors' psychosocial working conditions and quality of care within a theoretical framework. Only five of the studies found in the review did so. Theoretical frameworks are useful to explain the relationships, and to account for other factors such as confounders, moderators, mediation, and curvilinear effects. A better theoretical understanding also helps bridge the research and practitioner divide to inform and structure interventions. Second, none of the studies considered clinical outcomes, with the focus being on attitudinal and affective outcomes. Third, all these studies only functioned at the individual-level. This meant that events and outcomes at the hospital (or trust) level were not recognised. The latter two points are particularly important in approaching the issue of doctors' psychosocial working conditions and patient care from a systems perspective, evidenced by the system failures identified at Mid-Staffordshire and Morecambe Bay Trusts (Francis, 2013; Kirkup, 2015)

Recognising the issues identified in the systematic review, Chapter Five introduced the JD-R model (Bakker & Demerouti, 2017; Demerouti et al., 2001). It distinguished between job demands and job resources, explained its theoretical propositions, and provided evidence from other occupational samples. Consequently, in trying to examine the JD-R framework within a sample of doctors from English hospitals, the six research questions below were asked.

i. Do hospital doctors' job demands uniquely predict negative work-related wellbeing; and do job resources uniquely predict positive work-related wellbeing?
ii. Will hospital doctors' job resources moderate the relationship between job demands and negative work-related wellbeing?
iii. Will hospital doctors' job demands moderate the relationship between job resources and positive work-related wellbeing?
iv. Does work-related wellbeing mediate the relationship between hospital doctors' psychosocial working conditions and quality of care provided?
v. Will trust-level demands have the same impact within the JD-R as that of hospital doctors' job demands?
vi. Will hospital doctors' psychosocial working conditions and work-related wellbeing predict trust-level quality-of-care outcomes?

11.2 The Validity of the Job Demands-Resources Model among Hospital Doctors

Despite the growing popularity of the JD-R model, there has been little attempt to examine its key propositions among doctors as an occupational sample. To meet the three thesis aims, the JD-R model's propositions were tested within a sample of hospital doctors in England, and extended to integrate a multilevel perspective. This was done using the 2014 NHS Staff Survey, as well as existing data sources from the Care Quality Commission and different NHS agencies. Figure 11.1 illustrates the hypothesised model. Overall, the three studies in Chapters Eight, Nine, and Ten were able to test aspects of this model. The dual processes and the mediating role of work-related wellbeing were supported. Interactions between job demands and job resources were not. The relationships involving trust-level demands yielded different results than what was anticipated. The figures below allow the comparison of the hypothesised and final figures.

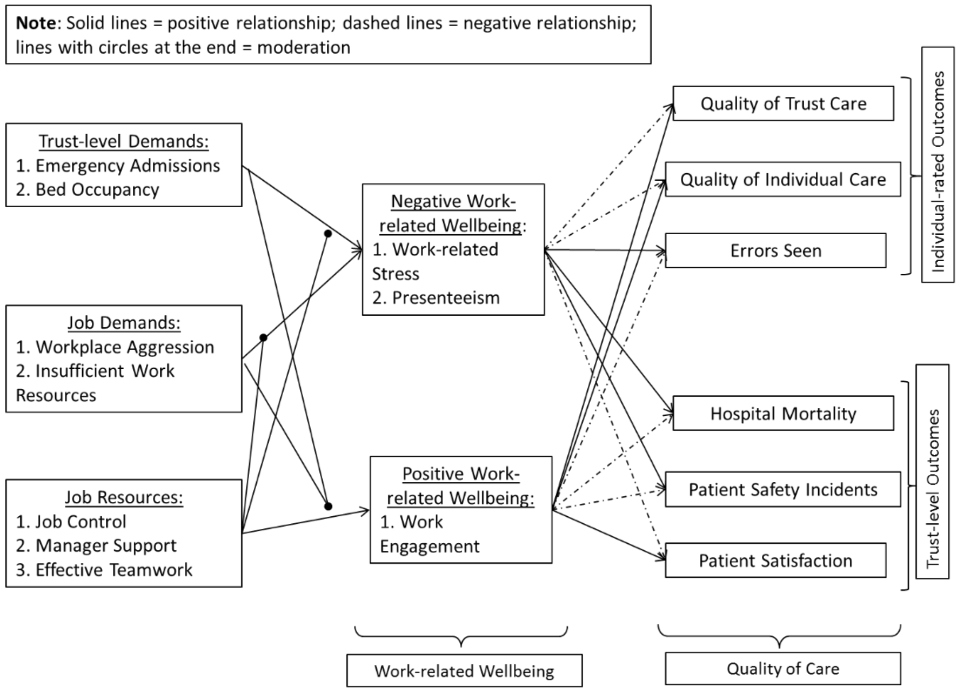

Figure 11.1: Hypothesised relationships between psychosocial working conditions, work-related wellbeing, and quality of care as proposed by the JD-R model

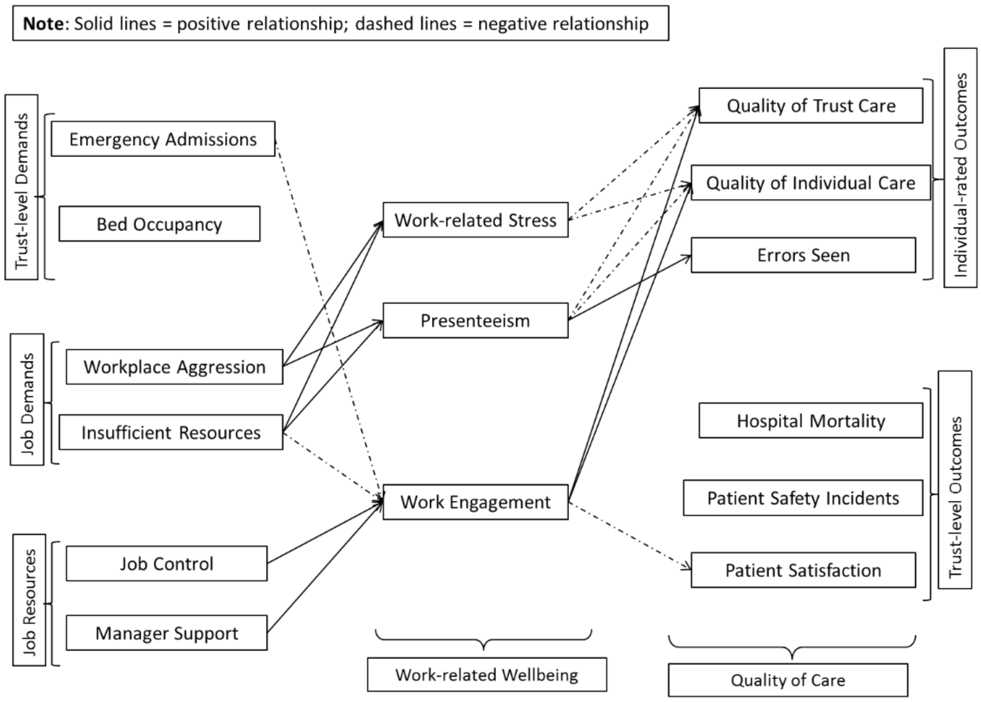

Figure 11.2: The mediated relationship found between psychosocial working conditions and quality of care

11.2.1 Dual processes

The first research question centred on the JD-R model's dual processes, and asked whether hospital doctors' negative work-related wellbeing is uniquely predicted by job demands, while positive work-related wellbeing is uniquely predicted by job resources. The studies in Chapters Eight to Ten found support for this. Hospital doctors' presenteeism and work-related stress were predicted by insufficient work resources and workplace aggression as per the health-impairment process. The motivational process was only partially supported, as manager support and job control, but not effective team practices, predicted work engagement. Moreover, a cross-path was observed where insufficient work resources predicted work engagement.

The presence of cross-paths. That insufficient work resources predicted work engagement is contrary to the JD-R model's proposed independent dual processes (Demerouti et al., 2001). However, the finding is consistent with past meta-analyses (LePine et al., 2005; Nahrgang et al., 2011) and longitudinal evidence (Hakanen et al., 2008) demonstrating a relationship between job demands and work engagement. This suggests that these pathways may not be completely independent and that some cross-paths may still occur. These cross-paths have been attributed to the distinction being made between challenge and hindrance demands (LePine et al., 2005), with the negative relationship observed here suggesting that insufficient work resources operates as a hindrance demand.

Doctors who perceive insufficient staffing and material resources likely interpret this as a barrier to achieving their work goals. This would, in turn, undermine their need for competence and autonomy, which are two of the three pillars of internal motivation according to self-determination theory (Deci & Ryan, 1985). Moreover, when faced with hurdles such as not having what is needed to achieve work goals (i.e., insufficient work resources), then workers are more likely to not engage cognitively or emotionally and instead utilise more emotion and passive-focused coping styles such as withdrawal and decreased engagement (Crawford et al., 2010; Harter et al., 2002). This relationship can also be explained using the social exchange theory (Blau, 1964), whereby the provision of adequate staffing and material resources is reciprocated with feelings of motivation and engagement. Empirical support that insufficient work resources is a hindrance demand is evident in Crawford et al.'s (2010) meta-analysis, where analyses showed that resource inadequacy was negatively correlated with work engagement.

The saliency of insufficient work resources. Aside from demonstrating a cross-path, insufficient work resources was the strongest predictor of work engagement in Chapter Nine, and a stronger predictor than manager support in Chapter Eight and Ten. This is in addition to this measure being the strongest predictor on six other outcome measures: work-related stress, presenteeism, quality of individual care, quality of trust care, trust-level patient safety incidents, and trust-level patient satisfaction. Collectively, these findings highlight the saliency of this job demand within the work environment. The provision of staff and material in the workplace can be a reflection of wider organisational demands (van Oostveen, Mathijssen, & Vermeulen, 2015). They are also likely to have an impact on other job demands and resources, such as workload, social support, and autonomy (Laschinger et al., 2012; M. C. W. Peeters & Le Blanc, 2001).This can be explained by psychosocial safety climate, where adequate staffing and material resources is a crucial manifestation of top management support and commitment, and is a precursor to job demands and resources (Dollard & Bakker, 2010). Consequently, insufficient work resources presents as a dominant psychosocial working condition with strong influence on nearly all outcomes.

The relevance of the dual processes. If work-related wellbeing extends beyond just negative wellbeing measures such as burnout and work-related stress, then it is important to understand what the antecedents to positive work-related wellbeing (e.g., work engagement) are. Although the results here suggest that insufficient work resources also predict work engagement, job resources (such as, manager support and job control) are imperative as well. These findings reinforce the argument that interventions should not only attempt to reduce job demands but to strengthen job resources in the workplace (Knight et al., 2017; K. Nielsen et al., 2017).

11.2.2 The psychosocial working conditions and quality-of-care relationship

Two chapters sought to test whether work-related wellbeing predicted quality of care. In Chapter Nine, all three work-related wellbeing measures predicted quality of trust and individual care, while presenteeism also predicted the number of errors seen. Work-related stress was the stronger predictor of quality of trust care while work engagement was the stronger predictor of quality of individual care. Despite the findings with self-reported care outcomes, Chapter Ten showed that the work-related wellbeing had a minimal impact on trust-level outcomes. The only consistent relationship was where work engagement negatively correlated with patient satisfaction. What this suggests is that the relationship between work-related wellbeing and quality of care depends on how these outcomes are measured. Hence, if certain quality-of-care outcomes have a stronger relationship with positive (e.g., quality of individual care with work engagement) or negative (e.g., quality of trust care with work-related stress) work-related wellbeing, then the dual processes provides value in anticipating the possible predictors of these work-related wellbeing states.

These points link in with the fourth research question, on whether work-related wellbeing mediates the relationship between doctors' psychosocial working conditions and quality of care provided. The answer to this question depends on the outcome measure employed, with work-related wellbeing operating as a mediator for quality of individual and trust care (Chapter Nine) and patient satisfaction with their doctor (Chapter Ten). This is congruent again with the health-impairment and motivational processes, where positive wellbeing is associated with higher performance; while the converse is observed with negative wellbeing. Despite this, mediational effects were not observed when the number of errors seen (Chapter Nine), and hospital mortality and the number of adverse events (Chapter Ten) were outcome measures.

In the same way that Bakker and Demerouti (2017) attributed the absence of cross-paths within the dual processes to suboptimal research designs, the absence of mediational effects for these specific outcome measures may be a reflection of methodological challenges. For example, Chapter Nine acknowledged the difficulty in measuring self-reported errors in the workplace. While high occurrence of errors could be a sign of an unsafe work environment, some workers underreport due to fear of reprisals (Probst & Estrada, 2010; Rosenman et al., 2006). Contradictory research highlights that increased reporting of errors can be a representation of a mature safety culture (Probst & Estrada, 2010; Raleigh et al., 2009). For hospital mortality and patient incidents, Chapter Ten attributed the absence of indirect effects to these trust-level outcomes being too distal to predicting and mediating constructs. Therefore, hospital mortality might be too blunt a measure to represent quality of care due to the increase in the number of competing causes, links in the causal chain, and other random factors that influence the relationship it has with its antecedents (Bottle et al., 2011; Shrout & Bolger, 2002). The same applies to patient safety incidents, which not only functions as a distal outcome, but is based on staff self-reported safety incidents.

The relevance of trust-level outcomes. If trust-level outcomes are too distal an outcome to be influenced by hospital doctors' work-related wellbeing, this could undermine any argument to enhance psychosocial working conditions or work-related wellbeing in order to improve quality of care. There are two points that counter this notion. First, it is important to highlight that hospital doctors' work-related wellbeing did predict one trust-level outcome - patient satisfaction with their doctor. This is not surprising, as among the three trust-level outcomes this was the only one which focused specifically on hospital doctors. As mentioned in Section 10.5.1, more proximal measures are better suited in demonstrating a significant relationship. This could be by using patient satisfaction with doctors rather than with their experience as a whole, as was done in this thesis; or, by examining mortality rates at the unit-level rather than the trust-level (Welp et al., 2015).

The second point to consider is that it would be somewhat naïve to assume that the work experiences of hospital doctors will alone influence trust-level outcomes. This is particularly as hospital doctors are the smallest occupational group within the NHS (NHS Staff Survey Co-ordination Centre, 2015). As a professional group, hospital doctors have been found to have more control and autonomy in their work, are more influential, and better remunerated than other professional groups (Chou, Li, & Hu, 2014; Escribà-Agüir, Martín-Baena, & Pérez-Hoyos, 2006; Lambrou, Kontodimopoulos, & Niakas, 2010). It could be that the experiences of nurses are more influential on patient care as they are a larger occupational group with substantially more patient contact than doctors (Bae, 2011; Buchan et al., 2016). Nevertheless, understanding how doctors perceive their work environment is still crucial in developing a comprehensive picture of working conditions within hospitals. After all, studies involving healthcare workers in general have found work-related wellbeing to predict hospital mortality rates (Virtanen et al., 2009; Welp et al., 2015), infection rates, and trusts' financial performance and absenteeism rates (Topakas et al., 2010a, 2010c). This only serves to reinforce the importance of systems thinking, where changes need to target multiple aspects of the work, including different professional groups.

11.2.3 When job demands and resources do not interact

The second and third research question focused on whether job demands and resources interacted with each other in relation to work-related wellbeing. Chapter Eight showed that not one of the hypothesised interactions was found. This was contrary to the JD-R model (Bakker & Demerouti, 2017). Job resources did not buffer the negative impact that job demands have on work-related stress or presenteeism. Neither did job demands enhance the strength of job resources' relationship with work engagement. Section 8.5.1 reviewed these findings in relation to challenge and hindrance demands, and the need for job demands and resources to match. Without repeating the discussion there, the implications for this in relation to the JD-R model are reviewed here.

The absence of interactions relates to the notion that the relevance of the job demand and resource is dependent on specific job characteristics that prevail within that organisational context (Bakker & Demerouti, 2007). For example, in the two studies known to have tested this interaction in a medical setting, none of the job resources that interacted with job demands were used in this thesis (Bakker et al., 2011; van Vegchel et al., 1999). These job resources were: development opportunities, feedback, supervisory coaching, and participation. Both these studies found that job autonomy, which has been argued to be conceptually similar to job control (M. A. G. Peeters & Rutte, 2005), did not fully interact with job demands. More specifically, job autonomy did not buffer high job demands' (work overload, emotional, and cognitive demands) impact on medical residents' work-home interference.

Similarly, Van Vegchel et al. (1999) found that job autonomy only buffered the influence of psychological, but not physical or emotional demands. This reinforces the notion that the type of job demands matters as well. Insufficient work resources and workplace aggression both correlated negatively with work engagement in this thesis. This suggests that they function as hindrance demands, impeding mastery and growth (LePine et al., 2005). The saliency for insufficient work resources observed in this thesis has already been discussed in Section 11.2.1 above. In the same way, the inherently negative nature of abuse, harassment, and bullying in the workplace (Henderson, 2003; Pellico et al., 2009) means this could also apply to workplace aggression. This suggests that hospital doctors' perception of job control, manager support, and effective team practices do not suffice in mitigating the impact of the job demands measured on work-related stress or presenteeism.

11.2.4 The multilevel perspective of the JD-R model

The complexity of the healthcare sector is such that the work experience of individuals, in this case hospital doctors, can be shaped by wider contextual factors, including: the economy, political decisions, funding, resource allocation, and organisational pressures, amongst others (Lowe & Chan, 2010; Montgomery et al., 2011; Powell et al., 2014). These experiences of individual doctors impact their individual performance levels, which could influence the performance of an entire trust when a large proportion of the workforce have similar work experiences. Although multilevel studies have increasingly started to explore the JD-R model at the micro level (i.e., within-individual; Breevaart, Bakker, Demerouti, & Derks, 2016), questions remain as to the relevance of the JD-R model to incorporate organisational demands and outcomes (Bakker & Demerouti, 2017; Schaufeli & Taris, 2014). From a multilevel perspective, this thesis sought to answer two research questions. Firstly, will trust-level demands have the same impact within the model as that of doctors' job demands; and will doctors' psychosocial working conditions and work-related wellbeing predict trust-level quality-of-care outcomes? The second question pertaining to trust-level outcomes has been covered in Section 11.2.2 above, and will not be revisited here.

By using multilevel modelling, the studies in Chapters Eight to Ten avoided violating the compatibility principle, which requires all variables within a model to operate at the same level of specificity (Ajzen, 2005). Instead, multilevel modelling separated the variance into a within-trust and between-trust component (Preacher et al., 2011, 2010). This then allowed measures that operate at the trust-level to be modelled against the between-trust variance components of individual-level job demands and resources, work-related wellbeing, and quality of care. This builds on the limitations of previous studies that broke the compatibility principle by aggregating individual-level job demands and resources to the group-level, before then relating them with team or organisational-level performance (Bakker, van Emmerik, et al., 2008; Harter et al., 2002; Xanthopoulou et al., 2009).

The results were not consistent with the JD-R model, despite the results across the studies in this thesis, which indicated that trust-level demands do have an influence on the work-related wellbeing of hospital doctors. The postulated dual processes were not found, as the number of emergency admissions predicted presenteeism, but not work-related stress. Bed occupancy rates did not predict either negative work-related wellbeing measure. Moreover, both trust-level demands demonstrated a cross-path as they predicted work engagement. In terms of interactions, none of the three individual-level job resources interacted with either trust-level demands. The only mediational role observed was for work engagement, which mediated the relationship between the number of emergency admissions with quality of individual and trust care, and between insufficient work resources and patient satisfaction with doctors.

Trust-level demands. The findings demonstrated a link between trust-level demands and the work-related wellbeing of doctors. While the distinction between challenge and hindrance demands did not originate from the JD-R model (LePine et al., 2005), it is evident where work engagement was positively related with bed occupancy rates and negatively correlated with the number of emergency admissions. Consequently, just as job demands at the individual-level can be construed as something that hinders or enhances work productivity and motivation, measures at the trust-level can also be construed in the same way. Therefore, more thought and concern should be given towards how these trust (or organisational) demands are operationalised and interpreted, including their impact on staff wellbeing. It is also worth questioning whether some of these relationships are even linear in nature. After all, bed occupancy rates has been postulated to have curvilinear properties (Madsen et al., 2014), while at the individual-level doctors' workload (Bertram et al., 1992) and autonomy (Stern et al., 2008) have been observed to have curvilinear relationships with quality-of-care outcomes.

It is important to clarify that this thesis did not look at the shared perception of job demands and resources or work-related wellbeing. Bed occupancy rates and number of emergency admissions are not set in individual perception but exist as structural measures of the work environment. This is not congruent with the definition of psychosocial working conditions used in the thesis, which focuses on an individual's perception of their work environment (see Section 2.2). An argument can be made that these measures should influence doctors' job demands and resources which in turn would influence their work-related wellbeing. Therefore, trust-level demands might be better modelled as antecedents to job demands and resources, rather than to hospital doctors' work-related wellbeing. This links in with the transactional theories of stress (Lazarus, 1999; Lazarus & Folkman, 1984). It pertains to how individuals appraise the environment they are in, whether it may be a challenge to them, and whether they have the capacity to deal with the situation. Appraisal is, therefore construed to be the mediating factor between the structural aspect of work and the subsequent impact on work-related wellbeing. This is also evident in Jex's (1998) review of the stressor–performance relationship, where this relationship was argued to be mediated by cognitions, emotions, and other psychological states. Going back to challenge and hindrance demands, if the situation is seen as one that is challenging but with which the individual can cope with, then this would likely elicit a positive response in terms of mood and motivation (i.e., challenge demand; Crawford et al., 2010; Lazarus & Folkman, 1984). However, when the discrepancy between perceived resources needed and available for the task is too large, then this would be followed by feelings of negative emotions and passive or emotional coping strategies. Therefore, the lack of support for the JD-R model involving trust-level demands may not necessarily be due to limitations with the model, but may be the result of them not being a suitable operationalisation of demands at the organisational-level.

Trust-level outcomes. As multilevel modelling allows the linking of between-trust variance for individual-level work-related wellbeing and job demands and resources, with trust-level outcomes, the thesis was able to respond to the call for more research involving objective outcomes (Bakker & Demerouti, 2014; Schaufeli & Bakker, 2004). It also moots any argument towards common method bias (Podsakoff et al., 2003). The trust-level outcome measures are also of interest to the Department of Health (2008) and NHS England (2014). This is because they function as performance indicators, increase the return-on-investment of data collection, allow the possibility of linking across different data sources, and inform evidence-based policy making (Downs et al., 2017; Medical Research Council, 2014; Nuffield Council on Bioethics, 2015; Welcome Trust, 2014). The results show that although stronger relationships are observed in relation to individual-level quality-of-care outcome measures, work-related wellbeing mediated the relationship that insufficient work resources, job control, and manager support had with patient satisfaction with doctors (see Section 10.5.1 for a discussion on the distal outcomes). With hindsight this difference is not surprising, and serves to highlight some of the issues that researchers need to consider when using organisational-level outcome measures.

11.3 Implications

Fundamentally, this thesis seeks to make a contribution to the scientific literature, to the work experience of hospital doctors, and to the patients receiving treatment. Implications have been discussed at various points in this thesis, including in the discussion above. This section draws these together to make explicit what this means for policy, practice, the NHS Staff Survey, and the use of theory in secondary data.

11.3.1 Policy implications

This section acknowledges the importance of policies towards enhancing psychosocial working conditions of doctors and frames this within the funding deficit in the NHS, the effectiveness of current policies, and the need for synergy between the wellbeing and safety agendas.

Financial deficits in the healthcare sector. The NHS' funding deficit is well-known (Dunn et al., 2016),. This may make it tempting for healthcare policy and decision makers to downgrade the work-related wellbeing and quality-of-care agenda in order to save costs (Leka, Jain, Zwetsloot, Andreou, & Hollis, 2016; Leka, Van Wassenhove, & Jain, 2015). Such cost cutting policies and measures could lead to lower psychosocial working conditions (Walters & Wadsworth, 2014), with adverse impact on doctors' work-related wellbeing and patient care. The challenge, therefore, is to emphasise that short-term savings will likely yield significant costs later on. Chapters Two and Three clearly summarise the literature linking challenging psychosocial working conditions on doctors' work-related wellbeing and quality of care, complementing the findings from this thesis. The issue is compounded by systematic reviews indicating that poor psychosocial working conditions yield significant financial cost to society in the form of increased medical expenditure and reduced productivity (Hassard et al., 2017a; Hassard, Teoh, Visockaite, Dewe, & Cox, 2017b).

Although various stakeholder groups in the healthcare sector have their own political agenda, the findings from this thesis are crucial to inform and lobby them to maintain and protect occupational safety and health (Anyfantis, Boustras, & Karageorgiou, 2016; Leka et al., 2016, 2015). These include among others: employer organisations (e.g., NHS Employers, individual trusts); trade unions (e.g., the BMA, the Royal Colleges); the national labour inspectorate (i.e., the HSE); and policy makers (e.g., the Department of Health). This is particularly relevant to the trade unions and professional groups that exist to protect the interests of their members. Some, such as the Royal College of Physicians (2015a, 2015b), have already raised concern about the challenging working conditions of doctors and their detrimental impact. For them, the findings from this thesis provide evidence specific to hospital doctors in England that support their argument. For others, like the Royal College of Psychiatrists, the work from this thesis is directly informing their strategy on why and how they should address the wellbeing of their members.

Effectiveness of current policies related to psychosocial working conditions. All NHS organisations are required to have a policy on managing work-related stress (NHS Litigation Authority, 2013). Hence, the issue is not about the development of policy but the implementation and understanding of it. Guidance from NHS Employers (2014) and the NHS Litigation Authority (2012, 2013) draws heavily on the work of the HSE (2017), including its definition of stress, the Management Standards, and the stress indicator tool. Despite this, there is little evidence risk assessments and primary interventions are being used although policies that advocate them exist. Examination of the four wellbeing case studies on the NHS Employers (2014) website reveals that the majority of interventions were in the form of workplace health promotion, including: increasing physical activity, health and wellbeing events, access to counselling service, and resiliency training. Only one of these trusts described training line managers to better manage wellbeing issues, using culture change champions, and improving communication channels between senior executives and staff.

This suggests that there is either a lack of awareness of these policies or that these policies are not properly understood. The evidence suggests that are both likely to be contributory factors. NHS managers have been observed to lack knowledge of local and national work-related stress initiatives (Griffith-Noble, 2010; Rodham & Bell, 2002), while those who most engage with training on psychosocial risk management are those already familiar with the issue (Stansfeld et al., 2015). The medical profession has also traditionally focused on treatment and rehabilitation rather than prevention (McGinnis, Williams-Russo, & Knickman, 2002), explaining the emphasis on rolling out support programmes for doctors in need (Roberts, 2016). This also reflects the dominance of secondary and tertiary-level interventions within trusts despite policies advocating risk assessments and prevention. Although it is plausible that the actual range of interventions may differ across trusts, that NHS Employers (2014) highlights trusts focusing on workplace health promotion merely serves to undermine the targeting of psychosocial working conditions. These interventions in practice are certainly valuable, but the discussion needs to shift from management to prevention. Individual-focused interventions typically place the responsibility of improving health and developing resilience on the individual, and away from the organisation and the key decision makers within the system (Leka et al., 2016). Therefore, policy makers need to ensure that policies and guidance issued are relevant and understood by those on the sharp end.

Synergy between wellbeing and patient care. Drives to improve quality of care and patient safety are not different to attempts to improve the work-related wellbeing of hospital doctors. Currently, both these areas function in parallel silos, echoing the wider policy and research fields where wellbeing and safety are distinct (Leka et al., 2016; Zwetsloot, Leka, &

Kines, 2017). Even within work-related wellbeing, a fragmented approach has resulted in separate efforts targeting stress, bullying, obesity, work-life balance, and mental health (Loretto et al., 2005; Sloan et al., 2014; Wells, 2011). The problem is that less importance is typically attached to work-related wellbeing compared to safety issues (Leka et al., 2016). Moreover, solely targeting one issue ignores how other issues are often intertwined. For example, medical residents in the United States report an unwillingness to utilise work-life policies as this increases the workload on their colleagues (Westercamp, Wang, & Fassiotto, 2017). Similarly, poorly designed rotas may comply with the European Working Time Directive, but are detrimental to doctors' rest and personal lives (Clarke, Pitcher, Lambert, & Goldacre, 2014).

From a systems perspective, both safety and wellbeing are predicted by organisational culture (Dixon-Woods et al., 2014; Dollard & Bakker, 2010; Zohar, 2010). The relationships found in this thesis are congruent with the perspective that a holistic approach to managing these issues should be considered (Leka et al., 2016). This could draw on *Vision Zero* policies, which aim to improve the system to develop better workplace safety and wellbeing (Zwetsloot, Kines, et al., 2017; Zwetsloot, Leka, et al., 2017). Similarly, the *Total Worker Health Programme* (NIOSH, 2017) and *WHO Healthy Workplace Model* (WHO, 2010) are about preventing worker illness and injury and enhancing sustainable health and wellbeing through the integration of health promotion with occupational safety and health protection. The evidence base for this is growing, with a systematic review of 17 Total Worker Health interventions demonstrating that all but one improved risk factors for work-related injuries and disease (Anger et al., 2015). While there have yet to be interventions that target both staff wellbeing and the improvement of patient safety or quality of care, the theoretical argument and evidence from other industries indicate that this approach has merit (Anger et al., 2015; Leka et al., 2016; Zwetsloot, Leka, et al., 2017). Consequently, a shift in thinking and policy is required to reflect the overlap between psychosocial working conditions, work-related wellbeing, and quality of care.

11.3.2 Practical implications

The findings from this thesis should elucidate to decision makers at the national and trust-level that the demands placed upon hospitals, and the resources they have to deal with them, have a real impact on the wellbeing of hospital staff and patient care. Foremost, this thesis provided a theoretical framework, namely the JD-R model (Bakker & Demerouti, 2017; Demerouti et al., 2001), in which to understand the relationships involving hospital doctors' psychosocial working conditions. Recognising that "there is nothing as practical as a good theory" (Lewin, 1943), the JD-R model can be used to better target change initiatives. Theory therefore is crucial to inform the pathways and work aspects that warrant change.

More than burnout. The findings from this thesis suggest a need to a move towards examining how work and organisational factors can be better developed to yield more positive responses (e.g., motivation and work engagement). This is to compliment the dominance of doctor burnout within the academic and practitioner literature evident in Chapters Two and Three. In addition to the moral argument, the evidence suggests that engaged and happy doctors are more likely to deliver better patient outcomes and experience (Scheepers et al., 2015). This is clear from this thesis, where work engagement was the most consistent predictor of both self-rated and trust-level quality of care. That job demands primarily predicted negative work-related wellbeing while job resources predicted work engagement indicates a differential effect between predictors and work-related wellbeing. These findings reinforce the argument that interventions should not only attempt to reduce job demands but to strengthen job resources in the workplace (Knight et al., 2017). The target of change should differ depending on the work-related wellbeing measure being targeted.

More than individual interventions. Interventions need to move beyond those that target change within the doctor (i.e., individual interventions) and consider instead approaches that target multiple job demands and resources in the work environment (i.e., trust-level interventions). However, to date few studies (e.g., Benning et al., 2011) in the healthcare sector have examined workplace-based psychosocial interventions. Interventions are still dominated by cognitive, behavioural, and mindfulness based interventions, with meta-analyses demonstrating these to reduce doctors and medical students' level of anxiety and stress (Regehr, Glancy, Pitts, & LeBlanc, 2014) as well as burnout (McCray, Cronholm, Bogner, Gallo, & Neill, 2008). When individual and organisational interventions are compared, one meta-analysis on doctor burnout found structural and organisational interventions were more effective in relation to overall burnout than individual interventions (C. P. West, Dyrbye, Erwin, & Shanafelt, 2016).

Trust-level interventions involving doctors are more likely to be impactful when they embrace good practice principles of risk assessments and a participatory approach (Cox, Taris, & Nielsen, 2010; Leka, Cox, & Zwetsloot, 2008; K. Nielsen & Miraglia, 2017). For example, a structured, participatory intervention based on continuous group meetings among German hospital doctors led to changes in work practices, increased training activities, and enhanced technical support (Weigl et al., 2013). This corresponded with doctors in the intervention departments reporting better support and less conflicting demands post-intervention. Patients also reported better satisfaction with the organisation of care. However, interventions do not have to be either at the individual or trust-level (C. P. West et al., 2016), as evidence beyond the healthcare sector indicate that a comprehensive approach involving both forms of interventions should yield the best improvements (Van der Klink, Blonk, Schene, & Van Dijk, 2001).

Most of the literature surrounding psychosocial interventions has been in relation to healthcare workers' work-related wellbeing and not in relation to quality of care or patient safety. Despite this, the extant evidence demonstrates that interventions to improve safety climate are associated with better safety behaviours (C. Snijders, Kollen, van Lingen, Fetter, & Molendijk, 2009), adverse outcomes (Pettker et al., 2009), and post-operative outcomes (Haynes et al., 2011). Climate reflects the *soft* aspects of the work environment, and has been found to be a precursor to job demands and resources (Dollard & Bakker, 2010). As discussed in the Policy section above, there has yet to interventions targeting both wellbeing and patient care. However, an evaluation of the Health Foundation's Safer Patient Initiatives across four UK hospitals found an improvement on some clinical processes, and a decline in the number of staff experiencing work-related stress (Benning et al., 2011). Therefore, more practical interventions that link work-related wellbeing and patient care are still needed. However, it is important to recognise that no "one size fits all" solution exists.

Beyond trusts. Hospital doctors' working conditions and pressures are not solely determined by their trusts. As junior doctors undergo regular training and examination, the requirements set out by the relevant Royal Colleges and their Local Education Training Board can influence their psychosocial work experience. In light of concern surrounding working conditions during the junior doctor contract negotiations in 2016, Health Education England (2017) formed a working group to address ten non-contract issues relating to training, the work environment, and doctors feeling valued. This included representatives from NHS Employers, the BMA Junior Doctors' Committee, Academy of Medical Royal Colleges, the General Medical Council, and junior doctors themselves.

This has already resulted in changes that focus on reducing demands and increasing the amount of control that junior doctors have. For example, notice of placements is increased from eight to 12 weeks, while rotas are to be published eight weeks rather than six weeks ahead.

Recruitment and induction process are to be standardised so that duplication is minimised while on rotation. Similarly, junior doctors who have caring responsibilities or health conditions that tie them to a specific location will receive priority allocation during recruitment phase. The working group's ongoing work will see them pilot and test other changes designed to improve the work and training experiences of junior doctors. The success of these changes has yet to be tested. However, by focusing on prevention and changes to the work and training environment, involving relevant appropriate stakeholder groups, encouraging junior doctor input, and being willing to pilot and test ideas, the work here is firmly embedded in the principles advocated within the occupational health psychology literature for successful interventions (Cox et al., 2010; K. Nielsen & Miraglia, 2017; Randall & Nielsen, 2010).

11.3.3 Implications for the NHS Staff Survey

The thesis demonstrates the suitability of the NHS Staff Survey as a data source on the working conditions of hospital doctors. The strengths and limitations of the NHS Staff Survey are considered in Chapter Six where it was introduced, as well as in the respective studies that used it for analyses. Its greatest advantage is its large and random sample, clustered into each English trust, that reflects the NHS staff population. In addition, although this thesis was restricted to cross-sectional analyses using the 2014 dataset, the annual surveying of staff means the dataset provides added utility for longitudinal analyses at the trust-level.

Advances in statistical analyses and software have increased the interest in techniques that overcome the difficulty in managing clustered data, while allowing the integration of individual and organisational-level data (Preacher et al., 2011; Zhang et al., 2009). The clustering of healthcare staff according to their occupational groups and trusts allows for multilevel analyses using the NHS Staff Survey. This is especially as the number of acute trusts surpasses the various minimum recommended group-level sample size requirements (Hox & Maas, 2001; Hox et al., 2010; Maas & Hox, 2005; Meuleman & Billiet, 2009). Examination of data at the trust-level also permits the linking with numerous other forms of data being collected about trusts. In Chapter Ten, the NHS Staff Survey was linked to hospital mortality, patient satisfaction with their doctor, and patient safety incidents. However, alternative measures include amongst others: infection rates, financial performance, Care Quality Commission ratings, and patient complaints (Shipton, Armstrong, West, & Dawson, 2008; Topakas, Admasachew, & Dawson, 2010b; M. West & Dawson, 2010).

The caveat here lies in the complexity of the multilevel model being developed. Chapter Ten acknowledged that the number of trusts provided from the NHS Staff Survey might generate too little statistical power when complex models are proposed (Section 10.2.3). Although the number of trusts could be increased by including mental health, ambulatory, and community trusts –this is contingent on the research question posed. This remains an issue within the wider sphere of multilevel modelling, particularly when all available groups are already sampled (e.g., clustering according to nations; Meuleman & Billiet, 2009). This also links with the NHS Staff Survey's inability to cluster below the trust-level to protect the anonymity of responders. Clustering at the ward or departmental-level may be useful in generating a larger group-level sample. It would also allow for the examination of measures that are more proximal to the individual being surveyed. While it is for the NHS Staff Survey Advisory Group (Picker Institute Europe, 2015) to determine the relevance of including items that allow clustering at the lower levels, future researchers need to consider what impact grouping at the trust-level has in terms of staff experience and the outcome measures included.

11.3.4 Applying theory in secondary data analyses

The thesis centres on the application of a theoretical framework (i.e., the JD-R model) onto an existing secondary dataset. However, the NHS Staff Survey could also allow other theoretical models to be tested, including the job demand-control-support model (Johnson & Hall, 1988) and self-determination theory (Deci & Ryan, 1985). Applying theory to secondary data analyses is typically a post hoc exercise where researchers have little control of the measures available within the dataset used. This, in turn, introduces a degree of measurement error as proxy variables are used to represent constructs of interest. For example, job demands in this thesis were represented by insufficient work resources and workplace aggression. This does not fully capture what job demands represent, as commonly used indicators such as workload and work pressure were not available (Bakker, 2014). Moreover, demands salient to the healthcare sector that were identified in Chapter Four (e.g., emotional conflict, work-life balance), were missing. Therefore, researchers need to be aware that poor operationalisation of constructs could undermine the validity of the theoretical model being tested with a particular dataset.

The examination of simple relationships between two constructs may be intuitive and could function within an a-theoretical framework. However, the healthcare sector, and organisations more generally, are far more complicated. In trying to analyse and link large datasets, theory provides a framework to identify relevant constructs and structure hypotheses testing. Otherwise, researchers could run multiple tests across different measures to obtain significant results that may have little practical significance (Schlomer & Copp, 2014). Although the JD-R model provides a useful framework at the individual-level, trust-level demands did not operate as anticipated. This reflects the general trend of stress theories which have yet to be rigorously examined from a multilevel perspective, and may not be able to accurately reflect the complex organisational systems that hospital doctors function in. Instead, rather than incorporating a systems perspective into the stress theories, researchers may want to consider whether theories from organisational sciences are better positioned to explain the context within the healthcare sector (Dow, DiazGranados, Mazmanian, & Retchin, 2013; Nilsen, 2015).

11.4 Limitations

The specific limitations for each study are discussed within the respective chapters. However, three of them warrant repeating here: heterogeneous doctor samples, the absence of causality, and low internal reliability for some of the measures. It is important to recognise that the thesis sample encompassed medical professionals from acute and specialist English trusts. This limits its generalisability to other groups and settings. No distinction is made to doctors' specialty (e.g., emergency medicine, paediatrics, dentistry) or seniority (e.g., foundation doctor, consultant), who experience different working conditions and different care challenges. The thesis is also limited as its study design cannot establish causality. Where the term predictor has been used, this is a reflection of the role of the measure within a regression model (Field, 2014) and not as a causal antecedent. It is likely that some form of reciprocal relationships exist between these constructs, as longitudinal studies have found doctors who are psychologically unwell to perceive poorer working conditions (Bakker et al., 2000; McManus et al., 2004), and that doctors' error rates predicted future levels of depressive symptoms (Shanafelt et al., 2010). Equally important to highlight is the low internal reliability scores observed for insufficient work resources, workplace aggression, and the number of errors seen. The lower the internal reliability the more measurement error is introduced (Hair et al., 2014), this biases the relationships involving the true score that the measure is supposed to reflect (Lance et al., 2006). Although future researchers may want to want use measures with more items or that have been demonstrated to have higher reliability (Kline, 2016), one limitation of using secondary data is that there is no control over the original measures chosen and used. Consequently, any interpretations of these results and implications need to be framed with these limitations in mind.

Additional limitations that relate to the thesis as a whole also need to be acknowledged. First, although the thesis aimed to test the JD-R model within this sample of hospital doctors, it was not able to test the whole model as one large structural model. Instead, different components of the model were tested across the three studies in Chapters Eight to Ten. Even here, the inclusion of trust-level outcomes in Chapter Ten necessitated the use of eight different models. This was largely due to the limited number of trusts at the group-level which impacted statistical power and model complexity. One alternate approach would be to test the entire model at the individual-level with both moderation and mediation as a single structural equation model. This, however, would not only have left out the clustering doctors but also the trust-level predictors and outcomes.

Second, as the JD-R model provided the framework for this thesis, one of the key limitations of the model applies to this thesis as well. This is that the JD-R provides an open, heuristic, and descriptive model that can be somewhat superficial (Schaufeli & Taris, 2014). It does not provide a psychological explanation on why the predicted relationships occur, relying instead on other psychological frameworks including conservation of resources (Hobfoll, 1989; Hobfoll & Shirom, 2001), self-determination theory (Deci & Ryan, 1985), social cognitive theory (Bandura, 2010), and job characteristic theory (Hackman & Oldham, 1980). Consequently, although this thesis to an extent can explain what happens, it is not able to explain why this is the case.

The third limitation pertains to the low intraclass correlations (ICC) and design effect (*deff)* values observed in this thesis. To recap, ICC represents the amount of variance due to between-group variation (Heck & Thomas, 2015); *deff* refers to the ICC in relation to average cluster size (B. O. Muthén & Satorra, 1995). An argument could have been made that multilevel modelling was not even needed in this thesis as the ICC values were always negligible, while not all *deff* values surpassed the required threshold (Hox, 2010). This means that that the assumption of independence could perhaps be met (Heck & Thomas, 2015). Consequently, a single-level structural equation model would have allowed for the more complex models mentioned in the previous limitation. However, the decision to carry on with multilevel analyses was based on the recommendation that as long as the data presents in hierarchical form then multilevel modelling should be used (Nezlek, 2008; Stride, 2016). This is as variation can still occur in the relationship between the outcome measure and other measures in the model. In addition, the absence of a multilevel perspective would not have allowed the inclusion of trust-level demands and outcomes.

The fourth limitation relates to the distinction between psychosocial and structural elements of workload. Although Parkes and Sparkes' (1998) classed psychosocial working conditions as the perception of the work environment (see Section 2.2), psychosocial working conditions typically refers to how work environments are managed, organised, and designed (Cox et al., 2000; Leka et al., 2008). Not only do structural aspects of workload fall under this definition, but numerous psychosocial taxonomies include structural demands such as shift-work, long work hours, night or unsociable shifts, and low wages (Cox et al., 2000; Dewe & Trenberth, 2004; HSE, 2017). This means that in practice the distinction between psychosocial and structural aspects of workload (and even the work environment) is not as simple as presented in this thesis. While some studies have observed perceived psychosocial aspects to predict more variance on outcomes than structural demands in the work environment (P. Tucker et al., 2015; Visser et al., 2003), this was not done in this thesis. Including and controlling for structural workload in the models would have allowed a better understanding of the influence that psychosocial working conditions have on hospital doctors' work-related wellbeing and the quality of care they provide, above and beyond the influence of these structural elements. Equally, psychosocial working conditions could be modelled as a mediator with structural workload as a predictor, testing the role of an individual's appraisal in explaining the pathways between work environment and outcomes (Jex, 1998; Lazarus, 1999). However, the secondary dataset which underpins this thesis did not provide data relating to structural workload. This remains as a limitation that should be modelled in future research investigating the impact of psychosocial working conditions.

11.5 Future Research Directions

Researchers looking to build on this thesis should consider the limitations acknowledged in the preceding section and in the respective studies. In particular, future researchers are encouraged to employ longitudinal designs to better understand the temporal relationships; use more precise and focused samples that account for the heterogeneity of hospital doctors; and explore the psychological mechanisms that explain the JD-R model's postulated pathways.

In addition, researchers should pay attention to the impact of using distal and proximal measures. As indicated throughout this discussion, distal measures likely explain why relationships involving some outcome measures were supported while others were not. This does not mean that distal outcomes (e.g., hospital mortality) are redundant, but more research is needed to explore how to theoretically and empirically close the gap between distal measures. One approach here is to reduce the group-level analyses from the trust-level, down to a lower-level, such as the ward or department. This should result in an individual's psychosocial working environment having a stronger impact on group-level outcome measures.

The job demands and resources used in this thesis were restricted by the NHS Staff Survey. Future studies may want to explore the other job demands and resources which were found in Chapter Four's systematic review, including time pressure, perceived workload, learning and development opportunities, and emotional demands. It is important that the local context be acknowledged, and here risk assessments can play an important role in identifying the salient psychosocial working conditions that warrant attention (Cox et al., 2010; Leka et al., 2008; K. Nielsen & Miraglia, 2017). This may in turn lead to relationships that more closely map onto those postulated by the JD-R model.

In terms of the JD-R model, there needs to be consideration as to why and how job resources would buffer job demands against negative wellbeing. In the same way, how and why do job resources and demands interact to enhance work engagement? Context is important, and it may be that other job resources that have been found to buffer job demands (development opportunities, feedback, supervisory coaching, and participation; Bakker et al., 2011; van Vegchel et al., 1999), could be considered. A second option is to examine the aggregated effect of more than one job resource. If both job demands here arguably are salient and strong, then job resources could only play a role when hospital doctors are able to draw on more than one form of resource. To date a latent factor of job resources has not yet been examined from a moderation perspective, although multiple studies have tested a latent factor of job resources (and demands) in terms of mediation (Bakker, Demerouti, de Boer, et al., 2003; Xanthopoulou, Bakker, Demerouti, et al., 2007) and direct effects (Hakanen et al., 2008; Schaufeli, 2015). Bakker and Demerouti (2017) identified the need to understand the effect of job demands accumulating.

Arguably, this proposal can also be framed around the effect of the accumulation of job resources, especially in the face of salient manifestations of job demands.

Finally, more contemporary developments within the JD-R model, namely the role of personal resources and job crafting (Bakker & Demerouti, 2014, 2017), warrant further attention. Personal resources (e.g., self-efficacy, optimism) function in a similar manner to job resources. It encompasses the individual's ability to successfully control their environments (Hobfoll, Johnson, Ennis, & Jackson, 2003), and allows workers to reduce the discrepancy between their expectations and goals (Luthans, Avey, Avolio, Norman, & Combs, 2006). Their role in interacting with job demands while buffering job demands means they can have a crucial role in understanding the validity of the JD-R among hospital doctors. The second development – job crafting, refers to the individual's ability to change their work environment by manipulating the task, the relationships around them, or how they appraise their work (Wrzesniewski & Dutton, 2001). Motivated employees are more likely to job craft, which leads to increased job resources, that in turn increases motivation and engagement; thus, this creates a continuous gain cycle. This is seen in a job crafting intervention which resulted in healthcare workers seeking more resources and challenging demands, while reducing their hindering demands (Gordon et al., 2017). This improved work-related wellbeing and subjective performance, but not objective performance. Therefore, job crafting could be particularly relevant in the healthcare sector, where hospital doctors have limited control over the demands and policies that affect their work environment.

11.5 Final Summary

This thesis brought together hospital doctors' psychosocial working conditions, their work-related wellbeing, and the quality of patient care, into one single theoretical framework - the job demands-resources model (Bakker & Demerouti, 2017; Demerouti et al., 2001). Using the 2014 NHS Staff Survey, it adds value to the extant literature by linking the JD-R model with existing quality indicators from the Care Quality Commission and various NHS agencies. Most of the JD-R model's propositions were supported. Job demands primarily predicted work-related stress and presenteeism, while work engagement was mainly predicted by job resources. Similarly, work-related wellbeing mediated most relationships between psychosocial working conditions and quality of care. It is clear that there is merit in expanding the research and practice involving hospital doctors beyond the commonly used construct of burnout. In particular, support for the JD-R model's motivational pathway supports a pathway that links job resources with work engagement and subsequently quality of care.

Crucially, by employing a multilevel perspective it was possible to situate these relationships within a systems perspective. Much of the research, in the wider occupational psychology and medical literature, has focused at the individual-level without acknowledging that the work-related wellbeing and performance of hospital doctors is a product of the context that they are in. Not only did the thesis demonstrate that some trust-level predictors influence hospital doctors' work-related wellbeing and quality of care, but that doctors' perception of their work environment related with some trust-level outcomes. This should educate healthcare policy and decision makers that the demands placed upon hospitals, and the resources they have to deal with them, have an impact on the wellbeing of hospital doctors and the quality of patient care.

The findings suggest that efforts to improve both hospital doctors' work-related wellbeing and the quality of care should focus on interventions that aim to reduce job demands while increasing job resources. The JD-R model provides a useful framework in which to understand the impact of psychosocial working conditions and to target interventions. Finally, based on the relationships between hospital doctors' psychosocial working environment with both work-related wellbeing and quality of care, a greater synergy between the wellbeing and safety agendas is recommended.

References

Academy of Medical Royal Colleges. (2012). *The benefits of consultant delivered care*. London, UK.

Admasachew, L., & Dawson, J. F. (2010). *Staff engagement in the NHS: A multilevel analysis*. Birmingham, UK: Aston University.

Admasachew, L., & Dawson, J. F. (2011). The association between presenteeism and engagement of National Health Service staff. *Journal of Health Services Research & Policy*, *16*(1), 29–33. https://doi.org/10.1258/jhsrp.2010.010085

Adriaenssens, J., De Gucht, V., & Maes, S. (2015). Determinants and prevalence of burnout in emergency nurses: A systematic review of 25 years of research. *International Journal of Nursing Studies*, *52*(2), 649–661. https://doi.org/10.1016/j.ijnurstu.2014.11.004

Ahmad, A. (2010). Work-family conflict among junior physicians: Its mediating role in the relationship between role overload and emotional exhaustion. *Journal of Social Sciences*, *6*(2), 265–271. https://doi.org/10.3844/jssp.2010.265.271

Aiken, L. H., Clarke, S. P., Sloane, D. M., Lake, E. T., & Cheney, T. (2008). Effects of hospital care environment on patient mortality and nurse outcomes. *JONA: The Journal of Nursing Administration*, *38*(5), 223–229. https://doi.org/10.1097/01.NNA.0000312773.42352.d7

Aiken, L. H., Sermeus, W., Van den Heede, K., Sloane, D. M., Busse, R., McKee, M., … Kutney-Lee, A. (2012). Patient safety, satisfaction, and quality of hospital care: Cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. *BMJ (Clinical Research Ed.)*, *344*(March), e1717. https://doi.org/10.1136/bmj.e1717

Ajzen, I. (2005). Laws of human behavior: Symmetry, compatibility, and attitude-behavior correspondence. In A. Beauducel, B. Biehl, M. Bosniak, W. Conrad, G. Schönberger, & D. Wagener (Eds.), *Multivariate research strategies* (pp. 3–19). Aachen, Germany: Shaker Verlag.

Alderson, P. (1998). The importance of theories in health care. *BMJ*, *317*(7164), 1007–1010. https://doi.org//10.1136/bmj.317.7164.1007

Alexander, C., & Fraser, J. (2004). Occupational violence in an Australian healthcare setting: Implications for managers. *Journal of Healthcare Management*, *49*(6), 377–390.

An, P. G., Manwell, L. B., Williams, E. S., Laiteerapong, N., Brown, R. L., Rabatin, J. S., … Linzer, M. (2013). Does a higher frequency of difficult patient encounters lead to lower quality care? *The Journal of Family Practice*, *62*(1), 24–29.

An, P. G., Rabatin, J. S., Manwell, L. B., Linzer, M., Brown, R. L., & Schwartz, M. D. (2009). Burden of difficult encounters in primary care: Data from the minimizing error, maximizing outcomes study. *Archives of Internal Medicine*, *169*(4), 410. https://doi.org/10.1001/archinternmed.2008.549

Anagnostopoulos, F., Liolios, E., Persefonis, G., Slater, J., Kafetsios, K., & Niakas, D. (2012). Physician burnout and patient satisfaction with consultation in primary health care settings: Evidence of relationships from a one-with-many design. *Journal of Clinical Psychology in Medical Settings*, *19*(4), 401–410. https://doi.org/10.1007/s10880-011-9278-8

Anger, W. K., Elliot, D. L., Bodner, T., Olson, R., Rohlman, D. S., Truxillo, D. M., … Montgomery, D. (2015). Effectiveness of Total Worker Health interventions. *Journal of Occupational Health Psychology*, *20*(2), 226–247. https://doi.org/10.1037/a0038340

Ansmann, L., Kowalski, C., Ernstmann, N., Ommen, O., Jung, J., Visser, A., & Pfaff, H. (2013). Do breast cancer patients receive less support from physicians in German hospitals with high physician workload? A multilevel analysis. *Patient Education and Counseling*, *93*(2), 327–334. https://doi.org/10.1016/j.pec.2013.06.016

Ansmann, L., Wirtz, M., Kowalski, C., Pfaff, H., Visser, A., & Ernstmann, N. (2014). The impact of the hospital work environment on social support from physicians in breast cancer care. *Patient Education and Counseling*, *96*(3), 352–360. https://doi.org/10.1016/j.pec.2014.07.016

Anyfantis, I., Boustras, G., & Karageorgiou, A.

(2016). Maintaining occupational safety and health levels during the financial crisis - A conceptual model. *Safety Science*. https://doi.org/10.1016/j.ssci.2016.02.014

Appleby, J., & Robertson, R. (2016). *Public satisfaction with the NHS in 2015: Results and trends form the British Social Attitudes Survey*. London, UK: The King's Fund. Retrieved from http://www.kingsfund.org.uk/file/15497

Arnetz, B. B. (2001). Psychosocial challenges facing physicians of today. *Social Science and Medicine*, *52*(2), 203–213. https://doi.org/10.1016/S0277-9536(00)00220-3

Arora, S., Sevdalis, N., Aggarwal, R., Sirimanna, P., Darzi, A., & Kneebone, R. (2010). Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy and Other Interventional Techniques*, *24*(10), 2588–2593. https://doi.org/10.1007/s00464-010-1013-2

Arrindell, W. A., & van der Ende, J. (1985). An empirical test to the utility of the observer-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, *9*, 165–178. https://doi.org/10.1177/014662168500900020

Asparouhov, T., & Muthén, B. O. (2006). *Constructing covariates in multilevel regression*. *Mplus Web Notes: No. 11*.

Asparouhov, T., & Muthén, B. O. (2010a). *Bayesian analysis of latent variables models using Mplus* (Mplus Web Notes). Retrieved from www.statmodel.com/download/BayesAdvantages18.pdf

Asparouhov, T., & Muthén, B. O. (2010b). *Multiple imputation with Mplus* (MPlus Web Notes). Retrieved from http://statmodel2.com/download/Imputations7.pdf

Asparouhov, T., & Muthén, B. O. (2010c). *Weighted least squares estimation with missing data* (Mplus Technical Appendix). Retrieved from http://www.statmodel.com/download/GstrucMissingRevision.pdf

Asparouhov, T., & Muthén, B. O. (2013). *Computing the strictly positive Satorra-Bentler chi-square test in Mplus* (Mplus Web Notes No. 12). Retrieved from http://www.statmodel.com/examples/webnotes/SB5.pdf

Aspden, P., Corrigan, J., Wolcott, J., & Erickson, S. M. (2004). *Patient safety: Achieving a new standard for care*. Washington, DC: The National Academies Press.

Bae, S.-H. (2011). Assessing the relationships between nurse working conditions and patient outcomes: systematic literature review. *Journal of Nursing Management*, *19*(6), 700–713. https://doi.org/10.1111/j.1365-2834.2011.01291.x

Bagenal, J., Moberly, T., & Godlee, F. (2015). Problems with the new junior doctor contract. *BMJ*, *5077*(September 2015), h5077. https://doi.org/10.1136/bmj.h5077

Bagnall, P. (2012). *Facilitators and barriers to leadership and quality improvement: The King's Fund Junior Doctor Project*. London, UK: The King's Fund.

Baker, G. R., Norton, P. G., Flintoft, V., Blais, R., Brown, A., Cox, J., … Tamblyn, R. (2004). The Canadian Adverse Events Study: the incidence of adverse Events among hospital patients in Canada. *CMAJ: Canadian Medical Association Journal*, *170*(11), 1678–1686. https://doi.org/10.1503/cmaj.1040498

Bakker, A. B. (2014). *The Job Demands–Resources Questionnaire*. Rotterdam, the Netherlands.

Bakker, A. B., & Bal, P. M. (2010). Weekly work engagement and performance: A study among starting teachers. *Journal of Occupational and Organizational Psychology*, *83*, 189–206. https://doi.org/10.1348/096317909X402596

Bakker, A. B., Boyd, C. M., Dollard, M. F., Gillespie, N., Winefield, A. H., & Stough, C. (2010). The role of personality in the job demands-resources model: A study of Australian academic staff. *Career Development International*, *15*(7), 622–636. https://doi.org/10.1108/13620431011094050

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology*, *22*(3), 309–328. https://doi.org/10.1108/02683940710733115

Bakker, A. B., & Demerouti, E. (2014). Job Demands-Resources Theory. In P. Chen & C. L. Cooper (Eds.), *Work and wellbeing: Wellbeing: A complete reference guide* (Vol. III, pp. 1–28). John Wiley & Sons.

https://doi.org/10.1002/9781118539415.wbwell019

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273–285. https://doi.org/10.1037/ocp0000056

Bakker, A. B., Demerouti, E., de Boer, E., & Schaufeli, W. B. (2003). Job demand and job resources as predictors of absence duration and frequency. *Journal of Vocational Behavior*, *62*(2), 341–356. https://doi.org/10.1016/S0001-8791(02)00030-1

Bakker, A. B., Demerouti, E., & Euwema, M. C. (2005). Job resources buffer the impact of job demands on burnout. *Journal of Occupational Health Psychology*, *10*(2), 170–180. https://doi.org/10.1037/1076-8998.10.2.170

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2014). Burnout and work engagement: The JD–R approach. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 389–411. https://doi.org/10.1146/annurev-orgpsych-031413-091235

Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2003). Dual processes at work in a call centre: An application of the job demands – resources model. *European Journal of Work and Organizational Psychology*, *12*(4), 393–417. https://doi.org/10.1080/13594320344000165

Bakker, A. B., Demerouti, E., Taris, T. W., Schaufeli, W. B., & Schreurs, P. J. G. (2003). A multigroup analysis of the job demands-resources model in four home care organizations. *International Journal of Stress Management*, *10*(1), 16–38. https://doi.org/10.1037/1072-5245.10.1.16

Bakker, A. B., Demerouti, E., & Verbeke, W. (2004). Using the job demands-resources model to predict burnout and performance. *Human Resource Management*, *43*(1), 83–104. https://doi.org/10.1002/hrm.20004

Bakker, A. B., Hakanen, J. J., Demerouti, E., & Xanthopoulou, D. (2007). Job resources boost work engagement, particularly when job demands are high. *Journal of Educational Psychology*, *99*(2), 274–284. https://doi.org/10.1037/0022-0663.99.2.274

Bakker, A. B., & Sanz-Vergel, A. I. (2013). Weekly work engagement and flourishing: The role of hindrance and challenge job demands. *Journal of Vocational Behavior*, *83*(3), 397–409. https://doi.org/10.1016/j.jvb.2013.06.008

Bakker, A. B., & Schaufeli, W. B. (2008). Positive organizational behavior: Engaged employees in flourishing organizations. *Journal of Organizational Behavior*, *29*(2), 147–154. https://doi.org/10.1002/job.515

Bakker, A. B., Schaufeli, W. B., Leiter, M. P., & Taris, T. W. (2008). Work engagement: An emerging concept in occupational health psychology. *Work & Stress*, *22*(3), 187–200. https://doi.org/10.1080/02678370802393649

Bakker, A. B., Schaufeli, W. B., Sixma, H. J., Bosveld, W., & Van Dierendonck, D. (2000). Patient demands, lack of reciprocity, and burnout: A five-year longitudinal study among general practitioners. *Journal of Organizational Behavior*, *21*(4), 425–441. https://doi.org/10.1002/(SICI)1099-1379(200006)21:4<425::AID-JOB21>3.0.CO;2-#

Bakker, A. B., ten Brummelhuis, L. L., Prins, J. T., & van der Heijden, F. M. M. A. (2011). Applying the job demands–resources model to the work–home interface: A study among medical residents and their partners. *Journal of Vocational Behavior*, *79*(1), 170–180. https://doi.org/10.1016/j.jvb.2010.12.004

Bakker, A. B., van Emmerik, H. I. J., & Euwema, M. C. (2006). Crossover of burnout and engagement in work teams. *Work and Occupations*, *33*(4), 464–489. https://doi.org/10.1177/0730888406291310

Bakker, A. B., van Emmerik, H. I. J., & Van Riet, P. (2008). How job demands, resources, and burnout predict objective performance: A constructive replication. *Anxiety, Stress & Coping*, *21*(3), 309–324. https://doi.org/10.1080/10615800801958637

Bakker, A. B., van Veldhoven, M., & Xanthopoulou, D. (2010). Beyond the demand-control model: Thriving on high job demands and resources. *Journal of Personnel Psychology*, *9*(1), 3–16. https://doi.org/10.1027/1866-5888/a000006

Baldisseri, M. R. (2007). Impaired healthcare professional. *Critical Care Medicine*, *35*(2 Suppl), S106-116. https://doi.org/10.1097/01.CCM.0000252918.87746.96

Baldwin, P. J., Dodd, M., & Wrate, R. W. (1997). Young doctors' health--I. How do working conditions affect attitudes, health and performance? *Social Science and Medicine*, *45*(1), 35–40. https://doi.org/10.1016/S0277-9536(96)00306-1

Bandura, A. (2010). Self-Efficacy. In *The Corsini Encyclopedia of Psychology* (Vol. 4, pp. 71–81). Hoboken, NJ: John Wiley & Sons, Inc.

Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational and Behavioral Statistics*, *6*(3), 267–285. https://doi.org/10.3102/10769986006003267

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*(2), 142–163. https://doi.org/10.1037/1082-989X.11.2.142

Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, *8*(3), 274–289. https://doi.org/10.1177/1094428105278021

Benning, A., Ghaleb, M., Suokas, A., Dixon-Woods, M., Dawson, J. F., Barber, N., … Lilford, R. (2011). Large scale organisational intervention to improve patient safety in four UK hospitals: mixed Method evaluation. *BMJ*, *342*(feb03 1), d195–d195. https://doi.org/10.1136/bmj.d195

Benson, S., Truskett, P. G., & Findlay, B. (2007). The relationship between burnout and emotional intelligence in Australian surgeons and surgical trainees. *ANZ Journal of Surgery*, *77*, 79. https://doi.org/10.1111/j.1445-2197.2007.04129_12.x

Bernburg, M., Vitzthum, K., Groneberg, D. A., & Mache, S. (2016). Physicians' occupational stress, depressive symptoms and work ability in relation to their working environment: A cross-sectional study of differences among medical residents with various specialties working in German hospitals. *BMJ Open*, *6*(6), e011369. https://doi.org/10.1136/BMJOPEN-2016-011369

Bertram, D. A., Hershey, C. O., Opila, D. A., & Quirin, O. (1990). A measure of physician mental work load in internal medicine ambulatory care clinics. *Medical Care*, *28*(5), 458–467. https://doi.org/10.1097/00005650-199005000-00005

Bertram, D. A., Opila, D. A., Brown, J. L., Gallagher, S. J., Schifeling, R. W., Snow, I. S., & Hershey, C. O. (1992). Measuring physician mental workload: Reliability and validity assessment of a brief instrument. *Medical Care*, *30*(2), 95–104. https://doi.org/10.1097/00005650-199202000-00001

Blau, P. M. (1964). *Exchange and Power in Social Life*. *Wiley*. New York, NY: Wiley.

Bliese, P. D., & Britt, T. W. (2001). Social support, group consensus and stressor-strain relationships: social context matters. *Journal of Organizational Behavior*, *22*(4), 425–436. https://doi.org/10.1002/job.95

Bloor, K., Freemantle, N., & Maynard, A. (2012). Trends in consultant clinical activity and the effect of the 2003 contract change: Retrospective analysis of secondary data. *Journal of the Royal Society of Medicine*, *105*(11), 472–479. https://doi.org/10.1258/jrsm.2012.120130

Boden, D. G., Agarwal, A., Hussain, T., Martin, S. J., Radford, N., Riyat, M. S., … Whale, C. I. (2016). Lowering levels of bed occupancy is associated with decreased inhospital mortality and improved performance on the 4-hour target in a UK district general hospital. *Emergency Medicine Journal*, *33*(2), 85–90. https://doi.org/10.1136/emermed-2014-204479

Boorman, S. (2009). *NHS Health and Well-being Review: Interim Report*. London, UK: Department of Health. Retrieved from http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_108910.pdf

Borenstein, M., Hedges, L. V., & Higgins, J. P. T. (2009). *Introduction to meta-analysis*.

Chichester, UK: Wiley.

Borg, M. . (2003). Bed occupancy and overcrowding as determinant factors in the incidence of MRSA infections within general ward settings. *Journal of Hospital Infection*, *54*(4), 316–318. https://doi.org/10.1016/S0195-6701(03)00153-1

Bottle, A., Jarman, B., & Aylin, P. P. (2011). Strengths and weaknesses of hospital standardised mortality ratios. *BMJ (Clinical Research Education)*, *342*, c7116. https://doi.org/10.1136/bmj.c7116

Bowman, N. A. (2012). Effect sizes and statistical methods for meta-analysis in higher education. *Research in Higher Education*, *53*(3), 375–382. https://doi.org/10.1007/s11162-011-9232-5

Breevaart, K., Bakker, A. B., Demerouti, E., & Derks, D. (2016). Who takes the lead? A multi-source diary study on leadership, work engagement, and job performance. *Journal of Organizational Behavior*, *37*(3), 309–325. https://doi.org/10.1002/job.2041

Briner, R. B. (2014). *What is employee engagement and does it matter? An evidence-based approach*. Bath, UK: University of Bath School of Management.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guildford Press.

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology*, *37*, 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Buchan, J., Seccombe, I., & Charlesworth, A. (2016). *Staffing matters; funding counts. Workforce profile and trends in the English NHS*. London, UK: The Health Foundation. Retrieved from http://www.health.org.uk/publication/staffing-matters-funding-counts

Buddeberg-Fischer, B., Klaghofer, R., Stamm, M., Siegrist, J., & Buddeberg, C. (2008). Work stress and reduced health in young physicians: Prospective evidence from Swiss residents. *International Archives of Occupational and Environmental Health*, *82*(1), 31–38. https://doi.org/10.1007/s00420-008-0303-7

Busch, C., Deci, N., & Laackmann, M. (2013). Teamwork and health: The influence of team characteristics on coping and health among low-qualified workers. *Gruppendynamik Und Organisationsberatung*, *44*(2), 133–151. https://doi.org/10.1007/s11612-013-0209-z

Buttigieg, S. C., West, M. a, & Dawson, J. F. (2011). Well-structured teams and the buffering of hospital employees from stress. *Health Services Management Research*, *24*(4), 203–212. https://doi.org/10.1258/hsmr.2011.011013

Byrne, B. M. (2012). *Structural equation modeling with Mplus*. New York, NY: Routledge.

Care Quality Commission. (2016). *2015 Adult Inpatient Survey Statistical Release* (Vol. 2016). Newcastle, UK. Retrieved from http://www.cqc.org.uk/sites/default/files/20150822_ip15_statistical_release_corrected.pdf

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116. https://doi.org/10.3844/jssp.2007.106.116

Carpenter, L. M., Swerdlow, A. J., & Fear, N. T. (1997). Mortality of doctors in different specialties: Findings from a cohort of 20000 NHS hospital consultants. *Occupational and Environmental Medicine*, *54*(6), 388–395. https://doi.org/10.1136/oem.54.6.388

Chambers, C., Frampton, C., & Barclay, M. (2017). Presenteeism in the New Zealand senior medical workforce - A mixed methods analysis. *New Zealand Medical Journal*, *130*(1449), 10–22.

Chen, C. S., Liu, T. C., Lin, H. C., & Lien, Y. C. (2008). Does high surgeon and hospital surgical volume raise the five-year survival rate for breast cancer? A population-based study. *Breast Cancer Research and Treatment*, *110*, 349–356. https://doi.org/10.1007/s10549-007-9715-4

Chou, L.-P., Li, C.-Y., & Hu, S. C. (2014). Job stress and burnout in hospital employees: comparisons of different medical professions in a regional hospital in Taiwan. *BMJ Open*, *4*(2), e004185. https://doi.org/10.1136/bmjopen-2013-004185

Christian, M. S., Bradley, J. C., Wallace, J. C., &

Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*(5), 1103–1127. https://doi.org/10.1037/a0016172

Clarke, R. T., Pitcher, A., Lambert, T. W., & Goldacre, M. J. (2014). UK doctors' views on the implementation of the European Working Time Directive as applied to medical practice: A qualitative analysis. *BMJ Open*, *4*, e004390. https://doi.org/10.1136/bmjopen-2013-004390

Cohen, J. (1960). A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. *Second Edition* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, S. (1980). After effects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin*, *88*(1), 82–108. https://doi.org/10.1016/j.theriogenology.2009.10.003

College of Emergency Medicine. (2014). Crowding in emergency departments. Retrieved May 26, 2017, from https://secure.rcem.ac.uk/code/document.asp?ID=6296%0A

Comrey, A. L., & Lee, H. B. (2012). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Costa, P. L., Passos, A. M., & Bakker, A. B. (2015). Direct and contextual influence of team conflict on team resources, team work engagement, and team performance. *Negotiation and Conflict Management Research*, *8*(4), 211–227. https://doi.org/10.1111/ncmr.12061

Cox, T. (1978). *Stress*. Sudbury, UK: MacMillan.

Cox, T. (1993). *Stress research and stress management: Putting theory to work*. Sudbury, UK: HSE Books.

Cox, T., & Griffiths, A. (1995). The nature and measurement of work stress: Theory and practice. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work: A Practical ergonomics methodology* (pp. 783–803). London, UK: Taylor & Francis.

Cox, T., & Griffiths, A. (2010). Work-related stress: A theoretical perspective. In S. Leka & J. Houdmont (Eds.), *Occupational Health Psychology* (pp. 31–56). Chichester, UK: Wiley-Blackwell.

Cox, T., Griffiths, A., & Rial-González, E. (2000). *Research on work-related stress*. Bilbao, Spain: European Agency for Safety and Health at Work. https://doi.org/http://europa.eu.int

Cox, T., Karanika-Murray, M., Griffiths, A., Wong, Y. Y. V., & Hardy, C. (2009). *Developing the management standards approach within the context of common health problems in the workplace: A Delphi study*. Buxton, UK: HSE Books.

Cox, T., Taris, T. W., & Nielsen, K. (2010). Organizational interventions: Issues and challenges. *Work & Stress*, *24*(3), 217–218. https://doi.org/10.1080/02678373.2010.519496

Coyle, J., & Williams, B. (1999). Seeing the wood for the trees: Defining the forgotten concept of patient dissatisfaction in the light of patient satisfaction research. *International Journal of Health Care Quality Assurance Incorporating Leadership in Health Services*, *12*(4), i–ix. https://doi.org/10.1108/13660759910298707

Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology*, *95*(5), 834–848. https://doi.org/10.1037/a0019364

Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, *12*(1), 45–57. https://doi.org/10.1037/1082-989X.12.1.45

Crow, R., Gage, H., Hampson, S., Hart, J., Kimber, A., Storey, L., & Thomas, H. (2002). The measurement of satisfaction with healthcare: Implications for practice from a systematic review of the literature. *Health Technology Assessment*, *6*(32).

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Cylus, J., Richardson, E., Findley, L., Longley, M., O'Neil, C., & Steel, D. (2015). United

Kingdom: Health system review. *Health Systems in Transition*, *17*(5), 1–125.

Davies, H. T. O., Hodges, C.-L., & Rundall, T. G. (2003). Views of doctors and managers on the doctor-manager relationship in the NHS. *BMJ (Clinical Research Ed.)*, *326*(7390), 626–628. https://doi.org/10.1136/bmj.326.7390.626

Dawson, J. F. (2009). *Does the experience of staff working in the NHS link to the patient experience of care? An analysis of links between the 2007 acute trust*. Birmingham, UK: Aston Business School.

De Leeuw, J. (2006). Nonlinear principal component analysis and related techniques. *Multiple Correspondence Analysis and Related Methods*, 107–133. https://doi.org/10.1109/IJCNN.2003.1223477

de Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J., & Boermeester, M. A. (2008). The incidence and nature of in-hospital adverse events: A systematic review. *Quality and Safety in Health Care*, *17*(3), 216–223. https://doi.org/10.1136/qshc.2007.023622

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Publishing Co.

Demerouti, E., & Bakker, A. B. (2011). The job demands–resources model: Challenges for future research. *SA Journal of Industrial Psychology*, *37*(2), 1–9. https://doi.org/10.4102/sajip.v37i2.974

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499–512. https://doi.org/10.1108/02683940710733115

Demerouti, E., Le Blanc, P. M., Bakker, A. B., Schaufeli, W. B., & Hox, J. (2009). Present but sick: A three-wave study on job demands, presenteeism and burnout. *Career Development International*, *14*(1), 50–68. https://doi.org/10.1108/13620430910933574

Department of Health. (2008). *High quality care for all: The NHS next stage review final report*. London, UK: The Stationary Office.

Department of Health. (2010). *Equity and excellence: Liberating the NHS (White Paper)*. London, UK: The Stationary Office. https://doi.org/10.1136/adc.2010.205294

Department of Health. (2017). Domain 2: Enhancing quality of life for people with long-term conditions. Retrieved July 22, 2017, from https://www.england.nhs.uk/resources/resources-for-ccgs/out-frwrk/dom-2/

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousands Oaks, CA: Sage Publications, Inc.

DeVellis, R. F. (2006). Classical test theory. *Medical Care*, *44*(11 Suppl 3), S50–S59. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Dewe, P., & Trenberth, L. (2004). Work stress and coping: Drawing together research and practice. *British Journal of Guidance & Counselling*, *32*(2), 143–156. https://doi.org/10.1080/03069880410001692201

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*(3), 434–449. https://doi.org/10.1007/s11747-011-0300-3

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(3), 327–346. https://doi.org/10.1207/S15328007SEM0903_2

Dixon-Woods, M., Baker, R., Charles, K., Dawson, J. F., Jerzembek, G., Martin, G. P., … West, M. A. (2014). Culture and behaviour in the English National Health Service: Overview of lessons from a large multimethod study. *BMJ Quality & Safety*, *0*, 1–10. https://doi.org/10.1136/bmjqs-2013-001947

Dollard, M. F., & Bakker, A. B. (2010). Psychosocial safety climate as a precursor to conducive work environments, psychological health problems, and employee engagement. *Journal of Occupational and Organizational Psychology*, *83*(3), 579–599. https://doi.org/10.1348/096317909x470690

Dollarhide, A. W., Rutledge, T., Weinger, M. B., Fisher, E. S., Jain, S., Wolfson, T., &

Dresselhaus, T. R. (2013). A real-time assessment of factors influencing medication events. *Journal for Healthcare Quality*, *36*(5), 5–12. https://doi.org/10.1111/jhq.12012

Dornan, T., Ashcroft, D. M., Lewis, P., Miles, J., Taylor, D., & Tully, M. (2010). An in depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education. EQUIP study. *Methods*, *44*(0), 1–215. https://doi.org/10.1177/1091581811407934

Dow, A. W., DiazGranados, D., Mazmanian, P. E., & Retchin, S. M. (2013). Applying organizational science to health care: A framework for collaborative practice. *Academic Medicine*, *88*(7), 952–957. https://doi.org/10.1097/ACM.0b013e31829523d1

Downs, J., Gilbert, R., Hayes, R. D., Hotopf, M., & Ford, T. (2017). Linking health and education data to plan and evaluate services for children. *Archives of Disease in Childhood*, *102*(7), 599–602. https://doi.org/10.1136/archdischild-2016-311656

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behaviour in context: A multilevel modelling approach. *Social Science and Medicine*, *42*(6), 817–830. https://doi.org/10.1016/0277-9536(95)00181-6

Dunn, P., McKenna, H., & Murray, R. (2016). *Deficits in the NHS 2016*. London, UK: The King's Fund. Retrieved from http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/Deficits_in_the_NHS_Kings_Fund_July_2016_1.pdf

Eastwood, J., Jalaludin, B., Kemp, L., Phung, H., Barnett, B., & Tobin, J. (2013). Social exclusion, infant behavior, social isolation, and maternal expectations independently predict maternal depressive symptoms. *Brain and Behavior*, *3*(1), 14–23. https://doi.org/10.1002/brb3.107

Eckleberry-Hunt, J., Kirkpatrick, H., & Barbera, T. (2017). The problems with burnout research. *Academic Medicine*, (Advanced Online Publication). https://doi.org/10.1097/ACM.0000000000001890

Einarsen, S., Hoel, H., Zapf, D., & Cooper, C. L. (2010). The concept of bullying and harassment at work: The European tradition. In S. Einarsen, H. Hoel, D. Zapf, & C. L. Cooper (Eds.), *Bullying and harassment in the workplace (2nd edition)* (pp. 3–32). Boca Raton, FL: Taylor & Francis. https://doi.org/10.1108/09596110010343666

Eisinga, R., Grotenhuis, M. Te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Escribà-Agüir, V., Martín-Baena, D., & Pérez-Hoyos, S. (2006). Psychosocial work environment and burnout among emergency medical and nursing staff. *International Archives of Occupational and Environmental Health*, *80*(2), 127–133. https://doi.org/10.1007/s00420-006-0110-y

Fahrenkopf, A. M., Sectish, T. C., Barger, L. K., Sharek, P. J., Lewin, D., Chiang, V. W., … Landrigan, C. P. (2008). Rates of medication errors among depressed and burnt out residents: Prospective cohort study. *BMJ*, *336*(7642), 488–491. https://doi.org/10.1136/bmj.39469.763218.BE

Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on SEM fit indices. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 56–83.

Feddock, C. A., Hoellein, A. R., Griffith, C. H., Wilson, J. F., Becker, N. S., Bowerman, J. L., & Caudill, T. S. (2005). Are continuity clinic patients less satisfied when residents have a heavy inpatient workload? *Evaluation & the Health Professions*, *28*(4), 390–399. https://doi.org/10.1177/0163278705281070

Federation of the Royal Colleges of Physicians of the UK. (2016). *Census of consultant physicians and higher specialty trainees in the UK 2014–15: Data and commentary*. London, UK: Royal College of Physicians. Retrieved from https://www.rcplondon.ac.uk/projects/outputs/2013-14-census-uk-consultants-and-higher-specialty-trainees

Field, A. (2014). *Discovering statistics using IBM SPSS statistics.* (Vol. 4). London, UK: Sage Publishers.

Firth-Cozens, J., & Greenhalgh, J. (1997). Doctors' perceptions of the links between stress and lowered clinical care. *Social Science and Medicine*, *44*(7), 1017–1022. https://doi.org/10.1016/S0277-9536(96)00227-4

Fletcher, K. E., Underwood, W., Davis, S. Q., Mangrulkar, R. S., McMahon, L. F., & Saint, S. (2005). Effects of work hour reduction on residents' lives: A systematic review. *JAMA*, *294*(9), 1088–1100. https://doi.org/10.1001/jama.294.9.1088

Flinn, F., & Armstrong, C. (2011). Junior doctors' extended work hours and the effects on their performance: The Irish case. *International Journal for Quality in Health Care*, *23*(2), 210–217. https://doi.org/10.1093/intqhc/mzq088

Flora, D. B., & Curran, P. J. (2004). An evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466.An

Flynn, T., Fletcher, G., McGeorge, P., Sutherland, A., & Patey, R. (2003). Anaesthetists' attitudes to teamwork and safety. *Anaesthesia*, *58*(3), 233–242.

Forster, A. J., Murff, H. J., Peterson, J. F., Gandhi, T. K., & Bates, D. W. (2003). The incidence and severity of adverse events affecting patients after discharge from the hospital. *Annals of Internal Medicine*, *138*(3), 161–167. https://doi.org/200302040-00007 [pii]

Francis, R. (2013). *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive summary*. London, UK: The Stationary Office. https://doi.org/10.1002/yd.20044

Frank, E., Biola, H., & Burnett, C. A. (2000). Mortality rates and causes among U.S. physicians. *American Journal of Preventive Medicine*, *19*(3), 155–159. https://doi.org/10.1016/S0749-3797(00)00201-4

Fuß, I., Nübling, M., Hasselhorn, H.-M., Schwappach, D., & Rieger, M. A. (2008). Working conditions and work-family conflict in German hospital physicians: Psychosocial and organisational predictors and consequences. *BMC Public Health*, *8*(1), 353. https://doi.org/10.1186/1471-2458-8-353

Gainsbury, S. (2016). *Feeling the crunch: NHS finances to 2020*. London, UK: The Nuffield Trust. Retrieved from http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/feeling_the_crunch_nhs_finances_to_2020_summary_web_correction.pdf

Gander, P., Purnell, H., Garden, A., & Woodward, A. (2007). Work patterns and fatigue-related risk among junior doctors. *Occupational and Environmental Medicine*, *64*(11), 733–738. https://doi.org/10.1136/oem.2006.030916

Garman, A. N., Corrigan, P. W., & Morris, S. (2002). Staff burnout and patient satisfaction: Evidence of relationships at the care unit level. *Journal of Occupational Health Psychology*, *7*(3), 235–241. https://doi.org/10.1037/1076-8998.7.3.235

Gaskin, J. (2017). Confirmatory factor analysis: Common method bias. Retrieved July 18, 2017, from http://statwiki.kolobkreations.com/index.php?title=Confirmatory_Factor_Analysis#Common_Method_Bias_.28CMB.29

General Medical Council. (2014). *National training survey 2014: Bullying and undermining*. Manchester, UK: Geneal Medical Council. Retrieved from http://www.gmc-uk.org/NTS_bullying_and_undermining_report_2014_FINAL.pdf_58648010.pdf

General Medical Council. (2016). *National training survey 2016*. Manchester, UK. Retrieved from http://www.gmc-uk.org/National_training_survey_2016_England.pdf_66922358.pdf

Gilboa, S., Shirom, A., Fried, Y., & Cooper, C. L. (2008). A meta-analysis of work demand stressors and job performance: Examining main and moderating effects. *Personnel Psychology*, *61*(2), 227–271. https://doi.org/10.1111/j.1744-6570.2008.00113.x

Goddard, A. F. (2010). Planning a consultant delivered NHS. *BMJ*, *341*(2), c6034. https://doi.org/10.1136/bmj.c6034

Goddard, A. F. (2016). Lessons to be learned from the UK junior doctors' strike. *JAMA*, *316*(14), 1445–1446. https://doi.org/10.1001/jama.2016.12029

Godlee, F. (2008). Understanding the role of the doctor. *BMJ*, *337*, a3035.

https://doi.org/10.1136/bmj.a3035

Goehring, C., Gallacchi, M. B., Künzi, B., & Bovier, P. A. (2005). Psychosocial and professional characteristics of burnout in Swiss primary care practitioners: A cross-sectional survey. *Swiss Medical Weekly*, *135*(7–8), 101–108. https://doi.org/2005/07/smw-10841

Goldberg, L. R., & Digman, J. M. (1994). Revealing structure in the data: Principles of exploratory factor analysis. In S. Strack & M. Lorr (Eds.), *Differentiating Normal and Abnormal Personality* (pp. 216–242). New York, NY: Springer.

González-Morales, M. G., Peiró, J. M., Rodríguez, I., & Bliese, P. D. (2012). Perceived collective burnout: A multilevel explanation of burnout. *Anxiety, Stress & Coping*, *25*(1), 43–61. https://doi.org/10.1080/10615806.2010.542808

Goodwin, L. (1999). The role of factor analysis in the estimation of construct validity. *Measurement in Physical Education and Exercise Science*, *3*(2), 85–100. https://doi.org/10.1207/s15327841mpee0302_2

Gordon, H. J., Demerouti, E., Le Blanc, P. M., Bakker, A. B., Bipp, T., & Verhagen, M. A. M. T. (2017). Individual job redesign: Job crafting interventions in healthcare. *Journal of Vocational Behavior*. https://doi.org/10.1016/j.jvb.2017.07.002

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gosden, T. (1999). How should we pay doctors? A systematic review of salary payments and their effect on doctor behaviour. *QJM*, *92*(1), 47–55. https://doi.org/10.1093/qjmed/92.1.47

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*(3–4), 325–338. https://doi.org/10.1093/biomet/53.3-4.325

Grant, A., Rix, A., Mattick, K., Jones, D., & Winter, P. (2013). *Identifying good practice among medical schools in the support of students with mental health concerns*. Manchester, UK: General Medical Council.

Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(2), 108–120. https://doi.org/10.1080/10705519709540064

Griffith-Noble, F. (2010). *A multi-method investigation of the psychosocial work environment and nature of work-related stress of NHS physiotherapists and occupational therapists*. University of Nottingham. Retrieved from http://eprints.nottingham.ac.uk/13279/2/537794_vol2.pdf

Grol, R., Mokkink, H., Smits, A., & Van, E. (1985). Work satisfaction of general practitioners and the quality of patient care. *Family Practice*, *2*(3), 128–135. https://doi.org/10.1093/fampra/2.3.128

Groß, S. E., Ernstmann, N., Jung, J., Karbach, U., Ansmann, L., Gloede, T. D., … Neumann, M. (2014). Can a stressed oncologist be good in a consultation? A qualitative study on the oncologists' perception. *European Journal of Cancer Care*, *23*(5), 594–606. https://doi.org/10.1111/ecc.12199

Grunfeld, E., Whelan, T. J., Zitzelsberger, L., Willan, A. R., Montesanto, B., & Evans, W. K. (2000). Cancer care workers in Ontario: Prevalence of burnout, job stress and job satisfaction. *CMAJ: Canadian Medical Association Journal*, *163*(2), 166–169. https://doi.org/10.1111/ecc.12098

Guthrie, E., Tattan, T., Williams, E., Black, D., & Bacliocotti, H. (1999). Sources of stress, psychological distress and burnout in psychiatrists: Comparison of junior doctors, senior registrars and consultants. *Psychiatric Bulletin*, *23*(4), 207–212. https://doi.org/10.1192/pb.23.4.207

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (New Intern). Essex, UK: Pearson Education Limited.

Hakanen, J. J., Bakker, A. B., & Demerouti, E. (2005). How dentists cope with their job demands and stay engaged: The moderating role of job resources. *European Journal of Oral Sciences*, *113*(6), 479–487. https://doi.org/10.1111/j.1600-0722.2005.00250.x

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology*, *43*(6), 495–513. https://doi.org/10.1016/j.jsp.2005.11.001

Hakanen, J. J., Schaufeli, W. B., & Ahola, K. (2008). The Job Demands-Resources model: A three-year cross-lagged study of burnout, depression, commitment, and work engagement. *Work & Stress*, *22*(3), 224–241. https://doi.org/10.1080/02678370802379432

Halbesleben, J. R. B., & Rathert, C. (2008). Linking physician burnout and patient outcomes: Exploring the dyadic relationship between physicians and patients. *Health Care Management Review*, *33*(1), 29–39. https://doi.org/10.1097/01.HMR.0000304493.87898.72

Hall, L. H., Johnson, J., Watt, I., Tsipa, A., & O'Connor, D. B. (2016). Healthcare staff wellbeing, burnout, and patient safety: A systematic review. *PLoS ONE*, *11*(7), 1–12. https://doi.org/10.1371/journal.pone.0159015

Hansen, N., Sverke, M., & Näswall, K. (2009). Predicting nurse burnout from demands and resources in three acute care hospitals under different forms of ownership: A cross-sectional questionnaire survey. *International Journal of Nursing Studies*, *46*(1), 95–106. https://doi.org/10.1016/j.ijnurstu.2008.08.002

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development. Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research*, *57*(2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8

Harley, K. T., Streja, E., Rhee, C. M., Molnar, M. Z., Kovesdy, C. P., Amin, A. N., & Kalantar-Zadeh, K. (2013). Nephrologist caseload and hemodialysis patient survival in an urban cohort. *Journal of the American Society of Nephrology : JASN*, *24*, 1678–1687. https://doi.org/10.1681/ASN.2013020123

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology*, *87*(2), 268–279. https://doi.org/10.1037//0021-9010.87.2.268

Hassard, J., Teoh, K. R.-H., Visockaite, G., Dewe, P., & Cox, T. (2017a). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology*, (Advanced Online Publication). https://doi.org/10.1037/ocp0000069

Hassard, J., Teoh, K. R.-H., Visockaite, G., Dewe, P., & Cox, T. (2017b). The financial burden of psychosocial workplace aggression: A systematic review of cost of illness studies. *Work & Stress*, (Advanced Online Publication). https://doi.org/10.1080/02678373.2017.1380726

Häusser, J. A., Mojzisch, A., Niesel, M., & Schulz-Hardt, S. (2010). Ten years on: A review of recent research on the Job Demand–Control (-Support) model and psychological well-being. *Work & Stress*, *24*(1), 1–35. https://doi.org/10.1080/02678371003683747

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

Hayes, B., Prihodova, L., Walsh, G., Doyle, F., & Doherty, S. (2017). What's up doc? A national cross- sectional study of psychological well-being of hospital doctors in Ireland. *BMJ Open*, (Advanced Online Publication). https://doi.org/10.1136/bmjopen-2017-018023

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A.-H. S., Dellinger, E. P., … Gawande, A. A. (2011). Changes in safety attitude and relationship to decreased postoperative morbidity and mortality following implementation of a checklist-based surgical safety intervention. *BMJ Quality & Safety*, *20*(1), 102–107. https://doi.org/10.1136/bmjqs.2009.040022

Health Education England. (2017). *Enhancing junior doctors' working lives: A progress report*. Winchester, UK: Health Education England.

Health Foundation. (2011). *Measuring safety culture. Research scan*. London. Retrieved from http://www.health.org.uk/sites/health/files/MeasuringSafetyCulture.pdf

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques*. New York, NY: Routledge.

Hedges, L. V., & Olkin, I. O. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press, Inc.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504.

Henderson, A. D. (2003). Nurses and workplace violence: Nurses' experiences of verbal and physical abuse at work. *Nursing Leadership*, *16*(4), 82–98.

Hershcovis, M. S. (2011). "Incivility, social undermining, bullying…oh my!": A call to reconcile constructs within workplace aggression research. *Journal of Organizational Behavior*, *32*(3), 499–519. https://doi.org/10.1002/job.689

Higgins, J. P. T., & Deeks, J. J. (2011). Chapter 7: Selecting studies and collecting data. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (5.0.0). The Cochrane Collaboration. Retrieved from http://handbook.cochrane.org/chapter_7/7_selecting_studies_and_collecting_data.htm

Hill, J., Rolfe, I. E., Pearson, S.-A., & Heathcote, A. (1998). Do junior doctors feel they are prepared for hospital practice? A study of graduates from traditional and non-traditional medical schools. *Medical Education*, *32*(1), 19–24. https://doi.org/10.1046/j.1365-2923.1998.00152.x

Hillhouse, J. J., Adler, C. M., & Walters, D. N. (2000). A simple model of stress , burnout and symptomatology in medical residents: A longitudinal study. *Psychology, Health & Medicine*, *5*(1), 63–73. https://doi.org/10.1080/135485000106016

Hillier, D. F., Parry, G. J., Shannon, M. W., & Stack, A. M. (2009). The effect of hospital bed occupancy on throughput in the pediatric emergency department. *Annals of Emergency Medicine*, *53*(6), 767–776.e3. https://doi.org/10.1016/j.annemergmed.2008.11.024

Hobfoll, S. E. (1989). Conservation of resources. A new attempt at conceptualizing stress. *The American Psychologist*, *44*(3), 513–524. https://doi.org/10.1037/0003-066X.44.3.513

Hobfoll, S. E. (2002). Social and psychological resources and adaptation. *Review of General Psychology*, *6*(4), 307–324. https://doi.org/10.1037/1089-2680.6.4.307

Hobfoll, S. E., Johnson, R. J., Ennis, N., & Jackson, A. P. (2003). Resource loss, resource gain, and emotional outcomes among inner city women. *Journal of Personality and Social Psychology*, *84*(3), 632–643. https://doi.org/10.1037/0022-3514.84.3.632

Hobfoll, S. E., & Shirom, A. (2001). Conservation of resources theory: Applications to stress and management in the workplace. In R. T. Golembiewski (Ed.), *Handbook of organizational behavior* (2nd ed., pp. 57–80). Newbury Park, CA: Dekker.

Hochschild, A. R. (1983). *The managed heart: commercialization of human feeling*. Berkeley, CA: University of California Press.

Hockey, G. J. (1993). Cognitive-energetical control mechanisms in the management of work demands and psychological health. In A. B. Baddely & L. Weiskrantz (Eds.), *Attention, selection, awareness and control: A tribute to Donald Broadbent* (pp. 328–345). New York, NY: Clarendon Press/Oxford University Press.

Hoff, T., Jameson, L., Hannan, E., & Flink, E. (2004). A review of the literature examining linkages between organizational factors, medical errors, and patient safety. *Medical Care Research and Review : MCRR*, *61*(1), 3–37. https://doi.org/10.1177/1077558703257171

Hogan, H., Healey, F., Neale, G., Thomson, R., Vincent, C., & Black, N. (2013). Preventable deaths due to problems in care in English acute hospitals: A retrospective case record review study. *BMJ Quality & Safety*, *21*(21), 737–745. https://doi.org/10.1136/bmjqs-2012-001159

Hogan, H., Zipfel, R., Neuburger, J., Hutchings, A., Darzi, A., & Black, N. (2015). Avoidability of hospital deaths and association with hospital-wide mortality ratios: Retrospective case record review and regression analysis. *BMJ (Clinical Research Ed.)*, *351*(February), 1–6. https://doi.org/10.1136/bmj.h3239

Howell, A.-M., Burns, E. M., Bouras, G., Donaldson, L. J., Athanasiou, T., & Darzi, A. (2015). Can patient safety incident reports be used to compare hospital safety? Results from a quantitative analysis of the English National Reporting and Learning System Data. *PLOS ONE*, *10*(12), e0144107. https://doi.org/10.1371/journal.pone.0144107

Howie, J. G. R., Hopton, J. L., Heaney, D. J., & Porter, A. M. D. (1992). Attitudes to medical care, the organization of work, and stress among general practitioners. *British Journal of General Practice*, *42*(358), 181–185.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Hove, UK: Routledge.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(2), 157–174. https://doi.org/10.1207/S15328007SEM0802_1

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*(2), 157–170. https://doi.org/10.1111/j.1467-9574.2009.00445.x

HSCIC. (2015a). *Hospital Episode Statistics Admitted Patient Care, England 2014-2015*. Leeds, UK: Health and Social Care Information Centre. Retrieved from http://www.hscic.gov.uk/catalogue/PUB19124/hosp-epis-stat-admi-summ-rep-2014-15-rep.pdf

HSCIC. (2015b). *Summary hospital-level mortality indicator (SHMI) – Deaths associated with hospitalisation* (Quarterly Report, July 2013-June 2014). Leeds, UK.

HSCIC. (2016). *Hospital Episode Statistics: Accident and Emergency Attendances in England 2014-15*. Leeds, UK: Health and Social Care Information Centre. Retrieved from http://content.digital.nhs.uk/catalogue/PUB19883/acci-emer-atte-eng-2014-15-rep.pdf

HSE. (2017). What are the Management Standards. Retrieved August 29, 2017, from http://www.hse.gov.uk/stress/standards/index.htm

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, Q., Schaufeli, W. B., & Taris, T. W. (2011). The Job Demands-Resources model: An analysis of additive and joint effects of demands and resources. *Journal of Vocational Behavior*, *79*(1), 181–190. https://doi.org/10.1016/j.jvb.2010.12.009

Hughes, D., & Clarke, V. (2016, February 29). Thousands of NHS nursing and doctor posts lie vacant. Retrieved from www.bbc.co.uk/news/health-35667939

Huth, E. J., & Murray, T. J. (2006). *Medicine in quotations: Views of health and disease through the ages*. Philadelphia, PA: American College of Physicians.

Ironson, G. H., Smith, P. C., Brannick, M. T., Gibson, W. M., & Paul, K. B. (1989). Construction of a job in general scale: A comparison of global, composite, and specific measures. *Journal of Applied Psychology*, *74*(2), 193–200. https://doi.org/10.1037/0021-9010.74.2.193

Iversen, A., Rushforth, B., & Forrest, K. (2009). How to handle stress and look after your mental health. *BMJ*, *338*(May), 1139–1144. https://doi.org/10.1136/bmj.b1368

Jagsi, R., & Surender, R. (2004). Regulation of junior doctors' work hours: An analysis of British and American doctors' experiences and attitudes. *Social Science and Medicine*, *58*(11), 2181–2191. https://doi.org/10.1016/j.socscimed.2003.08.016

Jansen, A., Beyer, A., Brandt, C., Höhne, M., Schreier, E., Schulzke, J., … Schneider, T. (2004). Die Norovirus-Epidemie in Berlin - Klinik, Epidemiologie und Prävention. *Zeitschrift Für Gastroenterologie*, *42*(4), 311–316. https://doi.org/10.1055/s-2004-812908

Jena, A. B., Baldwin, D. C., Daugherty, S. R., Meltzer, D. O., & Arora, V. M. (2010). Presenteeism among resident physicians. *JAMA*, *304*(11), 1166. https://doi.org/10.1001/jama.2010.1315

Jex, S. M. (1998). *Stress and job performance: Theory, research, and implications for managerial practice*. Thousands Oak, CA: Sage.

Johnson, J., & Hall, E. (1988). Job strain, work place social support, and cardiovascular disease: A cross-sectional study of a random sample of the Swedish working population. *American Journal of Public Health*, *78*(10), 1336–1342. https://doi.org/10.2105/AJPH.78.10.1336

Jollife, I. T. (2002). *Principle component analysis* (2nd ed.). New York, NY: Springer.

Joules, N., Williams, D. M., & Thompson, A. W. (2014). Depression in resident physicians: A systematic review. *Open Journal of Depression*, *3*(3), 89–100. https://doi.org/10.4236/ojd.2014.33013

Jourdain, G., & Chênevert, D. (2010). Job demands-resources, burnout and intention to leave the nursing profession: A questionnaire survey. *International Journal of Nursing Studies*, *47*(6), 709–722. https://doi.org/10.1016/j.ijnurstu.2009.11.007

Joyce, C. M., Schurer, S., Scott, A., Humphreys, J., & Kalb, G. (2011). Australian doctors' satisfaction with their work: Results from the MABEL longitudinal survey of doctors. *The Medical Journal of Australia*, *194*(1), 30–33.

Judge, T. A., & Klinger, R. (2008). Job satisfaction: Subjective well-being at work. *The Science of Subjective Well-Being*, 393–413. https://doi.org/10.1111/j.1475-682X.2000.tb00904.x

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*(3), 376–407. https://doi.org/10.1037/0033-2909.127.3.376

Kahn, R. L., & Byosiere, P. B. (1992). Stress in organizations. In M. D. Dunnette & L. M. Hughs (Eds.), *Handbook of Industrial and Organizational Psychology.* Palo Alto, CA: Consulting Psychologists Press.

Kaier, K., Mutters, N. T., & Frank, U. (2012). Bed occupancy rates and hospital-acquired infections—should beds be kept empty? *Clinical Microbiology and Infection*, *18*(10), 941–945. https://doi.org/10.1111/j.1469-0691.2012.03956.x

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151. https://doi.org/10.1177/001316446002000116

Karasek, R., & Theorell, T. (1990). *Healthy work: Stress, productivity and the reconstruction of working life*. New York, NY: Basic Books. https://doi.org/10.1016/0003-6870(92)90320-U

Kazanjian, A., Green, C., Wong, J., & Reid, R. (2005). Effect of the hospital nursing environment on patient mortality: A systematic review. *Journal of Health Services Research & Policy*, *10*(2), 111–117. https://doi.org/10.1258/1355819053559100

Kenny, D. A. (2016). Mediation. Retrieved May 11, 2017, from http://davidakenny.net/cm/mediate.htm#IM

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology, Vol. 1* (4th ed., pp. 233–265). New York, NY: McGraw-Hill.

Keogh, B. (2013). *Review into the quality of care and treatment provided by 14 hospital trusts in England: Overview report*. London, UK: The Stationary Office. https://doi.org/10.1613/jair.301

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *JRSM*, *96*(3), 118–121. https://doi.org/10.1258/jrsm.96.3.118

Khan, K. S., Riet, G. Ter, Glanville, J., Sowden, A. J., & Kleijnen, J. (2009). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*. York, UK: Centre for Reviews and Disseminations. Retrieved from http://eprints.whiterose.ac.uk/1139/

Kilminster, S. M., & Jolly, B. C. (2000). Effective supervision in clinical practice settings: A literature review. *Medical Education*, *34*(10), 827–840. https://doi.org/10.1046/j.1365-2923.2000.00758.x

King's Fund. (2011). *Defining and measuring the quality of general practice*. *Improving the quality of care in general practice*. London, UK: The King's Fund.

King's Fund. (2016). *How is the NHS performing? Quarterly Monitoring Report*. London, UK: The King's Fund. Retrieved from http://qmr.kingsfund.org.uk/2016/20/

Kirkup, B. (2015). *The report of the Morecambe Bay investigation*. London, UK: The Stationary Office. Retrieved from https://www.gov.uk/government/publications

Kivimäki, M., Leino-Arjas, P., Luukkonen, R., Riihimaki, H., Vahtera, J., & Kirjonen, J. (2002). Work stress and risk of cardiovascular mortality: Prospective cohort study of industrial employees. *BMJ*, *325*(7369), 857. https://doi.org/10.1136/bmj.325.7369.857

Kivimäki, M., Sutinen, R., Elovainio, M., Vahtera, J., Rasanen, K., Toyry, S., … Firth-Cozens, J.

(2001). Sickness absence in hospital physicians: 2 year follow up study on determinants. *Occupational and Environmental Medicine*, *58*(6), 361–366. https://doi.org/10.1136/oem.58.6.361

Kivimäki, M., Virtanen, M., Elovainio, M., Kouvonen, A., Väänänen, A., & Vahtera, J. (2006). Work stress in the etiology of coronary heart disease—A meta-analysis. *Scandinavian Journal of Work, Environment and Health*, *32*(6), 431–442. https://doi.org/10.5271/sjweh.1049

Klein, J., Frie, K. G., Blum, K., & von dem Knesebeck, O. (2011). Psychosocial stress at work and perceived quality of care among clinicians in surgery. *BMC Health Services Research*, *11*(1), 109–123. https://doi.org/10.1186/1472-6963-11-109

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. *Nature* (Vol. 156). New York, NY: Guildford Press.

Knight, C., Patterson, M., & Dawson, J. F. (2017). Building work engagement: A systematic review and meta-analysis investigating the effectiveness of work engagement interventions. *Journal of Organizational Behavior*, *38*(6), 792–812. https://doi.org/10.1002/job.2167

Koziol, N., & Athur, A. (2011). *An introduction to secondary data analysis* (Research Methodology Series). Nebreska, NA. Retrieved from http://r2ed.unl.edu/presentations/2011/RMS/120911_Koziol/120911_Koziol.pdf

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Krämer, T., Schneider, A., Spieß, E., Angerer, P., & Weigl, M. (2016). Associations between job demands, work-related strain and perceived quality of care: A longitudinal study among hospital physicians. *International Journal for Quality in Health Care*, *28*(6), 824–829. https://doi.org/10.1093/intqhc/mzw119

Krueger, L., Funk, C., Green, J., & Kuznar, K. (2013). Nurse-related variables associated with patient outcomes: A review of the literature 2006–2012. *Teaching and Learning in Nursing*, *8*(4), 120–127. https://doi.org/10.1016/j.teln.2013.05.001

Kumpunen, S., Currey, N., Edwards, N., Holmes, M., Price, H., & Ballard, T. (2015). *Collaboration in general practice: Surveys of GPs and CCGs*. London, UK: Royal College of General Practitioners.

Kuper, H., & Marmot, M. G. (2003). Job strain, job demands, decision latitude, and risk of coronary heart disease within the Whitehall II study. *Journal of Epidemiology and Community Health*, *57*(2), 147–153. https://doi.org/10.1136/jech.57.2.147

Lambrou, P., Kontodimopoulos, N., & Niakas, D. (2010). Motivation and job satisfaction among medical and nursing staff in a Cyprus public general hospital. *Human Resources for Health*, *8*(1), 26. https://doi.org/10.1186/1478-4491-8-26

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Lang, T. A., Hodge, M., Olson, V., Romano, P. S., & Kravitz, R. L. (2004). Nurse-patient ratios: A systematic review on the effects of nurse staffing on patient, nurse employee, and hospital outcomes. *The Journal of Nursing Administration*, *34*(7–8), 326–337. https://doi.org/10.1097/00005110-200407000-00005

Larson, E. B., & Yao, X. (2005). Clinical empathy as emotional labor in the patient-physician relationship. *Journal of the American Medical Association*, *293*(9), 1100–1106. https://doi.org/10.1001/jama.293.9.1100

Laschinger, H. K. S., Grau, Ashley, L., Finegan, J., & Wilk, P. (2012). Predictors of new graduate nurses' workplace well-being: Testing the job demands-resources model. *Health Care Management Review*, *37*(2), 175–186. https://doi.org/10.1097/HMR.0b013e31822aa456

Laubach, W., & Fischbeck, S. (2007). Job satisfaction and the work situation of physicians: A survey at a German University hospital. *International Journal of Public Health*, *52*(1), 54–59. https://doi.org/10.1007/s00038-006-5021-x

Lawrence, S., Sullivan, C., Patel, N., Spencer, L., Sinnott, M., & Eley, R. (2016). Admission of medical patients from the emergency department: An assessment of the attitudes, perspectives and practices of internal medicine and emergency medicine trainees. *Emergency Medicine Australasia*, *28*(4), 391–398. https://doi.org/10.1111/1742-6723.12604

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York, NY: McGraw-Hill.

Lazarus, R. S. (1999). *Stress and emotion.* New York, NY: Springer.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York, NY: Springer.

Lee, R. T., Seo, B., Hladkyj, S., Lovell, B. L., & Schwartzmann, L. (2013). Correlates of physician burnout across regions and specialties: A meta-analysis. *Human Resources for Health*, *11*(1), 48. https://doi.org/10.1186/1478-4491-11-48

Leka, S., Cox, T., & Zwetsloot, G. (2008). *The European framework for psychosocial risk management*. *PRIMA-EF. I-WHO Publications, Nottingham*. Nottingham, UK: I-WHO Publications.

Leka, S., & Jain, A. (2010). *Health impact of psychosocial hazards at work: An overview*. Geneva, Switzerland: WHO Press. Retrieved from http://apps.who.int/iris/bitstream/10665/44428/1/9789241500272_eng.pdf

Leka, S., Jain, A., Zwetsloot, G. I. J. M., Andreou, N., & Hollis, D. (2016). Future challenges of occupational safety and health policy-making in the UK. *Policy and Practice in Health and Safety*, *14*(1), 65–80. https://doi.org/10.1080/14773996.2016.1231871

Leka, S., Van Wassenhove, W., & Jain, A. (2015). Is psychosocial risk prevention possible? Deconstructing common presumptions. *Safety Science*, *71*(Part A), 61–67. https://doi.org/10.1016/j.ssci.2014.03.014

LePine, J. A., Podsakoff, N. P., & LePine, M. A. (2005). A meta-analytic test of the challenge stressor-hindrance stressor framework: An explanation for inconsistent relationships among stressors and performance. *Academy of Management Journal*, *48*(5), 764–775. https://doi.org/10.5465/AMJ.2005.18803921

Lewig, K. A., Xanthopoulou, D., Bakker, A. B., Dollard, M. F., & Metzer, J. C. (2007). Burnout and connectedness among Australian volunteers: A test of the Job Demands-Resources model. *Journal of Vocational Behavior*, *71*(3), 429–445. https://doi.org/10.1016/j.jvb.2007.07.003

Lewin, K. (1943). Psychology and the Process of Group Living. *The Journal of Social Psychology*, *17*(1), 113–131. https://doi.org/10.1080/00224545.1943.9712269

Lewis, P. J., Dornan, T., Taylor, D., Tully, M. P., Wass, V., & Ashcroft, D. M. (2009). Prevalence, incidence and nature of prescribing errors in hospital inpatients. *Drug Safety*, *32*(5), 379–389. https://doi.org/10.2165/00002018-200932050-00002

Li, B., Bruyneel, L., Sermeus, W., Van den Heede, K., Matawie, K., Aiken, L. H., & Lesaffre, E. (2013). Group-level impact of work environment dimensions on burnout experiences among nurses: A multivariate multilevel probit model. *International Journal of Nursing Studies*, *50*(2), 281–291. https://doi.org/10.1016/j.ijnurstu.2012.07.001

Lim, K. K. (2014). Impact of hospital mergers on staff job satisfaction: A quantitative study. *Human Resources for Health*, *12*(1), 70. https://doi.org/10.1186/1478-4491-12-70

Lin, H. C., & Lee, H. C. (2008). Caseload volume-outcome relation for pulmonary embolism treatment: Association between physician and hospital caseload volume and 30-day mortality. *Journal of Thrombosis and Haemostasis*, *6*(April), 1707–1712. https://doi.org/10.1111/j.1538-7836.2008.03098.x

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*(1), 114–121. https://doi.org/10.1037/0021-9010.86.1.114

Linting, M., Meulman, J. J., Groenen, P. J. F., & Van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods*, *12*(3), 336–358. https://doi.org/10.1037/1082-989X.12.3.336

Linzer, M., Gerrity, M., Douglas, J. A., McMurray, J. E., Williams, E. S., & Konrad, T. R. (2002). Physician stress: Results from the physician worklife study. *Stress and Health*, *18*(1), 37–42. https://doi.org/10.1002/smi.917

Linzer, M., Manwell, L. B., Williams, E. S., Bobula, J. A., Brown, R. L., Varkey, A. B., … Schwartz, M. D. (2009). Working conditions in primary care: Physician reactions and care quality. *Annals of Internal Medicine*, *151*(1), 28–36.

Locke, E. L. (1970). Job satisfaction and job performance : A theoretical analysis. *Organisational Behavior and Human Performance*, *5*, 484–500. https://doi.org/10.1016/0030-5073(70)90036-X

Loerbroks, A., Weigl, M., Li, J., & Angerer, P. (2016). Effort-reward imbalance and perceived quality of patient care: A cross-sectional study among physicians in Germany. *BMC Public Health*, *16*(April), 342. https://doi.org/10.1186/s12889-016-3016-y

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

Loretto, W., Popham, F., Platt, S., Pavis, S., Hardy, G., Macleod, L., & Gibbs, J. (2005). Assessing psychological well-being: A holistic investigation of NHS employees. *International Review of Psychiatry*, *17*(5), 329–336. https://doi.org/10.1080/09540260500238371

Lowe, G., & Chan, B. (2010). Using common work environment metrics to improve performance in healthcare organizations. *Healthcare Papers*, *10*(3), 43–47. https://doi.org/10.12927/hcpap.2010.21863

Luchman, J. N., & González-Morales, M. G. (2013). Demands, control, and support: A meta-analytic review of work characteristics interrelationships. *Journal of Occupational Health Psychology*, *18*(1), 37–52. https://doi.org/10.1037/a0030541

Luthans, F., Avey, J. B., Avolio, B. J., Norman, S. M., & Combs, G. M. (2006). Psychological capital development: toward a micro-intervention. *Journal of Organizational Behavior*, *27*(3), 387–393. https://doi.org/10.1002/job.373

Luthy, C., Perrier, A., Perrin, E., Cedraschi, C., & Allaz, A. F. (2004). Exploring the major difficulties perceived by residents in training: A pilot study. *Swiss Medical Weekly*, *134*(41–42), 612–617. https://doi.org/2004/41/smw-10795

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86–92. https://doi.org/10.1027/1614-1881.1.3.86

Maben, J., Peccei, R., Adams, M., & Robert, G. (2012). *Exploring the relationship between patients' experiences of care and the influence of staff motivation, affect and wellbeing*.

Mache, S., Danzer, G., Klapp, B. F., & Groneberg, D. A. (2013). Surgeons' work ability and performance in surgical care: relations between organisational predictors, work engagement and work ability. *Langenbeck's Archives of Surgery*, *398*(2), 317–325. https://doi.org/10.1007/s00423-012-1044-3

Mache, S., Vitzthum, K., Klapp, B. F., & Groneberg, D. A. (2012). Improving quality of medical treatment and care: Are surgeons' working conditions and job satisfaction associated to patient satisfaction? *Langenbeck's Archives of Surgery*, *397*(6), 973–982. https://doi.org/10.1007/s00423-012-0963-3

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128.

Madsen, F., Ladelund, S., & Linneberg, A. (2014). High levels of bed occupancy associated with increased inpatient and thirty-day hospital mortality in Denmark. *Health Affairs*, *33*(7), 1236–1244. https://doi.org/10.1377/hlthaff.2013.1303

Magnavita, N., Fileni, A., Magnavita, G., Mammi, F., Mirk, P., Roccia, K., & Bergamaschi, A. (2008). Work stress in radiologists. A pilot study. *La Radiologia Medica*, *113*(3), 329–346. https://doi.org/10.1007/s11547-008-0259-4

Manisera, M., Van der Kooij, A. J., & Dusseldorp, E. (2010). Identifying the component structure of satisfaction scales by nonlinear principal components analysis. *Quality Technology & Quantitative Management*, *7*(2), 97–115.

Mann, S. (2005). A healthcare model of emotional labour: An evaluation of the literature and development of a model. *Journal of Health Organization and Management*, *19*(4/5), 304–317. https://doi.org/10.1108/14777260510615369

Marcelino, G., Cerveira, J. M., Carvalho, I., Costa, J. A., Lopes, M., Calado, N. E., & Marques-Vidal, P. (2012). Burnout levels among Portuguese family doctors: A nationwide survey. *BMJ Open*, *2*(3), e001050. https://doi.org/10.1136/bmjopen-2012-001050

Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, *42*(5), 851–858. https://doi.org/10.1016/j.paid.2006.09.023

Markwell, A. L., & Wainer, Z. (2009). The health and wellbeing of junior doctors: Insights from a national survey. *Medical Journal of Australia*, *191*(8), 441–444. https://doi.org/mar11478_fm [pii]

Martin, G. P., & Learmonth, M. (2012). A critical account of the rise and spread of "leadership": The case of UK healthcare. *Social Science and Medicine*, *74*(3), 281–288. https://doi.org/10.1016/j.socscimed.2010.12.002

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*(1), 397–422. https://doi.org/10.1146/annurev.psych.52.1.397

McCray, L. W., Cronholm, P. F., Bogner, H. R., Gallo, J. J., & Neill, R. A. (2008). Resident physician burnout: Is there hope? *Family Medicine*, *40*(9), 626–632.

McGinnis, J. M., Williams-Russo, P., & Knickman, J. R. (2002). The case for more active policy attention to health promotion. *Health Affairs*, *21*(2), 78–93. https://doi.org/10.1377/hlthaff.21.2.78

McGowan, Y., Humphries, N., Burke, H., Conry, M., & Morgan, K. (2013). Through doctors' eyes: A qualitative study of hospital doctor perspectives on their working conditions. *British Journal of Health Psychology*, *18*(4), 874–891. https://doi.org/10.1111/bjhp.12037

McKay, K. A., & Narasimhan, S. (2012). Bridging the gap between doctors and policymakers. *Journal of Nursing Education and Practice*, *2*(4), 52–55. https://doi.org/10.5430/jnep.v2n4p52

McKinstry, B., Walker, J., Porter, M., Fulton, C., Tait, A., Hanley, J., & Mercer, S. W. (2007). The impact of general practitioner morale on patient satisfaction with care: A cross-sectional study. *BMC Family Practice*, *8*(57). https://doi.org/10.1186/1471-2296-8-57

McManus, I. C., Keeling, A., & Paice, E. (2004). Stress, burnout and doctors' attitudes to work are determined by personality and learning style: A twelve year longitudinal study of UK medical graduates. *BMC Medicine*, *2*, 12–29. https://doi.org/10.1186/1741-7015-2-29

McNeish, D. (2017). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 1–17. https://doi.org/10.1080/10705511.2017.1280797

Medical Research Council. (2014). *Maximising the value of UK population cohorts: MRC strategic review of the largest UK population cohort studies*. London, UK: MRC.

Medical Schools Council. (2014). *The consensus statement on the role of the doctor*. London, UK: Medical Schools Council. Retrieved from http://www.medschools.ac.uk/Publications/Pages/Role-of-the-doctor-consensus-statement-2014.aspx

Mehta, S. J. (2015). Patient satisfaction reporting and its implications for patient care. *The AMA Journal of Ethic*, *17*(7), 616–621. https://doi.org/10.1001/journalofethics.2015.17.7.ecas3-1507

Meijman, T. F., & Mulder, G. (1998). Psychological aspects of workload. In P. J. D. Drenth, H. Thierry, & C. J. De Wolff (Eds.), *Handbook of Work and Organizational Psychology* (pp. 5–33). Hove, UK: Psychology Press. https://doi.org/10.2307/2392800

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, *3*(1), 45–58. https://doi.org/10.18148/srm/2009.v3i1.666

Meulman, J. J., Van der Kooij, A. J., & Heiser, W. J. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, 49–70. https://doi.org/10.4135/9781412986311

Michie, S., & Williams, S. (2003). Reducing work related psychological ill health and sickness absence: A systematic literature review. *Occupational and Environmental Medicine*, *60*, 3–9.

Michtalik, H. J., Pronovost, P., Marsteller, J., Spetz, J., & Brotman, D. (2013). Developing a model for attending physician workload and outcomes. *JAMA Internal Medicine*, *173*(11), 1026–1028. https://doi.org/10.1001/jamainternmed.2013.405

Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences*, *42*(5), 869–874. https://doi.org/10.1016/j.paid.2006.09.022

Montgomery, A., Panagopoulou, E., Kehoe, I., & Valkanos, E. (2011). Connecting organisational culture and quality of care in the hospital: Is job burnout the missing link? *Journal of Health Organization and Management*, *25*(1), 108–123. https://doi.org/10.1108/14777261111116851

Montgomery, A., Spânu, F., Băban, A., & Panagopoulou, E. (2015). Job demands, burnout, and engagement among nurses: A multi-level analysis of ORCAB data investigating the moderating effect of teamwork. *Burnout Research*, *2*(2–3), 71–79. https://doi.org/10.1016/j.burn.2015.06.001

Moonesinghe, S. R., Lowery, J., Shahi, N., Millen, A., & Beard, J. D. (2011). Impact of reduction in working hours for doctors in training on postgraduate medical education and patients' outcomes: Systematic review. *BMJ (Clinical Research Education)*, *342*(January), d1580. https://doi.org/10.1136/bmj.d1580

Moore, C., Wisnivesky, J., Williams, S., & McGinn, T. (2003). Medical errors related to discontinuity of care from an inpatient to an outpatient setting. *Journal of General Internal Medicine*, *18*(8), 646–651. https://doi.org/10.1046/j.1525-1497.2003.20722.x

Morrow, V., Boddy, J., & Lamb, R. (2014). *The ethics of secondary data analysis* (NOVELLA working paper: Narrative Research in Action). London, UK. Retrieved from http://eprints.ncrm.ac.uk/3301/

Mulaik, S. A. (1990). Blurring the distinctions between components analysis and common factor analysis. *Multivariate Behavioral Research*, *25*(1), 53–59. https://doi.org/10.1207/s15327906mbr2501

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling : Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidia, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Press.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (Unpublished technical report). Retrieved from https://www.statmodel.com/download/Article_075.pdf

Muthén, B. O., & Kaplan, D. (1992). A comparison of some methodologies for the factor-analysis of nonnormal likert variables - A note on the size of the model. *British Journal of Mathematical & Statistical Psychology*, *45*, 19–30. https://doi.org/10.1111/j.2044-8317.1992.tb00975.x

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 216–316). Washington, DC: American Sociological Association.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide (1998-2017)* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Nahrgang, J. D., Morgeson, F. P., & Hofmann, D. A. (2011). Safety at work: A meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. *Journal of Applied Psychology*, *96*(1), 71–94. https://doi.org/10.1037/a0021484

Naruse, T., Taguchi, A., Kuwahara, Y., Nagata, S., Watai, I., & Murashima, S. (2012). Relationship between perceived time pressure during visits and burnout among home visiting nurses in Japan. *Japan Journal of*

Nursing Science, *9*(2), 185–194. https://doi.org/10.1111/j.1742-7924.2011.00201.x

Naveh, E., Katz-Navon, T., & Stern, Z. (2015). Resident physicians' clinical training and error rate: the roles of autonomy, consultation, and familiarity with the literature. *Advances in Health Sciences Education*, (20), 59–71. https://doi.org/10.1007/s10459-014-9508-6

Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology*, *91*(4), 946–953. https://doi.org/10.1037/0021-9010.91.4.946

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *The New England Journal of Medicine*, *346*(22), 1715–1722. https://doi.org/10.1056/NEJMsa012247

Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass*, *2*(2), 842–860. https://doi.org/10.1111/j.1751-9004.2007.00059.x

NHS Digital. (2017). Summary hospital-level mortality indicator. Retrieved April 5, 2017, from http://content.digital.nhs.uk/SHMI

NHS Employers. (2014). *Guidance on Prevention and Management of Stress At Work*. London, UK: NHS Employers.

NHS England. (n.d.). NHS Staff Survey in England. Retrieved August 8, 2016, from https://www.england.nhs.uk/statistics/statistical-work-areas/nhs-staff-survey-in-england/

NHS England. (2014). *High quality datasets: Our vision*. Leeds, UK.

NHS England. (2015a). *Average daily number of available and occupied beds open overnight by sector - October to December 2014*. Leeds, UK: NHS England. Retrieved from https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/

NHS England. (2015b). *NHS England Framework Agreement for Staff Survey Services Invitation to Tender: Document 3 – Statement of Requirements*.

NHS England. (2015c). *NHS inpatient elective admission events and outpatient referrals and attendances*. Leeds, UK: NHS England. Retrieved from https://www.england.nhs.uk/statistics/statistical-work-areas/hospital-activity/quarterly-hospital-activity/qar-data/

NHS England. (2016). Staff friends and family test data. Retrieved November 22, 2016, from https://www.england.nhs.uk/ourwork/pe/fft/staff-fft/data/

NHS Litigation Authority. (2012). *An Organisation-wide Document for the Management of Stress*. London, UK: NHS Litigation Authority.

NHS Litigation Authority. (2013). *NHS LA Risk Management Standards 2013-14*. London, UK: NHS Litigation Authority.

NHS National Reporting and Learning System. (2015). *Organisation Patient Safety Incident workbook*. Leeds, UK.

NHS Staff Survey Co-ordination Centre. (2014a). *Guidance for the NHS Staff Survey 2014*. Oxford, UK: Picker Institute Europe.

NHS Staff Survey Co-ordination Centre. (2014b). *Making sense of your staff survey data 2013*. Oxford. Retrieved from http://www.nhsstaffsurveys.com/Caches/Files/ST13. MAKING SENSE OF YOUR STAFF SURVEY DATA_FINAL.pdf

NHS Staff Survey Co-ordination Centre. (2015). *Making sense of your staff survey data 2014*. Oxford, UK: Picker Institute Europe. Retrieved from http://www.nhsstaffsurveys.com/Caches/Files/ST14. MAKING SENSE OF YOUR STAFF SURVEY DATA_V1.3.pdf

Nielsen, K., & Miraglia, M. (2017). What works for whom in which circumstances? On the need to move beyond the "what works?" question in organizational intervention research. *Human Relations*, *70*(1), 40–62. https://doi.org/10.1177/0018726716670226

Nielsen, K., Nielsen, M. B., Ogbonnaya, C., Känsälä, M., Saari, E., & Isaksson, K. (2017). Workplace resources to improve both employee well-being and performance: A systematic review and meta-analysis. *Work & Stress*, *31*(2), 101–120.

https://doi.org/10.1080/02678373.2017.1304463

Nielsen, M. B., Matthiesen, S. B., & Einarsen, S. (2010). The impact of methodological moderators on prevalence rates of workplace bullying. A meta-analysis. *Journal of Occupational and Organizational Psychology*, *83*(4), 955–979. https://doi.org/10.1348/096317909X481256

Niks, I. M. W., Gevers, J. M. P., De Jonge, J., & Houtman, I. L. D. (2016). The relation between off-job recovery and job resources: Person-level differences and day-level dynamics. *European Journal of Work and Organizational Psychology*, *25*(2), 226–238. https://doi.org/10.1080/1359432X.2015.1042459

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*(1), 53. https://doi.org/10.1186/s13012-015-0242-0

NIOSH. (2017). Total Worker Health. Retrieved August 30, 2017, from https://www.cdc.gov/niosh/twh/totalhealth.html

Nuffield Council on Bioethics. (2015). *The collection, linking and use of data in biomedical research and health care: Ethical issues*. London, UK: Nuffield.

Nunnally, J. C. (1978). *Psychometric Theory (2nd ed.)*. New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill. https://doi.org/10.1037/018882

Oliver, S., Peersman, G., Harden, A., & Oakley, A. (1999). Discrepancies in findings from effectiveness reviews: The case of health promotion for older people in accident and injury prevention. *Health Education Journal*, *58*(1), 66–77. https://doi.org/10.1177/001789699905800108

Orman, S. A., & Thornton, V. J. (2010). Analysis of junior doctor supervision in Australasian emergency departments. *Emergency Medicine Australasia*, *22*(4), 301–309. https://doi.org/10.1111/j.1742-6723.2010.01300.x

Orton, P., Orton, C., & Pereira Gray, D. (2012). Depersonalised doctors: A cross-sectional study of 564 doctors, 760 consultations and 1876 patient reports in UK general practice. *BMJ Open*, *2*(1), e000274. https://doi.org/10.1136/bmjopen-2011-000274

Oxtoby, K. (2015). Doctors need to resist "presenteeism". *BMJ: British Medical Journal*, *351*(8037), 12–13. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=111742792&site=ehost-live

Ožvačić Adžić, Z., Katić, M., Kern, J., Soler, J. K., Cerovečki, V., & Polašek, O. (2013). Is burnout in family physicians in Croatia related to interpersonal quality of care? *Arhiv Za Higijenu Rada I Toksikologiju*, *64*, 69–78. https://doi.org/10.2478/10004-1254-64-2013-2307

Panari, C., Guglielmi, D., Simbula, S., & Depolo, M. (2010). Can an opportunity to learn at work reduce stress?: A revisitation of the job demand-control model. *Journal of Workplace Learning*, *22*(3), 166–179. https://doi.org/10.1108/13665621011028611

Parkes, K. R., & Sparkes, T. J. (1998). *Organizational interventions to reduce work stress: Are they effective?* Norwich, UK: Health and Safety Executive.

Peadon, E., Caldwell, P. H., & Oldmeadow, W. (2010). "I enjoy teaching but...": Paediatricians' attitudes to teaching medical students and junior doctors. *Journal of Paediatrics and Child Health*, *46*(11), 647–652. https://doi.org/10.1111/j.1440-1754.2010.01823.x

Peeters, M. A. G., & Rutte, C. G. (2005). Time management behavior as a moderator for the job demand-control interaction. *Journal of Occupational Health Psychology*, *10*(1), 64–75. https://doi.org/10.1037/1076-8998.10.1.64

Peeters, M. C. W., & Le Blanc, P. M. (2001). Towards a match between job demands and sources of social support: A study among oncology care providers. *European Journal of Work and Organizational Psychology*, *10*(1), 53–72. https://doi.org/10.1080/13594320042000034

Pellico, L. H., Brewer, C. S., & Kovner, C. T. (2009). What newly licensed registered nurses have to say about their first experiences. *Nursing Outlook*, *57*(4), 194–203. https://doi.org/10.1016/j.outlook.2008.09.008

Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European*

Health Psychologist, *16*(2), 56–69.

Petrescu, M. (2013). Marketing research using single-item indicators in structural equation models. *Journal of Marketing Analytics*, *1*(7), 99–117. https://doi.org/10.1057/jma.2013.7

Petrou, P., Demerouti, E., Peeters, M. C. W., Schaufeli, W. B., & Hetland, J. (2012). Crafting a job on a daily basis: Contextual correlates and the link to work engagement. *Journal of Organizational Behavior*, *33*(8), 1120–1141. https://doi.org/10.1002/job.1783

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford, UK: Blackwell Publishing. https://doi.org/10.1027/1016-9040.11.3.244

Pettker, C. M., Thung, S. F., Norwitz, E. R., Buhimschi, C. S., Raab, C. A., Copel, J. A., … Funai, E. F. (2009). Impact of a comprehensive patient safety strategy on obstetric adverse events. *American Journal of Obstetrics and Gynecology*, *200*(5), 492.e1-492.e8. https://doi.org/10.1016/j.ajog.2009.01.022

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556. https://doi.org/10.3102/00346543074004525

Phipps, D. L., Malley, C., & Ashcroft, D. M. (2012). Job characteristics and safety climate: The role of effort-reward and demand-control-support models. *Journal of Occupational Health Psychology*, *17*(3), 279–289. https://doi.org/10.1037/a0028675

Picker Institute Europe. (2015). *National Health Service National Staff Survey, 2014. SN: 7776*. Essex, UK: UK Data Service. Retrieved from http://dx.doi.org/10.5255/UKDA-SN-7776-1

Pinder, R. J., Greaves, F. E., Aylin, P. P., Jarman, B., & Bottle, A. (2013). Staff perceptions of quality of care: An observational study of the NHS Staff Survey in hospitals in England. *BMJ Quality & Safety*, *22*(7), 563–570. https://doi.org/10.1136/bmjqs-2012-001540

Pines, J. M., Iyer, S., Disbot, M., Hollander, J. E., Shofer, F. S., & Datner, E. M. (2008). The effect of emergency department crowding on patient satisfaction for admitted patients. *Academic Emergency Medicine*, *15*(9), 825–831. https://doi.org/10.1111/j.1553-2712.2008.00200.x

Pituch, K., Stapleton, L., & Kang, J. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*. Retrieved from http://www.tandfonline.com/doi/full/10.1207/s15327906mbr4103_5

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Pottier, P., Dejoie, T., Hardouin, J. B., Le Loupp, A. G., Planchon, B., Bonnaud, A., & Leblanc, V. R. (2013). Effect of stress on clinical reasoning during simulated ambulatory consultations. *Medical Teacher*, *35*(6), 472–480. https://doi.org/10.3109/0142159X.2013.774336

Powell, M., Dawson, J. F., Topakas, A., Durose, J., & Fewtrell, C. (2014). Staff satisfaction and organisational performance: Evidence from a longitudinal secondary analysis of the NHS staff survey and outcome data. *Health Services and Delivery Research*, *2*(50), 1–306. https://doi.org/10.3310/hsdr02500

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*(2), 93–115. https://doi.org/10.1037/a0022658

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(2), 161–182. https://doi.org/10.1080/10705511.2011.557329

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

Prieto, L. L., Soria, M. S., Martínez, I. M., & Schaufeli, W. B. (2008). Extension of the Job Demands-Resources model in the prediction of burnout and engagement among teachers over time. *Psicothema*, *20*(3), 354–360.

Prins, J. T., Gazendam-Donofrio, S. M., Dillingh, G. S., Van De Wiel, H. B. M., van der Heijden, F.

M. M. A., & Hoekstra-Weebers, J. E. H. M. (2008). The relationship between reciprocity and burnout in Dutch medical residents. *Medical Education*, *42*(7), 721–728. https://doi.org/10.1111/j.1365-2923.2008.03041.x

Prins, J. T., Gazendam-Donofrio, S. M., Tubben, B. J., van der Heijden, F. M. M. A., van de Wiel, H. B. M., & Hoekstra-Weebers, J. E. H. M. (2007). Burnout in medical residents: A review. *Medical Education*, *41*(8), 788–800. https://doi.org/10.1111/j.1365-2923.2007.02797.x

Prins, J. T., Hoekstra-Weebers, J. E. H. M., Gazendam-Donofrio, S. M., Dillingh, G. S., Bakker, A. B., Huisman, M., … van der Heijden, F. M. M. A. (2010). Burnout and engagement among resident doctors in the Netherlands: A national study. *Medical Education*, *44*(3), 236–247. https://doi.org/10.1111/j.1365-2923.2009.03590.x

Prins, J. T., Hoekstra-Weebers, J. E. H. M., Gazendam-Donofrio, S. M., Van De Wiel, H. B. M., Sprangers, F., Jaspers, F. C. A., & van der Heijden, F. M. M. A. (2007). The role of social support in burnout among Dutch medical residents. *Psychology, Health & Medicine*, *12*(1), 1–6. https://doi.org/10.1080/13548500600782214

Prins, J. T., van der Heijden, F. M. M. A., Hoekstra-Weebers, J. E. H. M., Bakker, A. B., van de Wiel, H. B. M., Jacobs, B., & Gazendam-Donofrio, S. M. (2009). Burnout, engagement and resident physicians' self-reported errors. *Psychology, Health & Medicine*, *14*(6), 654–666. https://doi.org/10.1080/13548500903311554

Probst, T. M., & Estrada, A. X. (2010). Accident under-reporting among employees: Testing the moderating influence of psychological safety climate and supervisor enforcement of safety practices. *Accident Analysis & Prevention*, *42*(5), 1438–1444. https://doi.org/10.1016/j.aap.2009.06.027

Pym, H. (2016, September 24). Junior doctors call off five-day strikes over contracts. *BBC News Online*. Retrieved from http://www.bbc.co.uk/news/health-37463929

Quine, L. (2002). Workplace bullying in junior doctors: Questionnaire survey. *BMJ*, *324*(7342), 878–879. https://doi.org/10.1136/bmj.324.7342.878

Rabatin, J., Williams, E., Baier Manwell, L., Schwartz, M. D., Brown, R. L., & Linzer, M. (2015). Predictors and outcomes of burnout in primary care physicians. *Journal of Primary Care & Community Health*, 5–7. https://doi.org/10.1177/2150131915607799

Raleigh, V. S., Hussey, D., Seccombe, I., & Qi, R. (2009). Do associations between staff and inpatient feedback have the potential for improving patient experience? An analysis of surveys in NHS acute trusts in England. *Quality & Safety in Health Care*, *18*(5), 347–354. https://doi.org/10.1136/qshc.2008.028910

Randall, R., & Nielsen, K. (2010). Interventions to promote well-being at work. In S. Leka & J. Houdmont (Eds.), *Occupational Health Psychology* (pp. 88–123). Chichester, UK: Wiley- Blackwell.

Ratanawongsa, N., Roter, D., Beach, M. C., Laird, S. L., Larson, S. M., Carson, K. A., & Cooper, L. A. (2008). Physician burnout and patient-physician communication during primary care encounters. *Journal of General Internal Medicine*, *23*(10), 1581–1588. https://doi.org/10.1007/s11606-008-0702-1

Raykov, T., & Marcoulides, G. A. (2000). A method for comparing completely standardized solutions in multiple groups. *Structural Equation Modeling*, *7*(2), 292–308. https://doi.org/10.1207/S15328007SEM0702_9

Reed, D. A., Beckman, T. J., Wright, S. M., Levine, R. B., Kern, D. E., & Cook, D. A. (2008). Predictive validity evidence for medical education research study quality instrument scores: Quality of submissions to JGIM's medical education special issue. *Journal of General Internal Medicine*, *23*(7), 903–907. https://doi.org/10.1007/s11606-008-0664-3

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA*, *298*(9), 1002–1009. https://doi.org/10.1001/jama.298.9.1002

Reed, D. A., Fletcher, K. E., & Arora, V. M. (2010). Systematic review: Association of shift length, protected sleep time, and night float with patient care, residents' health, and education. *Annals of Internal Medicine*, *153*(12), 829–842.

https://doi.org/10.1059/0003-4819-153-12-201012210-00010

Regehr, C., Glancy, D., Pitts, A., & LeBlanc, V. R. (2014). Interventions to reduce the consequences of stress in physicians: A review and meta-analysis. *The Journal of Nervous and Mental Disease*, *202*(5), 353–359. https://doi.org/10.1097/NMD.0000000000000130

Revelle, W., & Zinbarg, R. (2009). Coefficients alpha, beta, omega and GLB: Comments on sijtsma. *Psychometrika*, *74*(1), 145–154.

Ripp, J., Babyatsky, M., Fallar, R., Bazari, H., Bellini, L., Kapadia, C., … Korenstein, D. (2011). The incidence and predictors of job burnout in first-year internal medicine residents: A five-institution study. *Academic Medicine: Journal of the Association of American Medical Colleges*, *86*(10), 1304–1310. https://doi.org/10.1097/ACM.0b013e31822c1236

Roberts, N. (2016, November 9). One in 10 doctors referred to burnout clinic as NHS declared "occupational health hazard." *GP Online*. Retrieved from http://www.gponline.com/one-10-doctors-referred-burnout-clinic-nhs-declared-occupational-health-hazard/article/1415121

Rodham, K., & Bell, J. (2002). Work stress: An exploratory study of the practices and perceptions of female junior healthcare managers. *Journal of Nursing Management*, *10*(1), 5–11. https://doi.org/10.1046/j.0966-0429.2001.00263.x

Rodriguez-Jareno, M. C., Demou, E., Vargas-Prada, S., Sanati, K. A., Kerjanc, A., Reis, P. G., … Serra, C. (2014). European Working Time Directive and doctors' health: A systematic review of the available epidemiological evidence. *BMJ Open*, *4*(7), e004916. https://doi.org/10.1136/bmjopen-2014-004916

Rojon, C., McDowall, A., & Saunders, M. N. K. (2011). On the experience of conducting a systematic review in industrial, work, and organizational psychology: Yes, it is worthwhile. *Journal of Personnel Psychology*, *10*(3), 133–138. https://doi.org/10.1027/1866-5888/a000041

Rosenman, K. D., Kalush, A., Reilly, M. J., Gardiner, J. C., Reeves, M., & Luo, Z. (2006). How much work-related injury and illness is missed by the current national surveillance system? *Journal of Occupational and Environmental Medicine*, *48*(4), 357–365. https://doi.org/10.1097/01.jom.0000205864.81970.63

Ross, S., Bond, C., Rothnie, H., Thomas, S., & Macleod, M. J. (2009). What is the scale of prescribing errors committed by junior doctors? A systematic review. *British Journal of Clinical Pharmacology*, *67*(6), 629–640. https://doi.org/10.1111/j.1365-2125.2008.03330.x

Royal College of Physicians. (2013). *The medical registrar: Empowering the unsung heroes of patient care*. London, UK: Royal College of Physicians. Retrieved from https://www.rcplondon.ac.uk/file/1793/download?token=8fbmTetN

Royal College of Physicians. (2015a). *The NHS: The doctors' view. The RCP's priorities for politicians*. London, UK: Royal College of Physicians. Retrieved from https://www.rcplondon.ac.uk/guidelines-policy/nhs-doctors-view

Royal College of Physicians. (2015b). *Work and wellbeing in the NHS: Why staff health matters to patient care*. London, UK. Retrieved from https://www.rcplondon.ac.uk/guidelines-policy/work-and-wellbeing-nhs-why-staff-health-matters-patient-care

Royal College of Physicians. (2016). *Underfunded. Underdoctored. Overstretched.* London, UK: Royal College of Physicians. Retrieved from https://www.rcplondon.ac.uk/guidelines-policy/underfunded-underdoctored-overstretched-nhs-2016

Royal College of Surgeons. (2010). *General paediatric surgery: Survey of service provision in district general gospitals in England*. London, UK.

Rush, B., Shiell, A., & Hawe, P. (2004). A census of economic evaluations in health promotion. *Health Education Research*, *19*(6), 707–719. https://doi.org/10.1093/her/cyg092

Russell, V., Wyness, L. a, McAuliffe, E., & Fellenz, M. (2010). The social identity of hospital consultants as managers. *Journal of Health Organization and Management*, *24*(3), 220–236. https://doi.org/10.1108/14777261011054581

Ryan, C., Ross, S., Davey, P., Duncan, E. M., Francis,

J. J., Fielding, S., … Bond, C. (2014). Prevalence and causes of prescribing errors: The PRescribing Outcomes for Trainee Doctors Engaged in Clinical Training (PROTECT) study. *PLoS ONE*, *9*(1), 1–9. https://doi.org/10.1371/journal.pone.0079802

Ryu, E. (2015). The role of centering for interaction of level 1 variables in multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 617–630. https://doi.org/10.1080/10705511.2014.936491

Salanova, M., Agut, S., & Peiró, J. M. (2005). Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate. *Journal of Applied Psychology*, *90*(6), 1217–1227. https://doi.org/10.1037/0021-9010.90.6.1217

Salisbury, C., Wallace, M., & Montgomery, A. A. (2010). Patients' experience and satisfaction in primary care: Secondary analysis using multilevel modelling. *BMJ (Clinical Research Ed.)*, *341*(oct12_1), c5004. https://doi.org/10.1136/bmj.c5004

Salyers, M. P., Bonfils, K. A., Luther, L., Firmin, R. L., White, D. A., Adams, E. L., & Rollins, A. L. (2016). The relationship between professional burnout and quality and safety in healthcare: A meta-analysis. *Journal of General Internal Medicine*. https://doi.org/10.1007/s11606-016-3886-9

Sargent, M., Sotile, W., & Sotile, M. (2004). Stress and coping among orthopaedic surgery residents and faculty. *Journal Bone & Joint Surgery*, *86*, 1579–1586.

Sari, A. B.-A., Sheldon, T. A., Cracknell, A., Turnbull, A., Dobson, Y., Grant, C., … Richardson, A. (2007). Extent, nature and consequences of adverse events: Results of a retrospective casenote review in a large NHS hospital. *Quality & Safety in Health Care*, *16*(6), 434–439. https://doi.org/10.1136/qshc.2006.021154

Schaufeli, W. B. (2015). Engaging leadership in the job demands-resources model. *Career Development International*, *20*(5), 446–463. https://doi.org/10.1108/CDI-02-2015-0025

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, *25*(3), 293–315. https://doi.org/10.1002/job.248

Schaufeli, W. B., Bakker, A. B., van der Heijden, F. M. M. A., & Prins, J. T. (2009). Workaholism among medical residents: It is the combination of working excessively and compulsively that counts. *International Journal of Stress Management*, *16*(4), 249–272. https://doi.org/10.1037/a0017537

Schaufeli, W. B., Bakker, A. B., & Van Rhenen, W. (2009). How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. *Journal of Organizational Behavior*, *30*(7), 893–917. https://doi.org/10.1002/job.595

Schaufeli, W. B., & Taris, T. W. (2014). A critical review of the job demands-resources model: Implications for improving work and health. In G. F. Bauer & O. Hämmig (Eds.), *Bridging occupational, organizational and public health: A transdisciplinary approach* (pp. 43–68). Dordrecht, Netherlands: Springer Science. https://doi.org/10.1007/978-94-007-5640-3

Scheepers, R. A., Boerebach, B. C. M., Arah, O. A., Heineman, M. J., & Lombarts, K. M. J. M. H. (2015). A systematic review of the impact of physicians' occupational well-being on the quality of patient care. *International Journal of Behavioral Medicine*, *22*(6), 683–698. https://doi.org/10.1007/s12529-015-9473-3

Scheurer, D., McKean, S., Miller, J., & Wetterneck, T. (2009). U.S. physician satisfaction: A systematic review. *Journal of Hospital Medicine*, *4*(9), 560–568. https://doi.org/10.1002/jhm.496

Schilling, P. L., Campbell, D. A., Englesbe, M. J., & Davis, M. M. (2010). A comparison of in-hospital mortality risk conferred by high hospital occupancy, differences in nurse staffing levels, weekend admission, and seasonal influenza. *Medical Care*, *48*(3), 224–232. https://doi.org/10.1097/MLR.0b013e3181c162c0

Schlomer, B. J., & Copp, H. L. (2014). Secondary data analysis of large data sets in urology: Successes and errors to avoid. *The Journal of Urology*, *191*(3), 587–596. https://doi.org/10.1016/j.juro.2013.09.091

Sebo, P., Gallacchi, M. B., Goehring, C., Künzi, B., &

Bovier, P. A. (2007). Use of tobacco and alcohol by Swiss primary care physicians: A cross-sectional survey. *BMC Public Health*, *7*, 5. https://doi.org/10.1186/1471-2458-7-5

Selič, P., Stegne-Ignjatović, T., & Klemenc-Ketiš, Z. (2012). Burnout among Slovenian family medicine trainees: A cross-sectional study. *Zdravniski Vestnik*, *81*(3), 218–224.

Selig, J. ., & Preacher, K. J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects. Retrieved May 17, 2017, from http://www.quantpsy.org/medmc/medmc.htm

Selye, H. (1956). *Stress of life*. New York, NY: McGraw-Hill.

Shackelton, R., Siegrist, J., Link, C., Marceau, L., von dem Knesebeck, O., & McKinlay, J. (2010). Work stress of primary care physicians in the US, UK and German health care systems. *Social Science and Medicine*, *71*(2), 298–304. https://doi.org/10.1016/j.socscimed.2010.03.043

Shanafelt, T. D., Balch, C. M., Bechamps, G. J., Russell, T., Dyrbye, L. N., Satele, D., … Freischlag, J. A. (2010). Burnout and medical errors among American surgeons. *Annals of Surgery*, *251*(6), 995–1000. https://doi.org/10.1097/SLA.0b013e3181bfdab3

Shanafelt, T. D., Balch, C. M., & Dyrbye, L. N. (2012). Special report: Suicide ideation among American surgeons. *JAMA Surgery*, *146*(June), 889–896. https://doi.org/10.1016/j.ysur.2011.12.003

Shanafelt, T. D., Boone, S., Tan, L., Dyrbye, L. N., Sotile, W., Satele, D., … Oreskovich, M. R. (2012). Burnout and satisfaction with work-life balance among US physicians relative to the general US population. *Archives of Internal Medicine*, *172*(18), 1377–1385. https://doi.org/10.1001/archinternmed.2012.3199

Shanafelt, T. D., Bradley, K. a, Wipf, J. E., & Back, A. L. (2002). Burnout and self-reported patient care in an internal medicine residency program. *Annals of Internal Medicine*, *136*(5), 358–367. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11874308

Shanafelt, T. D., Hasan, O., Dyrbye, L. N., Sinsky, C., Satele, D., Sloan, J., & West, C. P. (2015). Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. *Mayo Clinic Proceedings*, *90*(12), 1600–1613. https://doi.org/10.1016/j.mayocp.2015.08.023

Shanafelt, T. D., Sloan, J. A., & Habermann, T. M. (2003). The well-being of physicians. *The American Journal of Medicine*, *114*(6), 513–519. https://doi.org/10.1016/S0002-9343(03)00117-7

Shen, L.-L., Lao, L.-M., Jiang, S.-F., Yang, H., Ren, L.-M., Ying, D. G.-C., & Zhu, S.-Z. (2012). A survey of anxiety and depression symptoms among primary-care physicians in China. *International Journal of Psychiatry in Medicine*, *44*(3), 257–270. https://doi.org/10.2190/PM.44.3.f

Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*, *3*(2), 1–13. https://doi.org/10.3389/fpsyg.2012.00034

Shipton, H., Armstrong, C., West, M., & Dawson, J. F. (2008). The impact of leadership and quality climate on hospital performance. *International Journal for Quality in Health Care*, *20*(6), 439–445. https://doi.org/10.1093/intqhc/mzn037

Shirom, A., Nirel, N., & Vinokur, A. D. (2006). Overload, autonomy, and burnout as predictors of physicians' quality of care. *Journal of Occupational Health Psychology*, *11*(4), 328–342. https://doi.org/10.1037/1076-8998.11.4.328

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445. https://doi.org/10.1037//1082-989x.7.4.422

Sibbald, B., Bojke, C., & Gravelle, H. (2003). National survey of job satisfaction and retirement intentions among general practitioners in England. *BMJ*, *326*(January), 1–4. Retrieved from http://www.bmj.com/content/326/7379/22?variant=full-text

Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of*

Occupational Health Psychology, *1*(1), 27–41. https://doi.org/10.1037/1076-8998.1.1.27

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sjetne, I. S., Veenstra, M., & Stavem, K. (2007). The effect of hospital size and teaching status on patient experiences with hospital care: A multilevel analysis. *Medical Care*, *45*(3), 252–258. https://doi.org/10.1097/01.mlr.0000252162.78915.62

Skakon, J., Nielsen, K., Borg, V., & Guzman, J. (2010). Are leaders' well-being, behaviours and style associated with the affective well-being of their employees? A systematic review of three decades of research. *Work & Stress*, *24*(2), 107–139. https://doi.org/10.1080/02678373.2010.495262

Sloan, D., Jones, S., Evans, H., Chant, L., Williams, S., & Peel, P. (2014). *Implementing NICE public health guidance for the workplace: A national organisational audit of NHS trusts in England*. London, UK: Royal College of Physicians.

Snijders, C., Kollen, B. J., van Lingen, R. A., Fetter, W. P. F., & Molendijk, H. (2009). Which aspects of safety culture predict incident reporting behavior in neonatal intensive care units? A multilevel analysis. *Critical Care Medicine*, *37*(1), 61–67. https://doi.org/10.1097/CCM.0b013e31819300e4

Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. *Encyclopedia of Statistics in Behavioral Science*, *3*, 1570–1573. https://doi.org/10.1002/0470013192.bsa492

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (Vol. 13, pp. 290–312). Washington, DC: American Sociological Association.

Soh, M., Zarola, A., Palaiou, K., & Furnham, A. (2016). Work-related well-being. *Health Psychology Open*, *3*(1), 205510291662838. https://doi.org/10.1177/2055102916628380

Spector, P. E., Dwyer, D. J., & Jex, S. M. (1988). Relation of job stressors to affective, health, and performance outcomes: A comparison of multiple data sources. *Journal of Applied Psychology*, *73*(1), 11–19. https://doi.org/10.1037/0021-9010.73.1.11

Stack, S. (2004). Suicide risk among physicians: A multivariate analysis. *Achives of Suicide Research*, *8*(3), 287–292. https://doi.org/10.1080/13811110490436954

Stansfeld, S. A., & Candy, B. (2006). Psychosocial work environment and mental health- A meta-analytic review. *Scandinavian Journal of Work, Environment and Health*, *32*(6), 443–462. https://doi.org/10.5271/sjweh.1050

Stansfeld, S. A., Kerry, S., Chandola, T., Russell, J., Berney, L., Hounsome, N., … Bhui, K. (2015). Pilot study of a cluster randomised trial of a guided e-learning health promotion intervention for managers based on management standards for the improvement of employee well-being and reduction of sickness absence: GEM Study. *BMJ Open*, *5*(10), e007981. https://doi.org/10.1136/bmjopen-2015-007981

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

Stern, Z., Katz-Navon, T., & Naveh, E. (2008). The influence of situational learning orientation, autonomy, and voice on error making: The case of resident physicians. *Management Science*, *54*(9), 1553–1564. https://doi.org/10.1287/mnsc.1080.0862

Stride, C. B. (2016). *Multilevel Modellig Using Mplus*. Sheffield, UK: Figure It Out.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson Education.

Taris, T. W. (2006). Is there a relationship between burnout and objective performance? A critical review of 16 studies. *Work & Stress*, *20*(4), 316–334. https://doi.org/10.1080/02678370601065893

Taris, T. W., & Schreurs, P. J. G. (2007). How may nonresponse affect findings in organizational surveys? The tendency-to-the-positive effect. *International Journal of Stress Management*, *14*(3), 249–259. https://doi.org/10.1037/1072-5245.14.3.249

Taylor, C., Graham, J., Potts, H. W. W., Richards, M. A., & Ramirez, A. J. (2005). Changes in mental health of UK hospital consultants since the mid-1990s. *The Lancet*, *366*(9487), 742–744. https://doi.org/10.1016/S0140-6736(05)67178-4

Teoh, K. R.-H. (2016). You mean the data has already been collected? Using secondary data in psychology. *PsyPAG Quarterly*, *100*, 48–50. Retrieved from http://www.psypag.co.uk/wp-content/uploads/2016/09/PsyPag-100-web.pdf

Teoh, K. R.-H., Coyne, I., Devonish, D., Leather, P., & Zarola, A. (2016). The interaction between supportive and unsupportive manager behaviors on employee work attitudes. *Personnel Review*, *45*(6), 1386–1402. https://doi.org/10.1108/PR-05-2015-0136

Teoh, K. R.-H., Hassard, J., & Cox, T. (2015). *A systematic review of the impact of doctors' working conditions on quality of care provided*. *PROSPERO International Prospective Register of Systematic Reviews*. York, UK. Retrieved from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015024695

Thomas, N. K. (2004). Resident Burnout. *JAMA*, *292*(23), 2880–2889. https://doi.org/10.1001/jama.292.23.2880

Tims, M., Bakker, A. B., & Derks, D. (2013). The impact of job crafting on job demands, job resources, and well-being. *Journal of Occupational Health Psychology*, *18*(2), 230–240. https://doi.org/10.1037/a0032141

Tofighi, D., & Thoemmes, F. (2014). Single-level and multilevel mediation analysis. *The Journal of Early Adolescence*, *34*(1), 93–119. https://doi.org/10.1177/0272431613511331

Topakas, A., Admasachew, L., & Dawson, J. F. (2010a). *F - Outcomes of staff engagement in the NHS: A trust level analysis*. Birmingham, UK: Aston Business School.

Topakas, A., Admasachew, L., & Dawson, J. F. (2010b). *H - Staff Advocacy and NHS Trust Performance: Advocacy of Treatment as a predictor of trust level outcomes*. Birmingham, UK: Aston Business School.

Topakas, A., Admasachew, L., & Dawson, J. F. (2010c). *J - Employee health and well-being in the NHS: A trust level analysis*. Birmingham, UK: Aston Business School.

Topakas, A., Admasachew, L., & Dawson, J. F. (2011). *K - NHS Staff Survey scores as predictors of trust outcomes: A multi-method longitudinal analysis*. Birmingham, UK: Aston Business School.

Toral-Villanueva, R., Aguilar-Madrid, G., & Juárez-Pérez, C. A. (2009). Burnout and patient care in junior doctors in Mexico City. *Occupational Medicine*, *59*(1), 8–13. https://doi.org/10.1093/occmed/kqn122

Torrente, P., Salanova, M., Llorens, S., & Schaufeli, W. B. (2012). Teams make it work: How team work engagement mediates between social resources and performance in teams. *Psicothema*, *24*(1), 106–112.

Trufelli, D. C., Bensi, C. G., Garcia, J. B., Narahara, J. L., Abrão, M. N., Diniz, R. W., … Del Giglio, A. (2008). Burnout in cancer professionals: A systematic review and meta-analysis. *European Journal of Cancer Care*, *17*(6), 524–531. https://doi.org/10.1111/j.1365-2354.2008.00927.x

Tucker, M. K., Jimmieson, N. L., & Oei, T. P. (2013). The relevance of shared experiences: A multi-level study of collective efficacy as a moderator of job control in the stressor-strain relationship. *Work & Stress*, *27*(1), 1–21. https://doi.org/10.1080/02678373.2013.772356

Tucker, P., Bejerot, E., & Akerstedt, T. (2012). Sleep quality as a mediator in the relationship between doctors' worktime control and patient safety. *Journal of Sleep Research*, *21*(1), 273. https://doi.org/10.1371/journal.pone.0023952.S33

Tucker, P., Bejerot, E., Kecklund, G., Aronsson, G., & Åkerstedt, T. (2015). The impact of work time control on physicians' sleep and well-being. *Applied Ergonomics*, *47*, 109–116. https://doi.org/10.1016/j.apergo.2014.09.001

Van den Eynden, V., Corti, L., Woollard, M., & Bishop, L. (2011). *Managing and sharing data: A best practice guide for researchers*. Colchester, Essex: UK Data Archive. Retrieved from http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

van den Hombergh, P., Künzi, B., Elwyn, G., van Doremalen, J., Akkermans, R., Grol, R., & Wensing, M. (2009). High workload and job stress are associated with lower practice performance in general practice: An observational study in 239 general practices in the Netherlands. *BMC Health Services Research*, *9*, 118. https://doi.org/10.1186/1472-6963-9-118

Van der Doef, M., & Maes, S. (1999). The Job Demand-Control (-Support) Model and psychological well-being: A review of 20 years of empirical research. *Work & Stress*, *13*(2), 87–114. https://doi.org/10.1080/026783799296084

Van der Klink, J. J. L., Blonk, R. W. B., Schene, A. H., & Van Dijk, F. J. H. (2001). The benefits of interventions for work-related stress. *American Journal of Public Health*, *91*(2), 270–276. https://doi.org/10.2105/AJPH.91.2.270

van Emmerik, H. I. J., Bakker, A. B., & Euwema, M. C. (2009). Explaining employees' evaluations of organizational change with the job-demands resources model. *Career Development International*, *14*(6), 594–613. https://doi.org/10.1108/13620430910997312

van Horn, J. E., Taris, T. W., Schaufeli, W. B., & Schreurs, P. J. G. (2004). The structure of occupational well-being: A study among Dutch teachers. *Journal of Occupational and Organizational Psychology*, *77*(3), 365–375. https://doi.org/10.1348/0963179041752718

van Oostveen, C. J., Mathijssen, E., & Vermeulen, H. (2015). Nurse staffing issues are just the tip of the iceberg: A qualitative study about nurses' perceptions of nurse staffing. *International Journal of Nursing Studies*, *52*(8), 1300–1309. https://doi.org/10.1016/j.ijnurstu.2015.04.002

van Vegchel, N., de Jonge, J., Bakker, A. B., & Schaufeli, W. B. (1999). Testing global and specific indicators of rewards in the Effort--Reward Imbalance Model: Does it make any difference? *European Journal of Cognitive Psychology*, *11*(4), 403–421. https://doi.org/10.1080/713752330

Velicer, W., & Jackson, D. (2004). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*, 1–28. https://doi.org/10.1207/s15327906mbr2501

Victoria State Government. (2013). *Junior doctors' contribution to patient safety and quality improvement*. Victoria, Australia: VMIA.

Vincent, C. (2001). Adverse events in British hospitals: Preliminary retrospective record review. *BMJ*, *322*(7285), 517–519. https://doi.org/10.1136/bmj.322.7285.517

Vincent, C. (2010). *Patient safety* (2nd ed.). London, UK: Wiley- Blackwell.

Virtanen, M., Kurvinen, T., Terho, K., Oksanen, T., Peltonen, R., Vahtera, J., … Kivimäki, M. (2009). Work hours, work stress, and collaboration among ward staff in relation to risk of hospital-associated infection among patients. *Medical Care*, *47*(3), 310–318. https://doi.org/10.1097/MLR.0b013e3181893c64

Visser, M. R. M., Smets, E. M. A., Oort, F. J., & de Haes, H. C. J. M. (2003). Stress, satisfaction and burnout among Dutch medical specialists. *Canadian Medical Association Journal*, *168*(3), 271–275.

Vogt, K., Hakanen, J. J., Brauchli, R., Jenny, G. J., & Bauer, G. F. (2016). The consequences of job crafting: A three-wave study. *European Journal of Work and Organizational Psychology*, *25*(3), 353–362. https://doi.org/10.1080/1359432X.2015.1072170

Volpe, F. M., de Miranda Magalhães, A. C., & Rocha, A. R. (2013). High bed occupancy rates: Are they a risk for patients and staff? *International Journal of Evidence-Based Healthcare*, *11*(4), 312–316. https://doi.org/10.1111/1744-1609.12046

von dem Knesebeck, O., Klein, J., Frie, K. G., Blum, K., & Siegrist, J. (2010). Psychosocial stress among hospital doctors in surgical fields: Results of a nationwide survey in Germany. *Deutsches Arzteblatt International*, *107*(14), 248–253. https://doi.org/10.3238/arztebl.2010.0248

Vree, F. E., Cohen, S. L., Chavan, N., & Einarsson, J. I. (2011). The impact of surgeon volume on perioperative outcomes in hysterectomy. *Journal of Minimally Invasive Gynecology*, *18*, 174–181. https://doi.org/10.1016/j.jmig.2011.08.241

Wallace, J. E., Lemaire, J. B., & Ghali, W. A. (2009). Physician wellness: A missing quality indicator. *The Lancet*, *374*(9702), 1714–1721. https://doi.org/10.1016/S0140-6736(09)61424-0

Walters, D., & Wadsworth, E. (2014). Contexts and determinants of the management of occupational safety and health in European workplaces. *Policy and Practice in Health and*

Safety, *12*(2), 109–130. https://doi.org/10.1080/14774003.2014.11667806

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology*, *82*(2), 247–252. https://doi.org/10.1037/0021-9010.82.2.247

Warr, P. (1994). A conceptual framework for the study of work and mental health. *Work & Stress*, *8*(2), 84–97. https://doi.org/10.1080/02678379408259982

Watt, I., Nettleton, S., & Burrows, R. (2008). The views of doctors on their working lives: A qualitative study. *Journal of the Royal Society of Medicine*, *101*(12), 592–597. https://doi.org/10.1258/jrsm.2008.080195

Weigl, M., Hornung, S., Angerer, P., Siegrist, J., & Glaser, J. (2013). The effects of improving hospital physicians working conditions on patient care: A prospective, controlled intervention study. *BMC Health Services Research*, *13*, 401–409. https://doi.org/10.1186/1472-6963-13-401

Weigl, M., Schneider, A., Hoffmann, F., & Angerer, P. (2015). Work stress, burnout, and perceived quality of care: A cross-sectional study among hospital pediatricians. *European Journal of Pediatrics*, *174*(9), 1237–1246. https://doi.org/10.1007/s00431-015-2529-1

Welcome Trust. (2014). *Enhancing discoverability of public health and empidemiology research data* (Publich Health Research Data Forum). London, UK.

Wellens, B. T., & Smith, A. P. (2006). Combined workplace stressors and their relationship with mood, physiology, and performance. *Work & Stress*, *20*(3), 245–258. https://doi.org/10.1080/02678370601022712

Wells, J. (2011). The impact of stress amongst health professionals. *Journal of Mental Health*, *20*(2), 111–114. https://doi.org/10.3109/09638237.2011.556161

Welp, A., Meier, L. L., & Manser, T. (2015). Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in Psychology*, *5*(January), 1–13. https://doi.org/10.3389/fpsyg.2014.01573

Wen, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure. *Psychological Methods*, *20*(2), 193–203. https://doi.org/10.1037/met0000029

West, C. P., Dyrbye, L. N., Erwin, P. J., & Shanafelt, T. D. (2016). Interventions to prevent and reduce physician burnout: A systematic review and meta-analysis. *The Lancet*, *6736*(16). https://doi.org/10.1016/S0140-6736(16)31279-X

West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., & Shanafelt, T. D. (2006). Association of perceived medical errors with resident distress and empathy. *JAMA*, *296*(9), 1071–1078. https://doi.org/10.1001/jama.296.9.1071

West, C. P., Tan, A. D., Habermann, T. M., Sloan, J. A., & Shanafelt, T. D. (2010). Association of resident fatigue and distress with perceived medical errors. *JAMA*, *30*(4), 217. https://doi.org/10.1001/jama.2009.1389

West, C. P., Tan, A. D., & Shanafelt, T. D. (2012). Association of resident fatigue and distress with occupational blood and body fluid exposures and motor vehicle incidents. *Mayo Clinic Proceedings*, *87*(12), 1138–1144. https://doi.org/10.1016/j.mayocp.2012.07.021

West, M., & Dawson, J. F. (2010). *NHS Staff Management and Health Service Quality*. Birmingham, UK: Aston Business School, Aston University.

West, M., Dawson, J. F., Admasachew, L., & Topakas, A. (2011). *NHS staff management and health service quality: Results from the NHS Staff Survey and related data*. Birmingham, UK: Aston Business School.

Westercamp, N., Wang, R. S., & Fassiotto, M. (2017). Resident perspectives on work-life policies and implications for burnout. *Academic Psychiatry*. https://doi.org/10.1007/s40596-017-0757-6

Westman, M., Bakker, A. B., Roziner, I., & Sonnentag, S. (2011). Crossover of job demands and emotional exhaustion within teams: A longitudinal multilevel study. *Anxiety, Stress & Coping*, *24*(5), 561–577. https://doi.org/10.1080/10615806.2011.558191

WHO. (2010). *Healthy workplaces: A model for action*. Geneva, Switzerland.

Williams, E. S., Konrad, T. R., Linzer, M., McMurray, J., Pathman, D. E., Gerrity, M., … Douglas, J. A. (2002). Physician, practice, and patient characteristics related to primary care physician physical and mental health: Results from the Physician Worklife Study. *Health Services Research*, *37*(1), 121–143. https://doi.org/10.1111/1475-6773.00007

Williams, E. S., Manwell, L. B., Konrad, T. R., & Linzer, M. (2007). The relationship of organizational culture, stress, satisfaction, and burnout with physician-reported error and suboptimal patient care: Results from the MEMO study. *Health Care Management Review*, *32*(3), 203–212. https://doi.org/10.1097/01.HMR.0000281626.28363.59

Williams, E. S., & Skinner, A. C. (2003). Outcomes of physician job satisfaction: A narrative review, implications, and directions for future research. *Health Care Management Review*, *28*(2), 119–139.

Winefield, H. R., & Veale, B. M. (2002). Work stress and quality of work performance in Australian general practitioners. *Australian Journal of Primary Health*, *8*(2), 59–65.

Wischmeyer, P. E., Johnson, B. R., Wilson, J. E., Dingmann, C., Bachman, H., Roller, E., … Henthorn, T. K. (2007). A survey of propofol abuse in academic anesthesia programs. *Survey of Anesthesiology*, *105*, 1066–1071. https://doi.org/10.1097/SA.0b013e3181669bc0

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. London, UK: Sage.

Wong, C. A., & Cummings, G. G. (2007). The relationship between nursing leadership and patient outcomes: A systematic review. *Journal of Nursing Management*, *15*(5), 508–521. https://doi.org/10.1111/j.1365-2834.2007.00723.x

Woodrow, C., & Guest, D. E. (2012). Public violence, staff harassment and the wellbeing of nursing staff: An analysis of national survey data. *Health Services Management Research*, *25*(1), 24–30. https://doi.org/10.1258/hsmr.2011.011019

World Health Organization. Preamble to the Constitution of WHO as adopted by the International Health Conference, New York, 19 June - 22 July 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of WHO, no. 2, p. 100) and entered into force on 7 April (1948).

Wrzesniewski, A., & Dutton, J. E. (2001). Crafting a job: As active employees revisioning crafters of their work. *Academy of Management Review*, *26*(2), 179–201. https://doi.org/10.2307/259118

Wu, A. W., Folkard, S., McPhee, S. J., & Lo, B. (2003). Do house officers learn from their mistakes? *Quality and Safety in Health Care*, *12*(3), 221–226. https://doi.org/10.1136/qhc.12.3.221

Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2007). The role of personal resources in the job demands-resources model. *International Journal of Stress Management*, *14*(2), 121–141. https://doi.org/10.1037/1072-5245.14.2.121

Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2009). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational & Organizational Psychology*, *82*(1), 183–200. https://doi.org/10.1348/096317908X285633

Xanthopoulou, D., Bakker, A. B., Dollard, M. F., Demerouti, E., Schaufeli, W. B., Taris, T. W., & Schreurs, P. J. G. (2007). When do job demands particularly predict burnout?: The moderating role of job resources. *Journal of Managerial Psychology*, *22*(8), 766–786. https://doi.org/10.1108/02683940710837714

Yasunaga, H., Nishii, O., Hirai, Y., Ochiai, K., Matsuyama, Y., & Ohe, K. (2009). Impact of surgeon and hospital volumes on short-term postoperative complications after radical hysterectomy for cervical cancer. *Journal of Obstetrics and Gynaecology Research*, *35*(4), 699–705. https://doi.org/10.1111/j.1447-0756.2009.01027.x

Zadow, A., & Dollard, M. F. (2015). Psychosocial safety climate. In *The Wiley Blackwell Handbook of the Psychology of Occupational Safety and Workplace Health* (pp. 414–436). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118979013.ch18

Zantinge, E. M., Verhaak, P. F., de Bakker, D. H., van der Meer, K., & Bensing, J. M. (2009). Does burnout among doctors affect their involvement in patients' mental health problems? A study of videotaped consultations. *BMC Family Practice*, *10*(1), 60. https://doi.org/10.1186/1471-2296-10-60

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, (12), 695–719. https://doi.org/10.5465/AMBPP.2008.33716518

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Zis, P., Anagnostopoulos, F., & Sykioti, P. (2014). Burnout in medical residents: A study based on the Job Demands-Resources Model. *The Scientific World Journal*, 1–10. https://doi.org/10.1155/2014/673279

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident; Analysis and Prevention*, *42*(5), 1517–1522. https://doi.org/10.1016/j.aap.2009.12.019

Zwaan, L. (2012). *Diagnostic reasoning and diagnostic error in medicine*. Vrije Universiteit.

Zwetsloot, G. I. J. M., Kines, P., Wybo, J.-L., Ruotsala, R., Drupsteen, L., & Bezemer, R. A. (2017). Zero Accident Vision based strategies in organisations: Innovative perspectives. *Safety Science*, *91*, 260–268. https://doi.org/10.1016/j.ssci.2016.08.016

Zwetsloot, G. I. J. M., Leka, S., & Kines, P. (2017). Vision Zero: From accident prevention to the promotion of health, safety and well-being at work. *Policy and Practice in Health and Safety*, *0*(0), 1–13. https://doi.org/10.1080/14773996.2017.1308701

Appendix

Appendix I: Extraction Form for Systematic Review

Study Information

Article name	
Article authors	
Reviewer	
Date examined	

Study Background

Country	
Research question	
Study design	
Predictor variable	
Outcome measure	
Mediating variable (if applicable)	
Theoretical framework (e.g. JDCS, ERI, JDR)	

Sample

Sample size	
Recruitment/ sampling method	
Are the participants from more than one site	
Response rate	

Measures

Construct	Measurement method/ instrument name	Self-report	Internal reliability

Findings

Predictor	Dependent	Effect Size	n

Other findings:

Limitations of study

Study quality- MERSQI

Domain	Score	Allocated score	Source
Study design - Single group cross-sectional - Single group pretest and posttest - Non-randomised, two groups - Randomised controlled experiment	 1 1.5 2 3		
Sampling Institutions - Single institution - Two institutions - More than two institutions Response rate - <50% OR not reported - 50-74% - 75-100%	 0.5 1 1.5 0.5 1 1.5		
Type of outcome data - Assessment by study participant - Objective measurement	 1 3		
Validity of psychosocial measure Internal structure (internal consistency, interrater reliability, factor analysis) - Not applicable - Not reported - Reported Content validity	 0 1 		

- Not applicable - Not reported - Reported	 0 1		
Relations to other variables (criterion, concurrent, and predictive validity) - Not applicable - Not reported - Reported	 0 1		
Data analysis Appropriateness of data analysis - Data analysis inappropriate for study design or type of data - Data analysis appropriate for study design or type of data Sophistication of data analysis - Descriptive statistics only (frequencies, measures of central tendency) - Beyond descriptive analysis (comparisons, correlations, relationships between variables)	 0 1 1 2		
Highest outcome level - Satisfaction, attitudes, perception - Knowledge, skills - Behaviours - Patient/ health care outcomes	 1 1.5 2 3		
Total			

Additional Comments

Where there any challenges or difficulties with regards to the filling out this form?

Is there any other feedback regarding the study that you with to include?

Appendix II: Job Demands and Job Resources' Relationship with Clinical Excellence

Author	Main Findings	Relationship Found
Job demands and clinical excellence		
Perceived Workload		
Mache et al. (2013)	Quantitative demands (*r*=-.18) and cognitive demands (*r*=-.31) both negatively correlated with work ability.	Yes
Bertram et al.(1990)	Mental workload negatively associated with quality of care (*r*=-.46).	Yes
Bertram et al. (1992)	Mental workload and mental workload squared curvilinear associated with self-rated performance (*r*=-.67) and observer-rated technical performance (*r*=-.38; *r*=.45).	Yes
Shirom et al. (2006)	Overload negatively associated with quality of care (β=-.15).	Yes
Bernburg et al. (2016)	Quantitative demands (*r*=-.28) negatively correlated with work ability.	Yes
Demanding Patients		
An et al. (2013)	Physicians with a high burden of difficult encounters had a 7.68% lower quality care rate than those with a lower burden. No significant differences for specific error quality of care for diabetes and hypertension.	Mixed
Time Pressure		
Krämer et al. (2016)	Time pressure at positively predicted quality of care one year later (β=-.19).	Yes
Linzer et al. (2009)	Only two out of nine relationships between time pressure and quality of care were significant.	Mixed
Emotional Demands		
Mache et al. (2013)	Emotional demands (*r*=-.21) and demands hiding emotion (*r*=-.19) both negatively correlated with work ability.	Yes
Bernburg et al. (2016)	Emotional demands (*r*=-.20) negatively correlated with work ability.	Yes
Krämer et al. (2016)	Emotional demands did not predict quality of care one year later.	No
Social Conflict		
Krämer et al. (2016)	Social conflict predicted quality of care one year later (β=-.15).	No
Higher-Order Job Demands		
Weigl et al. (2015)	Effort negatively predicted quality of care (β=-.49).	Yes
Loerbroks et al. (2016)	Effort negatively predicted quality of care (β=-.24).	Yes
Mache et al. (2013)	Latent factor of four job demands explained an additional 10% of the variance towards work ability.	Yes
Job resources and clinical excellence		
Autonomy		
Mache et al. (2013)	Degree of freedom (*r*=.32) correlated positively with work ability.	Yes
Bernburg et al. (2016)	Degree of freedom (*r*=.15) positively correlated with work ability.	Yes
Shirom et al. (2006)	Job autonomy positively correlated with quality of care (β=.37).	Yes
Job Control		
Mache et al. (2013)	Influence at work (*r*=.39) correlated positively with work ability.	Yes
Bernburg et al. (2016)	Influence at work (*r*=.15) positively correlated with work ability.	Yes
Linzer et al. (2009)	Work control was significantly associated with diabetes quality of care (β=8.41) but not with hypertension or overall quality of care.	Mixed
Learning and Development		
Mache et al. (2013)	Possibilities for development (*r*=.36) and feedback at work (*r*=.27) both correlated positively with work ability.	Yes

Bernburg et al. (2016)	Possibilities for development (*r*=.14) and feedback at work (*r*=.12) both correlated positively with work ability.	Yes
Social support - Colleagues		
Mache et al. (2013)	Social relationships (*r*=.20) and social support (*r*=.41) both correlated positively with work ability.	Yes
Bernburg et al. (2016)	Social relationships (*r*=.11) and social support (*r*=.15) both correlated positively with work ability.	Yes
Supervisor support		
Mache et al. (2013)	Quality of leadership (*r*=.25) correlated positively with work ability.	Yes
Bernburg et al. (2013)	Quality of leadership (*r*=.09) did not correlate with work ability.	No
Higher-order job resources		
Weigl et al. (2015)	Rewards (perceived salary, promotion prospects, esteem, and job security) was positively correlated with quality of care (β=.44).	Yes
Mache et al. (2013)	Latent factor of eight job resources (influence at work, degree of freedom of work, possibilities for development, quality of leadership, social support, feedback at work, social relations, & sense of community) explained an additional 18% of the variance towards work ability.	Yes
Loerbroks et al. (2016)	Rewards (perceived salary, promotion prospects, esteem, and job security) was positively correlated with quality of care (β=.20).	Yes

Appendix III: Job Demands and Job Resources' Relationship with Patient Safety

Author	Main Findings	Relationship Found
Job demands and patient safety		
Perceived Workload		
Dollarhide *et al*. (2013)	Perceived workload higher on a medication event day (M=35.9) than a medication non-event day (M=26.6).	Yes
Baldwin et al. (1997)	Feeling overwhelmed (r=.22) was associated with the number of mistakes made in the previous year.	Yes
Zwaan (2012)	Residents who reported higher subjective workload were associated with more adverse outcomes (OR=1.10).	Yes
Demanding Patients		
An et al. (2013)	Physicians with a high burden of difficult encounters had a 5.57% lower error rate than those with a lower burden. No significant differences for specific error rates for diabetes and hypertension.	Mixed
Time Pressure		
Linzer et al. (2009)	Neither time pressure nor pace were associated with any of the outcome measures.	Mixed
Job resources and patient safety		
Autonomy		
Naveh et al. (2015)	Autonomy (r=-.01) was not significantly related with error rate.	No
Stern *et al*. (2008)	Autonomy predicted treatment errors (β=-3.28).	Yes
Job control		
Tucker *et al.* (2012)	Work time control predicted concerns on patient safety (β=- .18).	Yes
Linzer et al. (2009)	Work control was not correlated with any of the four error measures	No
Learning and Development		
Naveh et al. (2015)	Consultation with physicians was correlated with error rate (r=-.14).	Yes
Baldwin et al. (1997)	Effective learning and skill use (r=-.180) were associated with the number of mistakes made in the previous year.	Yes

Appendix IV: Job Demands and Job Resources' Relationship with Patient Experience

Author	Main Findings	Relationship Found
Job demands and patient experience		
Perceived Workload		
Feddock *et al*. (2005)	Patient satisfaction with clinic visit lower when seen by a resident with heavier workload than one with a lighter workload on two out of the seven items.	Mixed
Ansmann et al. (2013)	No relationship was observed between work overload and patient satisfaction.	No
Ansmann et al. (2014)	No relationship between psychological job demands and patient satisfaction with support.	No
Time Pressure		
Ansmann et al. (2013)	Perceived lack of time for patient care was negatively associated with patient satisfaction (OR 1.62).	Yes
Perceived Physical Job Demands		
Ansmann et al. (2014)	Perceived physical activity demands (β=-0.44), but not work posture demands, was correlated with patient satisfaction with support.	Mixed
Higher-order Job Demands		
Mache et al. (2012)	Latent factor of three job demands negatively correlated with patient satisfaction (r=-.38)	Yes
Job resources and patient experience		
Job control		
Ansmann et al. (2014)	No relationship between decision latitude and patient satisfaction with support.	No
McKinstry et al. (2007)	Work control not related with patient satisfaction with communication or enablement.	No
Social support - Colleagues		
Ansmann et al. (2014)	Social capital (β=0.279), but not social support from colleagues, was associated with patient satisfaction.	Mixed
McKinstry et al. (2007)	Support not related with patient satisfaction with communication or enablement	No
Supervisor support		
Ansmann et al. (2014)	Supervisor support (r=.137) not associated with patient satisfaction.	No
Higher-order Job Resources		
Mache et al. (2012)	Latent factor of eight job resources (influence at work, degree of freedom of work, possibilities for development, quality of leadership, social support, feedback at work, social relations, & sense of community) positively correlated with patient satisfaction (r=.420).	Yes

UK Data Archive SN 7776 - National Health Service National Staff Survey, 2014

National NHS Staff Survey 2014

What is this survey and why are we asking you to complete it?

This is an independent survey of your experience of working in your organisation. The overall aim is to gather information that will help to improve the working lives of staff in the NHS and so help to provide better care for patients.

Your organisation will be able to use the results of the survey to improve local working conditions and practices and to increase involvement and engagement with staff. Other organisations, including NHS commissioners, the Care Quality Commission, the Department of Health, and NHS England, will make use of the results.

Please complete the survey for your current job, or the job you do most of the time. If you work across two or more employers in the NHS, please answer in relation to the organisation that pays your salary. Please read each question carefully, but give your immediate response by ticking the box which best matches your personal view.

Who will see my answers?

The survey is being conducted by [insert] and the NHS Staff Survey Co-ordination Centre on behalf of your organisation and NHS England.

Your answers will be treated in confidence. No one in your organisation will be able to identify individual responses. The bar code / number below is only used by [insert] to identify which staff should be sent a reminder and will not be available to staff in your organisation.

The survey findings will be analysed by [insert] and the NHS Staff Survey Co-ordination Centre and the results will be presented in a summary report in which no individual, or their responses, can be identified.

Please return this questionnaire, in the envelope provided, to:

[insert]
[insert]
[insert]
[insert]
[insert]

If you have any queries about this questionnaire please contact the [insert] helpline on [insert] or go to www.nhsstaffsurveys.com

NHS

YOUR PERSONAL DEVELOPMENT

1. Have you had any training, learning or development (paid for or provided by your organisation) in the following areas?	Yes, in the last 12 months	Yes, more than 12 months ago	No	Not applicable to me
Please include any taught courses or more informal ways of learning such as supervised on-the-job training, e-learning, shadowing, reading journals / manuals etc.				
a. Health and safety training	☐ 1	☐ 2	☐ 3	☐ 9
b. Equality and diversity training	☐ 1	☐ 2	☐ 3	☐ 9
c. How to prevent or handle violence and aggression to staff, patients / service users	☐ 1	☐ 2	☐ 3	☐ 9
d. Infection control (e.g. guidance on hand-washing, MRSA, waste management, disposal of sharps / needles)	☐ 1	☐ 2	☐ 3	☐ 9
e. How to handle confidential information about patients / service users	☐ 1	☐ 2	☐ 3	☐ 9
f. How to deliver a good patient / service user experience	☐ 1	☐ 2	☐ 3	☐ 9
g. Any other job-relevant training, learning or development	☐ 1	☐ 2	☐ 3	☐ 9

2. To what extent do you agree or disagree with the following statements?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
My training, learning and development has helped me to...					
a. ...do my job more effectively.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. ...stay up-to-date with professional requirements.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. ...deliver a better patient / service user experience.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

3a. In the last 12 months, have you had an appraisal, annual review, development review, or Knowledge and Skills Framework (KSF) development review?

1 ☐ Yes 2 ☐ No 3 ☐ Can't remember

If YES, please answer parts b to f below; if NO, go to Question 4

b. Did it help you to improve how you do your job?	1 ☐ Yes	2 ☐ No
c. Did it help you agree clear objectives for your work?	1 ☐ Yes	2 ☐ No
d. Did it leave you feeling that your work is valued by your organisation?	1 ☐ Yes	2 ☐ No
e. Were any training, learning or development needs identified?	1 ☐ Yes	2 ☐ No

If YES to Question 3e, please answer part f below; if NO, go to Question 4

f. Did your manager support you to receive this training, learning or development?	1 ☐ Yes	2 ☐ No

YOUR JOB

4. The following questions are about team working and relate to the group of people that you work with most closely.

a. Do you work in a team? 1 ☐ Yes 2 ☐ No

If NO, go to Question 5; if YES, please answer the following questions about the main team or group you work in:

Team members...	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
b. ...have a set of shared objectives.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. ...often meet to discuss the team's effectiveness.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. ...have to communicate closely with each other to achieve the team's objectives.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

5. For each of the statements below, how often do you feel this way about your job?	Never	Rarely	Sometimes	Often	Always
a. I look forward to going to work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. I am enthusiastic about my job.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. Time passes quickly when I am working.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

6. To what extent do you agree or disagree with the following statements about your job?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a. I have clear, planned goals and objectives for my job.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. I always know what my work responsibilities are.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. I am trusted to do my job.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. I am able to do my job to a standard I am personally pleased with.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

7. To what extent do you agree or disagree with the following statements about your work?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a. There are frequent opportunities for me to show initiative in my role.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. I am able to make suggestions to improve the work of my team / department.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. I am involved in deciding on changes introduced that affect my work area / team / department.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. I am able to make improvements happen in my area of work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e. I am unable to meet all the conflicting demands on my time at work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
f. I have adequate materials, supplies and equipment to do my work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
g. There are enough staff at this organisation for me to do my job properly.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

8. How satisfied are you with each of the following aspects of your job?	Very dissatisfied	Dissatisfied	Neither satis. nor dissatisfied	Satisfied	Very satisfied
a. The recognition I get for good work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. The support I get from my immediate manager.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. The freedom I have to choose my own method of working.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. The support I get from my work colleagues.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e. The amount of responsibility I am given.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
f. The opportunities I have to use my skills.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
g. The extent to which my organisation values my work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
h. My level of pay.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

9. Do the following statements apply to you and your job?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Not applicable to me
a. I am satisfied with the quality of care I give to patients / service users.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I feel that my role makes a difference to patients / service users.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
c. I am able to deliver the patient care I aspire to.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

YOUR MANAGERS

10. To what extent do you agree or disagree with the following statements about your immediate manager?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
My immediate manager...					
a. ...encourages those who work for her/him to work as a team.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. ...can be counted on to help me with a difficult task at work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. ...gives me clear feedback on my work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. ...asks for my opinion before making decisions that affect my work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e. ...is supportive in a personal crisis.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

11. To what extent do you agree or disagree with the following statements about senior managers where you work?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a. I know who the senior managers are here.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. Communication between senior management and staff is effective.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. Senior managers here try to involve staff in important decisions.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. Senior managers act on staff feedback.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e. Senior managers are committed to patient care.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

YOUR ORGANISATION

12. To what extent do these statements reflect your view of your organisation as a whole?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a. Care of patients / service users is my organisation's top priority.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. My organisation acts on concerns raised by patients / service users.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. I would recommend my organisation as a place to work.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. If a friend or relative needed treatment I would be happy with the standard of care provided by this organisation.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

13. Patient / service user experience measures	Yes	No	Don't know	Not applicable to me
a. Is patient / service user experience feedback collected within your directorate / department? (e.g. Friends and Family Test, patient surveys etc.)	☐ 1	☐ 2	☐ 3	☐ 9

If YES, please answer parts b and c below; if NO, go to Question 14

To what extent do you agree with the following statements about feedback from patients / service users?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	Don't know
b. I receive regular updates on patient / service user experience feedback in my directorate / department (e.g. via line managers or communications teams).	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 9
c. Feedback from patients / service users is used to make informed decisions within my directorate / department.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 9

YOUR HEALTH, WELL-BEING AND SAFETY AT WORK

14. To what extent do you agree or disagree with the following statements?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a. In general, my job is good for my health.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. My immediate manager takes a positive interest in my health and well-being.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. My organisation takes positive action on health and well-being.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

15a. In the last three months have you ever come to work despite not feeling well enough to perform your duties?

1 ☐ Yes 2 ☐ No

If YES, please answer parts b to d below; if NO, go to Question 16

b. Have you felt pressure from **your manager** to come to work? 1 ☐ Yes 2 ☐ No

c. Have you felt pressure from **colleagues** to come to work? 1 ☐ Yes 2 ☐ No

d. Have you put **yourself** under pressure to come to work? 1 ☐ Yes 2 ☐ No

16. During the last 12 months have you felt unwell as a result of work related stress?

1 ☐ Yes 2 ☐ No

17. In the last month have you seen any errors, near misses, or incidents that could have hurt...

a. Staff 1 ☐ Yes 2 ☐ No

b. Patients / service users 1 ☐ Yes 2 ☐ No

If YES to either a or b above, please answer part c below; if NO, go to Question 18

c. The **last** time you saw an error, near miss or incident that could have hurt **staff or patients / service users**, did you or a colleague report it?

1 ☐ Yes, I reported it 2 ☐ Yes, a colleague reported it 3 ☐ No 4 ☐ Don't know

18. To what extent do you agree or disagree with the following?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
a. My organisation treats staff who are involved in an error, near miss or incident fairly.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b. My organisation encourages us to report errors, near misses or incidents.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. My organisation treats reports of errors, near misses or incidents confidentially.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d. My organisation blames or punishes people who are involved in errors, near misses or incidents.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e. When errors, near misses or incidents are reported, my organisation takes action to ensure that they do not happen again.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
f. We are informed about errors, near misses and incidents that happen in the organisation.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
g. We are given feedback about changes made in response to reported errors, near misses and incidents.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

19. Raising concerns about unsafe clinical practice	Yes	No	Don't know		
a. If you were concerned about unsafe clinical practice, would you know how to report it?	☐ 1	☐ 2	☐ 9		
To what extent do you agree with the following statements about unsafe clinical practice?	**Strongly disagree**	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
b. I would feel secure raising concerns about unsafe clinical practice.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c. I am confident that my organisation would address my concern.	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

20. In the last 12 months how many times have you personally experienced physical violence at work from...?

a. Patients / service users, their relatives or other members of the public

1 ☐ Never 2 ☐ 1-2 3 ☐ 3-5 4 ☐ 6-10 5 ☐ More than 10

b. Managers / team leader or other colleagues

1 ☐ Never 2 ☐ 1-2 3 ☐ 3-5 4 ☐ 6-10 5 ☐ More than 10

c. The last time you experienced physical violence at work, did you or a colleague report it?

1 ☐ Yes, I reported it 2 ☐ Yes, a colleague reported it 3 ☐ No 4 ☐ Don't know 9 ☐ Not applicable

21. In the last 12 months how many times have you personally experienced harassment, bullying or abuse at work from...?

a. Patients / service users, their relatives or other members of the public

1 ☐ Never 2 ☐ 1-2 3 ☐ 3-5 4 ☐ 6-10 5 ☐ More than 10

b. Managers / team leader or other colleagues

1 ☐ Never 2 ☐ 1-2 3 ☐ 3-5 4 ☐ 6-10 5 ☐ More than 10

c. The last time you experienced harassment, bullying or abuse at work, did you or a colleague report it?

1 ☐ Yes, I reported it 2 ☐ Yes, a colleague reported it 3 ☐ No 4 ☐ Don't know 9 ☐ Not applicable

22. Does your organisation act fairly with regard to career progression / promotion, regardless of ethnic background, gender, religion, sexual orientation, disability or age?

1 ☐ Yes 2 ☐ No 9 ☐ Don't know

23. In the last 12 months have you personally experienced discrimination at work from any of the following?

a. Patients / service users, their relatives or other members of the public 1 ☐ Yes 2 ☐ No

b. Manager / team leader or other colleagues 1 ☐ Yes 2 ☐ No

If YES to either a or b above, please answer part c below; if NO, go to Question 24

c. On what grounds have you experienced discrimination? *Please tick all that apply*

1 ☐ Ethnic background 3 ☐ Religion 5 ☐ Disability 7 ☐ Other *(please specify)*

2 ☐ Gender 4 ☐ Sexual orientation 6 ☐ Age

BACKGROUND INFORMATION

We would like to know a bit more about you so that we can compare the experiences of different types of staff.

24. About you

a. **Gender:** 1 ☐ Male 2 ☐ Female

b. **Age:** 1 ☐ 16-20 2 ☐ 21-30 3 ☐ 31-40 4 ☐ 41-50 5 ☐ 51-65 6 ☐ 66+

25. Working hours

a. How many hours a week are you contracted to work?

1 ☐ Up to 29 hours   2 ☐ 30 or more hours

b. On average, how many *additional* PAID hours do you work per week for this organisation, over and above your contracted hours? *Please include paid overtime, bank shifts, and additional paid hours on-call.*

1 ☐ 0 hours   2 ☐ Up to 5 hours   3 ☐ 6-10 hours   4 ☐ 11 or more hours

c. On average, how many *additional* UNPAID hours do you work per week for this organisation, over and above your contracted hours? *Please include unpaid overtime and additional unpaid hours on-call.*

1 ☐ 0 hours   2 ☐ Up to 5 hours   3 ☐ 6-10 hours   4 ☐ 11 or more hours

26. What is your ethnic background?

White

01 ☐ British
02 ☐ Irish
03 ☐ Any other White background

Mixed

04 ☐ White and Black Caribbean
05 ☐ White and Black African
06 ☐ White and Asian
07 ☐ Any other mixed background

Asian/Asian British

08 ☐ Indian
09 ☐ Pakistani
10 ☐ Bangladeshi
11 ☐ Any other Asian background

Black/Black British

12 ☐ Caribbean
13 ☐ African
14 ☐ Any other Black background

Chinese and other ethnic background

15 ☐ Chinese
16 ☐ Any other ethnic background *(please specify)*

27. Which of the following best describes how you think of yourself?

1 ☐ Heterosexual (straight)   2 ☐ Gay Man   3 ☐ Gay Woman (lesbian)
4 ☐ Bisexual   5 ☐ Other   6 ☐ I would prefer not to say

28. What is your religion?

1 ☐ No religion
2 ☐ Christian
3 ☐ Buddhist
4 ☐ Hindu
5 ☐ Jewish
6 ☐ Muslim
7 ☐ Sikh
8 ☐ Any other religion *(please specify)*
9 ☐ I would prefer not to say

29. Disability

a. Do you have a long-standing illness, health problem or disability?   1 ☐ Yes   2 ☐ No

By long-standing, we mean that it has lasted, or will last, for at least 12 months
If YES, please answer part b below; if NO, go to Question 30

b. Has your employer made adequate adjustment(s) to enable you to carry out your work?

1 ☐ Yes   2 ☐ No   3 ☐ No adjustment required

30. Do you have face-to-face contact with patients / service users as part of your job?

1 ☐ Yes, frequently   2 ☐ Yes, occasionally   3 ☐ No

31. How many years have you worked for this organisation?

If your organisation has merged with another or changed its name, please include in your answer all the time you have worked with this organisation and its predecessors

1 ☐ Less than 1 year
2 ☐ 1-2 years
3 ☐ 3-5 years
4 ☐ 6-10 years
5 ☐ 11-15 years
6 ☐ More than 15 years

32. What is your occupational group?

Please tick one box only

Allied Health Professionals / Healthcare Scientists / Scientific and Technical

01 ☐ Occupational Therapy
02 ☐ Physiotherapy
03 ☐ Radiography
04 ☐ Pharmacy
05 ☐ Clinical Psychology
06 ☐ Psychotherapy
07 ☐ Arts therapy *(e.g. art, music, drama therapy)*
08 ☐ Other qualified Allied Health Professionals *(e.g. dietetics, speech and language therapy, complementary therapy)*
09 ☐ Support to Allied Health Professionals *(e.g. support worker, therapy helper, therapy assistant or student)*
10 ☐ Other qualified Scientific and Technical or Healthcare Scientists *(e.g. haematology, clinical biochemistry, microbiology)*
11 ☐ Support to healthcare scientists *(e.g. technicians, assistants or students)*

Medical and Dental

12 ☐ Medical / Dental - Consultant
13 ☐ Medical / Dental - In Training *(e.g. Foundation Y1 & Y2, StRs (incl FTSTAs & LATs), SHOs, SpRs / SpTs / GPRs)*
14 ☐ Medical / Dental - Other *(e.g. Staff and Associate Specialists / Non-consultant career grade)*

Ambulance (operational)

15 ☐ Emergency Care Practitioner
16 ☐ Paramedic
17 ☐ Emergency Care Assistant
18 ☐ Ambulance Technician
19 ☐ Ambulance Control Staff *(e.g. call handler, dispatchers, PTS controllers)*
20 ☐ Patient Transport Service *(e.g. ambulance drivers, support staff)*

Public Health

21 ☐ Public Health / Health Improvement

Commissioning

22 ☐ Commissioning managers / support staff

Registered Nurses and Midwives

23 ☐ Adult / General
24 ☐ Mental health
25 ☐ Learning disabilities
26 ☐ Children
27 ☐ Midwives
28 ☐ Health Visitors
29 ☐ District / Community
30 ☐ Other Registered Nurses

Nursing or Healthcare Assistants

31 ☐ Nursing auxiliary / Nursing assistant / Healthcare assistant *(including Health / Clinical / Nursing Support Worker)*

Social Care

32 ☐ Approved social workers / Social workers / Residential social workers
33 ☐ Social care managers
34 ☐ Social care support staff

Wider Healthcare Team

35 ☐ Admin & Clerical *(including Medical Secretary)*
36 ☐ Central Functions / Corporate Services *(e.g. HR, Finance, Information Systems, Information Technology)*
37 ☐ Maintenance / Ancillary *(e.g. housekeeping, domestic staff, maintenance, facilities, estates)*

General Management

38 ☐ General Management *(N.B. If you are a manager and can choose a group from elsewhere in the list, please select that other occupational group)*
39 ☐ Other occupational group *(please specify)*

If you have any additional comments about working in this organisation, please write these on a separate sheet and attach them to this questionnaire

Appendix VI: Direct Effects onto Work-related Wellbeing (MSEM)

This appendix is linked with the study in Chapter Eight, where multilevel structural equation modelling (MSEM) was not possible to test for interactions. The table below can be compared against Table 8.3, and demonstrates that while effects are consistent, workplace aggression had stronger effects in MSEM than in multilevel modelling. The model fit for the table below was good (RMSEA= .049; CF=: .988; TLI= .984).

Table IV. *Standardised coefficients for direct effects onto work-related wellbeing*

Measure	Work-related stress	Presenteeism	Work engagement
Tenure (w)	.08***	.06***	-.07***
Specialist (b)	-.201**	.20	.33**
Beds (b)	.19*	-.79*	.17**
Insufficient work resources (w)	.16***	.07**	-.19***
Workplace aggression (w)	.52***	.46***	-.23***
Manager support (w)	-.06**	-.03	.07***
Effective team practices (w)	-.02	-.04*	.10***
Job control (w)	-.13***	-.08***	.34***
Bed occupancy (b)	-.13**	.02	.33**
Emergency admissions (b)	-.02**	.62*	-.50**

Note: ****p*<.001, ***p*<.01, **p*<.01; (b) = trust-level predictor; (w) = individual-level predictor.