KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Public Health and Primary Care

MULTILEVEL AND STRUCTURAL EQUATION MODELS IN HEALTH SERVICES RESEARCH

APPLICATION TO RESEARCH ON THE ORGANIZATIONAL CONTEXT OF NURSING CARE

Luk BRUYNEEL

Jury:

Promoter:	prof. dr. Walter Sermeus
Co-promoter:	prof. dr. Emmanuel Lesaffre
Chair:	prof. dr. Rik Lories
Secretary :	prof. dr. Lode Godderis
Jury members:	prof. dr. Hans De Witte
	prof. dr. Sanna Salanterä
	prof. dr. Peter Van Bogaert
	prof. dr. Roger Watson

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Biomedical Sciences

2015

I dedicate this work to my wife and son. Lore, thank you for always being there for me. I'm proud of everything you have accomplished. Joppe, it is priceless to see you master some major learning events as a toddler.

I love both of you more and more each day.

I thank my family for supporting me throughout my life. Your care allowed me to successfully complete this journey.

To my lost grandparents who left us too soon, I hope this work makes you proud.

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my PhD supervisors, Professor Walter Sermeus and Professor Emmanuel Lesaffre. Professor Sermeus, as founder and coordinator of the international RN4CAST research consortium, you have provided the core knowledge, direction and data needed for many early career researchers to initiate doctoral studies in health services research and statistics. A special thank you for allowing an incubation period for my research skills to develop effectively and for managing my career. Professor Lesaffre, you facilitated originality in my work and taught me how to distill complex statistical material into readable narrative. I admire your capability to train so many young researchers with diverse backgrounds in various research areas. To both, I hope we can continue this successful collaboration.

Professor Kris Vanhaecht, I would like to express my deepest appreciation for the many helpful discussions we had on numerous research and personal topics. These discussions shaped my career ambitions. Doctor Koen Van den Heede, your mentorship in the early days of my career as a researcher was paramount. Daan Aeyels, thank you for being a great friend and making me part of your family. Thank you all for being exemplary researchers. Most of all, thank you for showing that great researchers can be great husbands and fathers.

Professor Linda Aiken, it is an honor to work with the world authority in this field. You are a pioneer and game changer, a continuous source of information, and amazing at giving invaluable constructive criticism to my work. I will always be grateful to you and your great time at the University of Pennsylvania: Timothy Cheney, Professor Ann Kutney-Lee, Professor Eileen Lake, Professor Matthew McHugh, Professor Doug Sloane, and Professor Herbert Smith.

Professor Anne Marie Rafferty, you are my favorite professor of enthusiasm. Thank you for your invaluable contributions to the RN4CAST study and thank you for taking care for me during my research stay in London.

Doctor Baoyue Li, we produced some very nice manuscripts during our three-year collaboration. Thank you for your brilliant comments and suggestions.

I owe a great debt of gratitude to all members from the RN4CAST consortium: Professor Sabina De Geest, doctor PD René Schwendimann and doctor Dietmar Ausserhofer (University of Basel, Switzerland), Professor Tischelman and doctor Rikard Lindqvist (Karolinska Institute, Sweden), Professor Rafferty (King's College London, UK), Jane Ball and Professor Griffiths (University of Southampton, UK), Professor John Mantas (University of Athens, Greece), Professor Reinhard Busse (Berlin University of Technology, Germany), Professor Juha Kinnunen and Professor Anneli Ensio (University of Eastern Finland, Finland), Professor Tomasz Brzostek (Jagiellonian University, Poland), Professor Morena-Casbas (Institute of Health Carlos III, Spain), Professor Anne Scott and Professor Anne Matthews (Dublin City University, Ireland), doctor Maud Heinen, doctor Lisette Schoonhoven and Professor Theo Van Achterberg (Radboud University Nijmegen/KU Leuven) and doctor Ingeborg Stromseng Sjetne (Norwegian Knowledge Centre for the Health Services, Norway).

Marleen Van Eygen and Mieke Florquin, it was a joy to work with you as students in our master programme. The qualitative study design and your experience as nurse managers provided innovative perspectives to the RN4CAST data. Thank you for sharing your knowledge and experience with me.

Professor Martin Mckee, thank you for subtly raising concerns about measurement non-invariance in cross-national research during our RN4CAST meeting in Spain in 2010. It was during this time that the hypotheses on measurement invariance in Chapter 6 of this PhD study were initially conjectured.

I would like to express my gratitude to Professor Bart Meuleman for co-supervising my master thesis in statistics and for co-authoring Chapter 6 of this PhD study. But above all, I would like to thank you for introducing me to the world of structural equation models, which has been of great value to this PhD study.

I would like to thank Professor Hans De Witte, Professor Lode Godderis, Professor Sanna Salanterä, Professor Peter Van Bogaert, and Professor Roger Watson for serving as committee members. Special

thanks go to Professor Philip Moons for coordinating the committee members' feedback and for chairing the public defense.

There have been many others at KU Leuven who have helped me along the way that I like to express my warm thanks to: Silke Apers, Seher Arat, Ellen Coeckelberghs, doctor Eline Dancet, Kristel De Vliegher, doctor Svin Deneckere, doctor Mieke Deschodt, Els Devriendt, Nathalie Duerinckx, Eva Goossens, doctor Deborah Hilderson, Astrid Janssens, doctor Marieke Kroezen, doctor Cathy Lodewijckx, doctor Shahenaz Najjar, Thijs Nelis, Marc Nickmilder, Kristien Scheepmans, Deborah Seys, Kristien Van der Elst, Liz Van der Veken, doctor Pieter Van Herck, Eva Van Gerven and Ellen Vlaeyen.

Last but not least, I would like to extend my thanks to Anja Bruers, Evelyn Cousy, Heidi Demin, Alexandra Pardon, Marjolijn Seuckens, Caroline Thys and Mia Veuchelen for their support in RN4CAST, help in managing my work as an ombudsperson, organizing my public defense, and so many other things.

Thank you

LIST OF ABBREVIATIONS

AHRQ	Agency for Healthcare Research and Quality
BSEM	Bayesian Structural Equation Model
CFA	Confirmatory Factor Analysis
CVI	Content Validity Index
EC	European Commission
EFA	Exploratory Factor Analysis
FP7	Seventh Framework Programme of the European Commission
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
HFE	Human Factors and Engineering
ICC	Intraclass Correlation Coefficient
IHI	Institute for Healthcare Improvement
IHOS	International Hospital Outcomes Study
IMF	International Monetary Fund
IOM	Institute of Medicine
MBI	Maslach Burnout Inventory
MCR	Multilevel Covariance Regression
MGCFA	Multiple Group Confirmatory Factor Analysis
MIMIC	Multiple Indicators Multiple Causes
MVP	Multivariate Probit Model
NDNQI	National Database of Nursing Quality Indicators
NPSA	National Patient Safety Agency
NQF	National Quality Forum
NWI	Nursing Work Index

NWI-R	Nursing Work Index – Revisited
OECD	Organisation for Economic Cooperation and Development
PES-NWI	Practice Environment Scale of the Nursing Work Index
PPP	posterior predictive p-values
QHOM	Quality Health Outcomes Model
RICH	Rationing of Nursing Care in Switzerland
RN4CAST	Registered Nurse Forecasting
SEIPS	System Engineering Initiative for Patient Safety
SEM	Structural Equation Model
SoPK	System of Profound Knowledge
TCAB	Transforming Care at the Bedside
WHO	World Health Organization

TABLE OF CONTENTS

TABLES AND FIGURES

Tables

Figures

1 GENERAL INTRODUCTION AND RESEARCH OBJECTIVES

1.1 Systems thinking and patient safety research

In September 1999, the Institute of Medicine (IOM) issued its 'To Err is Human: Building a Safer Health System' report (Institute of Medicine (U.S.) Committee on Quality of Health Care in America, 2000). Based on extrapolations from observational studies using chart reviews in New York (Brennan et al., 1991) and Colorado and Utah (Thomas et al., 2000) the IOM report indicated that between 44,000 and 98,000 people die in U.S. hospitals each year as a result of medical errors that could have been prevented. Patient safety problems in hospital care received increased recognition after the release of this seminal report (Leape & Berwick, 2005). 'To Err is Human' would be the first of eleven books within the IOM Quality Chasm Series. The second book was titled 'Crossing the Quality Chasm' and focused on strategies for health care systems to innovate and improve care (Institute of Medicine (U.S.) Committee on Quality of Health Care in America, 2001). Many countries and international institutions put patient safety as a top priority as a direct result of the concerns that were outlined in these IOM reports. The U.S. based Institute for Healthcare Improvement (IHI) started offering programs for developing Patient Safety Officers since the early 2000's. In the UK, the National Patient Safety Agency (NPSA) was established in 2001 (and abolished in 2012) with a mandate to identify patient safety issues and find appropriate solutions. The Organisation for Economic Cooperation and Development (OECD) initiated the Health Care Quality Indicators Project in 2002, which aims to measure and compare the quality of health services in its Member States. Patient safety is one of their key areas of focus. In 2004, the World Health Organization (WHO) launched the World Alliance for Patient Safety to reduce the adverse health and social consequences of unsafe health care. This research team has identified how priorities for patient safety research differ between developing, transitional and developed countries (Bates, Larizgoitia, Prasopa-Plaizier, & Jha, 2009). A last example is the Australian Commission on Safety and Quality in Health Care, established in 2006 to lead and coordinate safety and quality improvement nationally.

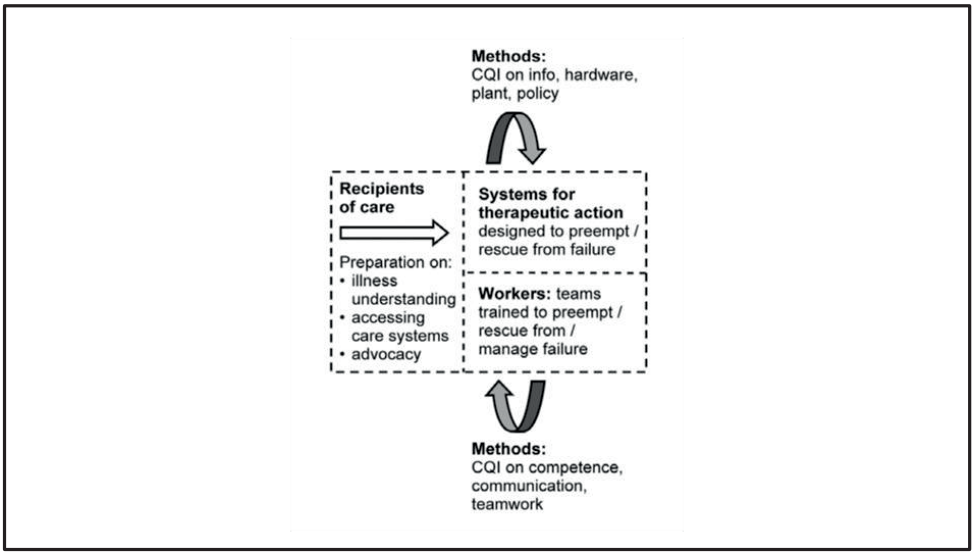
An interrupted time series analysis found that the To Err is Human report led to a large increase in patient safety publications and research awards. But more importantly, this analysis showed that the IOM initiated a paradigm shift in the patient safety literature from reports of malpractice to reports on organizational culture (Stelfox, Palmisani, Scurlock, Orav, & Bates, 2006). In the words of Wachter (2012):

"The modern patient safety movement replaces the blame and shame game with an approach known as systems thinking". This systems thinking approach is acknowledged in the Agency for Healthcare Research and Quality's (AHRQ) definition of patient safety as "A discipline in the health care sector that applies safety science methods toward the goal of achieving a trustworthy system of health care delivery. Patient safety is also an attribute of health care systems; it minimizes the incidence and impact of, and maximizes recovery from, adverse events" (Emanuel et al., 2008).

Many excellent models and tools have been proposed to identify, measure and manage underlying patient safety issues using a systems orientation approach. About fifty years ago, Donabedian (1966, 1988) advanced the idea that the quality of medical care can be assessed by examining healthcare structures, processes and outcomes. Structures describe the setting in which care takes place, including for example the adequacy of facilities and equipment and the qualification of the medical staff. Processes denote the transactions between patients and providers throughout the delivery of healthcare and relate to the question whether medicine is properly practiced. Outcomes refer to the effects of health care on recovery, restoration of function and survival. Donabedian's perspective was one-directional, implying that structures affect processes, which in turn affect outcomes. The Quality Health Outcomes Model (QHOM) added dynamism to Donabedian's work by recognizing feedback that occurs among clients, the system in which care is provided, and interventions (Mitchell, Ferketich, & Jennings, 1998). Carayon's (2006) Systems Engineering Initiative for Patient Safety (SEIPS) model also builds on Donabedian's influential work. Carayon's conceptions of studying patient safety in a systems framework are anchored in a human factors and ergonomics (HFE) approach. HFE is defined as "The scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance" (International Ergonomics Association, 2014). Whereas Donabedian's model solely focuses on patient outcomes, the SEIPS model also specifies the components of the care system that can contribute to employee and organizational outcomes. Such outcomes include employee satisfaction, burnout, safety and turnover, which can correlate with patient outcomes. It also puts increased attention on structures and looks at other processes besides care processes, such as

information flow, purchasing and cleaning. Several examples of HFE-based interventions for patient safety using the SEIPS model are described by Carayon, Xie and Kianfar (2014). SEIPS 2.0 (Holden et al., 2013) added three salient components to Carayon's work, including configuration (potential interaction between all components of the model), engagement (work can involve professionals, patients, or can be a collaboration between professionals and patients) and adaptation (feedback mechanisms). One of the developers' conclusions is that these additions allow the SEIPS 2.0 model to test emerging complex research questions on cross-level effects (e.g. organizational-level factors affecting individual-level factors) and the replication of effects across levels (e.g. parallel relationships at individual and organizational levels), pointing to surge of interest in healthcare towards multilevel analyses of complex phenomena. As explained later on, the challenges associated with the use of multilevel modeling techniques present an excellent opportunity for this PhD study to contribute to this field of research.

Figure 1.1. Patient safety model of health care (Emanuel et al. 2008)



Drawing on the ideas of Donabedian and other theoreticians, a team of patient safety experts described an overarching patient safety model of health care with four components: workers in health care, recipients of care, systems for therapeutic action, and methods for feedback and continuous improvement (Figure 1.1). This is an important effort to give a simple illustration of what constitutes

patient safety, where patient safety happens (systems for therapeutic action), how patient safety is achieved (methods for feedback and continuous improvement), and who are the people involved in patient safety (recipients and workers). The dotted lines mean that all four components interact with each other as well as with the environment (Emanuel et al., 2008).

Donabedian's notions were described to be crosscutting these four domains. Other models on the other hand have specifically been conceived as fitting unambiguously to one or several domains, such as the System of Profound Knowledge (SoPK), which is the result of Deming's work on quality, management and leadership (Deming, 1993). This model holds four areas for driving quality improvement. First, appreciation for a system refers to defining a system as an interdependent group of items, people or processes working together toward a common purpose. This implies a need for cooperation and integration instead of competition. Second, knowledge of variation is about recognizing and understanding different types of variation, their causes and their implications. The third area is building knowledge. With more knowledge about the system come tailored interventions. The fourth and last area is that of psychology. Successful management begins with acknowledging that people are different, and recognizing and optimizing workers' skills within an organization. Emanuel and colleagues placed all four of Deming's areas for quality improvement under the methods component of their patient safety model. Another well-established model, the Organizational Accident Model (Vincent, Taylor-Adams, & Stanhope, 1998) also found its way to the overarching patient safety model. The Organizational Accident Model is based on Reason's (2000) Swiss Cheese Model of Accident Causation. Using a human factors approach, Reason studied the conditions under which humans work. He concentrated on system defenses, barriers and safeguards as key concepts of a system approach. System failures can be caused by active failures (committed by people) or by latent conditions (resident pathogens within the system). Opportunities for active or latent errors to contribute to an accident are called windows of opportunity. If these windows of opportunity align at all levels in an organization, an accident occurs. Reason pictured this accident trajectory model as aligning holes in slices of Swiss cheese. In establishing high-reliability organizations, which Reason defines as "systems operating in hazardous conditions that have fewer than their fair share of adverse events", a reporting culture of how and why defenses fail is a key element for error

management. Vincent et al. translated this model to a medical setting. They identified seven crucial elements that influence safety. These seven elements and their respective categorization in the overarching patient safety model from Emanuel et al. are: patient characteristics (recipients of care), team factors and individual staff factors (workers in health care), organizational and management factors, work environment, task factors and institutional context (systems for therapeutic action). Vincent and co-workers recommend quality and safety initiatives to be multifactorial by addressing as much of these seven elements jointly, and added that interventions may oppose themselves at several points in the organizational hierarchy. The idea that different levels of an organization can be the focus of interest will be a major topic of interest in this PhD study.

The above models stand out in terms of redirecting conversations on patient safety from an individual perspective to a systems perspective. They have offered new ways of studying and improving quality and safety in health care organizations and provide the foundations for many models of nursing care delivery that are presented in the next section.

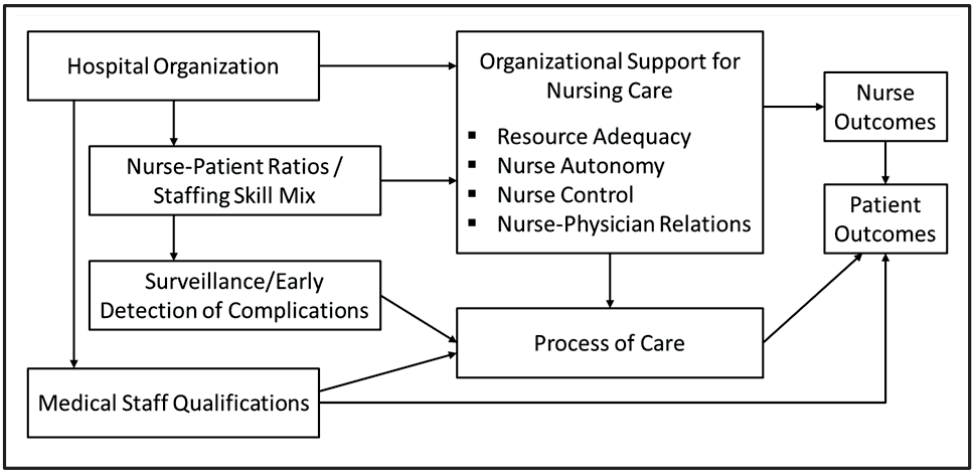
1.2 Models considering patient safety from a nursing perspective

A select number of health services researchers initiated research with a specific focus of interest on the extent to which human nursing capital and the organizational practices of nursing are part of the explanation for patient, nurse and organizational outcomes. Experts in this field have provided a theoretical basis for guiding these studies. The most widely cited theoretical work is Aiken, Sochalski and Lake's (1997) early conceptual framework that set out operant mechanisms relating organizational attributes, with a particular focus on nursing's contributions, to outcomes. Their model is theoretically founded on Donabedian's conceptions, including the dynamism between model components that was proposed by Mitchell and colleagues. The framework posits that hospital and nursing unit level organizational models that positively influence nurse autonomy, control by nurses of resources at the unit level, and relations between nurses and physicians, yield better nurse, patient and organization outcomes. In this model, nursing is seen as an on-going surveillance system. This system is better in detecting early signs of complications under better patient-to-nurse ratios and when more nurses are

educated to a bachelor's degree. This thinking was integrated into an international research agenda which initially put strong emphasis on analyzing effects of patient-to-nurse ratios and nurse education levels on patient outcomes. Measuring the nursing work environment (organizational support for nursing care) required the development of a method and instrument (Aiken, Clarke, & Sloane, 2002). This organizational construct of major interest has since been defined as "the organizational characteristics of a work setting that facilitate or constrain professional nursing practice" (Lake, 2002, 2007). Research into the quality of nursing work environments was the result of another seminal IOM Quality Chasm report, 'Keeping Patients Safe: Transforming the Work Environment of Nurses', which called for reinforcing patient safety defense in nurses' work environments (Institute of Medicine (U.S.) Committee on the Work Environment for Nurses and Patient Safety, 2004).

Figure 1.2 displays the conceptual framework from Aiken, Clarke, and Sloane (2002). Swiss researchers further developed the model and hypothesized that inadequate staffing and poor practice environments are linked with a higher degree of daily tasks that nurses withhold or fail to carry out. This is labelled rationing of care, which in turn relates to patient satisfaction and clinical outcomes. It is called the Rationing of Nursing Care in Switzerland (RICH) model (Schubert et al., 2008; Schubert, Glass, Clarke, Schaffert-Witvliet, & De Geest, 2007).

Figure 1.2 Conceptual framework linking the organization of nursing care to nurse and patient outcomes (Aiken et al. 2002)



Other health services researchers have established complementary organizational models to guide outcomes research in nursing. Within Donabedian's process component, the Nursing Role Effectiveness Model defined independent role functions (for which nurses are held accountable), dependent role functions (implementation of medical orders and treatments) and interdependent role functions (shared responsibilities). Effectively engaging in these roles is theorized to depend on unit structural variables such as nurse staffing and nurse leadership (Doran, Sidani, Keatings, & Doidge, 2002; Irvine, Sidani, & Hall, 1998). The Nursing Worklife Model relates the multidimensionality of nurses' perceptions of professional practice environments (leadership, nurse participation in hospital affairs, collegial nurse-physician relations, adequate staffing, nursing models of care) to nurse burnout (Manojlovich & Laschinger, 2007, 2008). The authors put strong emphasis on testing this theoretical model using path analysis to describe dependencies among the practice environment dimensions. Last, Nursing Services Delivery Theory (Meyer & O'Brien-Pallas, 2010) draws on Katz and Kahn's (1978) work on Open Systems Theory, which acknowledges the environment's influence on systems. The authors define a dynamic interaction between input, throughput and output. In line with Open Systems Theory, Nursing Services Delivery Theory recognizes that workers are clustered in work groups that are nested in the system or even other levels within the system. Consistent with SEIPS 2.0, Nursing Services Delivery Theory explicates the study of multiple levels of phenomena and the examination of cross-level effects. It also calls for researchers to examine 'why' and 'how' hypotheses to probe more deeply into the nature of known relationships.

Aiken and colleagues' framework is by far the most widely empirically tested model. It initiated an international program of research in 1997 that today generated evidence in 27 countries in which a common core of instrumentation was used to study organizational features of nursing care (Armenia, Belgium, Botswana, Canada, China, Cyprus, England, Finland, Germany, Greece, Iceland, Ireland, Japan, the Netherlands, New Zealand, Norway, Poland, Portugal, Russia, Scotland, South Africa, South Korea, Spain, Sweden, Switzerland, Thailand, and the U.S.). The challenges and rewards of developing this unique international consortium are described by Aiken et al. (1997), Sochalski and Aiken (1999), Cheung and Aiken (2008), Clarke and Aiken (2008) and Sermeus et al. (2011). This consortium of researchers is largely

responsible for documenting in an empirical way that there is large variation in every aspect of nursing organization within and between countries. Many of these studies have documented that this variation is linked to a wide range of nurse and patient outcomes. The next section provides a detailed overview of landmark studies on the organization of nursing care and consequences for patients and nurses in the U.S. and Europe.

1.3 Hallmarks of nursing and patient oriented health services research

The past decade has witnessed a large increase in studies bringing together the issues of patient safety and the organizational context of nursing care. This research was largely sparked by a 1996 IOM report calling for more rigorous research on the association of nurse staffing and patient outcomes in hospitals, since the evidence supporting such relationship was inconclusive (Institute of Medicine (U.S.) Committee on the Adequacy of Nursing Staff in Hospitals and Nursing Homes, 1996). The section below provides a summary of the main studies in this field. It will be evident from this overview that important explanations for avoidable poor nurse and patient outcomes are poor nurse work environments, fewer baccalaureate nurses, and less favorable nurse staffing.

In their ground-breaking study, Needleman et al. (2002) used administrative data from 799 U.S. hospitals and found that higher nurse staffing is associated with shorter length of stay and lower complication rates (e.g. lower rates of pneumonia, upper gastrointestinal bleeding and urinary tract infection). Also at that moment, the International Hospital Outcomes Study (IHOS) was initiated to study the work settings of nurses in the U.S., Canada, England, Scotland and Germany. The IHOS showed that nurses across these five countries reported deficiencies in the organization of hospital nursing care, resulting in high levels of job dissatisfaction, burnout, intention to leave the hospital and low ratings of quality of care (Aiken et al., 2001). A landmark U.S. IHOS study by Aiken, Clarke, Sloane, Sochalski and Silber (2002) showed that nurse staffing explained variation in patient mortality across hospitals after controlling for patient (age, sex, surgery types, comorbidity) and hospital (size, teaching status, technology) characteristics. Especially patient characteristics are important for risk-adjustment (Charlson, Pompei, Ales, & MacKenzie, 1987; Elixhauser, Steiner, Harris, & Coffey, 1998; Silber, Rosenbaum, & Ross,

1995). These findings were based on data for 168 Pennsylvania hospitals and 232,342 surgical patients, of whom 4,535 died within 30 days of being admitted for the number of patients per nurse ranging from 4 to 8 between hospitals. The authors estimated that 535 less patients might have died if all hospitals would have had a ratio of 4 patients per nurse in each hospital. A much acclaimed follow-up analysis would show that also the proportion of hospital nurses holding a bachelor's degree (versus coming from associate degree of hospital school programs) is associated to risk-adjusted patient mortality. A 10% increase in the proportion of nurses holding a bachelor's degree was associated with a 5% decrease in the likelihood of patients dying within 30 days of admission. The authors then estimated that another 190 patients less might have died at a ratio of 4 patients per nurse and 60% bachelor educated nurses in each hospital (Aiken, Clarke, Cheung, Sloane, & Silber, 2003). Similar effects of nurse staffing and education on patient mortality were shown in Canadian settings (Alberta and Ontario), by Tourangeau et al. (2007) and Estabrooks et al. (2005). Years later, a secondary data analysis of the U.S. IHOS data indicated that, next to nurse staffing and nurse education, also nurse work environments impact patient mortality and nurse job satisfaction, burnout, and intention to leave the hospital (Aiken, Clarke, Sloane, Lake, & Cheney, 2008; Friese, Lake, Aiken, Silber, & Sochalski, 2008). Next, in what is referred to by Wachter (2012) as the most methodologically rigorous study to date in this field, Needleman and colleagues (2011) examined day-to-day and shift-to-shift variation in nurse staffing at the nursing unit level in a single hospital. They found an association between nurse staffing below target levels and increased patient mortality. Another key study in 2011 showed that both better staffing and better education have a much bigger effect on mortality in hospitals that have a good work environment. Nursing education was associated with better patient outcomes, independent of the quality of the work environment, but this effect was larger in hospitals with a good work environment. Likewise, staffing has a very large effect on reducing mortality in hospitals with a good work environment but no effect in hospitals with a poor work environment (Aiken et al., 2011). Next to this series of patient mortality studies, there were initiatives to link nursing to positive patient outcomes. Strong evidence showed that increased nurse staffing, better work environments and better nurse outcomes are associated to higher patient satisfaction with care in U.S. hospitals (Kutney-Lee et al., 2009; Vahey, Aiken, Sloane, Clarke, & Vargas, 2004).

These North-American research findings have clearly defined points of leverage for preventing poor patient outcomes that are related to the organization of nursing care. Thirteen U.S. states have enacted legislation or adopted regulations around nurse staffing ratios (American Nurses Association, 2014b), and 16 states have restricted mandatory overtime (American Nurses Association, 2014a). Aiken et al. (2010) showed that hospital minimum nurse staffing ratios mandated in California are associated with lower patient mortality. Also, nearly 400 hospitals in the U.S. have achieved 'Magnet status', so-called due to the ability to attract and retain nurses because of good work environments. These hospitals have lower levels of nurse job dissatisfaction and burnout (Kelly, McHugh, & Aiken, 2012) and lower levels of patient mortality (McHugh et al., 2013). In addition to these standards, there has been an effort to develop quality and safety measures linked to the adequacy of nursing care. The National Database of Nursing Quality Indicators (NDNQI) provides quarterly reporting of structure, process and outcome indicators to evaluate nursing care at the unit level (Montalvo, 2007). All measures related to nurse staffing, skill mix and work environment are endorsed by the National Quality Forum in its national voluntary consensus standards for nursing-sensitive care (National Quality Forum, 2004).

Also in Europe, researchers that were part of the original IHOS or extensions thereof, showed evidence relating features in the organization of nursing care to nurse and patient outcomes. Single country studies in England, Belgium and Switzerland partly confirmed the direction and consistency of the effects found in North-American studies. English researchers examined the effects of hospital-wide nurse staffing levels on a series of patient outcomes and nurse wellbeing by linking cross-sectional survey data with discharge datasets. They compared hospitals from the upper quartile ('most favorable staffing levels') with hospitals from the lowest quartile ('less favorable staffing levels') and found consistently better patient outcomes and nurse wellbeing in hospitals from the upper quartile. Patients in hospitals in the upper quartile were 26% more likely to die overall than those in hospitals in the lowest quartile (Rafferty et al., 2007). Evidence from Switzerland relies on a single-center cohort study conducted in a medical intensive care unit. These authors found that lower nurse staffing is associated with increased risk for late-onset ventilator associated pneumonia. An association with early-onset ventilator associated pneumonia was not found. The authors categorized nurse staffing in four groups, the cut-off values being

arbitrarily the 25th, the 50th and the 75th percentile of the patient-to-nurse ratio's distribution (Hugonnet, Uçkay, & Pittet, 2007). Using the RICH model, which is an extension of the IHOS as mentioned in the previous section, other Swiss researchers found a link between rationing of nursing care and lower patient satisfaction, increased nurse-reported medication errors, patient falls, nosocomial infections, pressure ulcers, critical incidents involving patients over the previous year (Schubert et al., 2008) and patient mortality (Schubert, Clarke, Aiken, & de Geest, 2012). A Belgian research group (Van den Heede, Sermeus, et al., 2009) initially found no significant association between the acute hospital-level risk-adjusted nurse staffing measures and patient outcomes. To better understand the dynamics of their constructs of interest in acute hospitals, the authors later examined the association on the unit level using multilevel models. In their sample of post-operative cardiac surgery patients they showed that better nurse staffing and education levels are significantly associated with decreased mortality (Van den Heede, Lesaffre, et al., 2009). Similar findings compared to the U.S. are also found in Europe for what concerns the relationship between worse work environments and negative outcomes for nurses, for example in Belgium (Bruyneel, Van den Heede, Diya, Aiken, & Sermeus, 2009; Van Bogaert, Clarke, Vermeyen, Meulemans, & Van de Heyning, 2009), Iceland (Gunnarsdóttir, Clarke, Rafferty, & Nutbeam, 2009) and Finland (Tervo-Heikkinen, Partanen, Aalto, & Vehviläinen-Julkunen, 2008). The same Finish researchers also confirmed U.S. findings that patients are more satisfied with better nurse staffing (Tervo-Heikkinen, Kvist, Partanen, Vehviläinen-Julkunen, & Aalto, 2008).

Funded under the Seventh Framework Programme of the European Commission, the Registered Nurse Forecasting (RN4CAST) consortium extended this line of research to a broader European context. The consortium brings together researchers from twelve European countries (Belgium, England, Finland, Germany, Greece, Ireland, the Netherlands, Norway, Poland, Spain, Sweden, Switzerland). Also included are the U.S. and three international cooperating partner countries from the European Union (Botswana, China, and South Africa). RN4CAST studies the effects of nursing workforce dynamics such as the number of nurse staff, skill-mix and working environment on nurse wellbeing and patient outcomes (Sermeus et al., 2011). A full overview of the sampling design and instrumentation is given in Chapter 2.

A first major report from the RN4CAST study linked nurses' reports on the characteristics of their work setting to patients' experiences with hospital care. Findings showed that nurses who reported better working conditions (work environment, staffing) also had patients who were more satisfied with their hospital stay. There was also large agreement between nurses and patients as to which hospitals provided good care. This association was shown repeatedly in all countries (Aiken et al., 2012). This report also confirmed associations between better work environment, better staffing, and better nurse outcomes. A secondary data analysis by Heinen et al. (2013) studied the specifics of this relationship with regards to nurses' intention to leave the hospital, showing that particularly the following system level factors are associated with intention to leave: nurse-physician relations, leadership, and participation in hospital affairs. Also in the South African sub study of RN4CAST it was the practice environment that mainly influenced nurses' intention to leave, job dissatisfaction, burnout, and several nurse-perceived quality outcomes (Coetzee, Klopper, Ellis, & Aiken, 2013). Findings from the Chinese branch of the RN4CAST study showed that better hospital work environments and having a higher proportion of bachelor-prepared nurses were significantly associated with patients' willingness to recommend the hospital and patients rating hospitals highly. There was no association between higher patient-to-nurse ratios and outcomes of patient satisfaction (You et al., 2013).

A second major RN4CAST report underlined the association between the same nurse setting characteristics and hospital discharge data from 422,730 surgical patients. This study demonstrated that every one patient increase in patient to nurse ratios was associated with a 7% increase in deaths, while having a better educated nurse workforce is associated with fewer deaths. Every 10% increase in bachelor's degree nurses is associated with a 7% decline in mortality (Aiken et al., 2014)

A third multicountry RN4CAST analysis worth mentioning, especially because of its contribution to building theory, is work that includes the concept of missed nursing care. The aspect of missed care closely relates to Schubert's previously introduced concept of rationing of care. Another term that has been used interchangeably is nursing care left undone (Aiken et al., 2001). All three notions point to the omission of necessary nursing care. RN4CAST researchers theorized that missed nursing care underlies the relationship between the organizational context of nursing and nurse and patient outcomes. This idea was

modeled partly, showing that less care is being omitted in hospitals with a more favorable work environment, lower patient to nurse ratios and lower proportion of nurses frequently carrying out non-nursing tasks (Ausserhofer et al., 2013). The idea of including the amount of non-nursing tasks in this model is also new, although it had already been measured and reported in the original IHOS (Aiken et al., 2001). Findings on the effect of staffing and work environment on missed nursing care are in line with Schubert's findings and confirm U.S. (Kalisch, Tschannen, & Lee, 2011) and Lebanese (Kalisch, Doumit, Lee, & Zein, 2013) country reports on the antecedents of omission of nursing care that were reported around the same time.

In conclusion, although the evidence at a certain point in time stemmed predominantly from U.S. research – which limited the policy impact in Europe – it is no longer in dispute that findings linking nursing to patient outcomes are universal. The RN4CAST study quantified in a European setting the core logic of previous U.S. research linking nursing to patient outcomes. These study findings pave the way for a renewed discussion on the professional profile of nursing in Europe. The next section describes research objectives to further advance this field of research.

1.4 Research objectives

The objectives of this PhD study are in many ways an innovative extension of the RN4CAST study. This will be accomplished in four thematic chapters that transcend the boundaries of nursing research through the introduction of advanced statistical concepts. The first part of this PhD study is energized by a strong intent to bridge theory articulation and empirical testing: newly theorized concepts in nursing outcomes research are studied (Chapter 3); hypotheses are set at multiple systems levels (Chapter 4); and a deeper and more nuanced analysis of mere direct relations is explicated (Chapter 5). In the second part of this PhD study, it is evaluated how key concepts of nursing organization are similar or different in meaning and measurement across nurse characteristics and management level (Chapter 6). The extent to which respondents across groups perceive and interpret the content of survey items in the same way, is called measurement invariance (Horn & McArdle, 1992). The remainder of this section presents the specific research objectives of this PhD study. Coupled with these objectives are the multifaceted

RN4CAST study design and a number of advanced statistical methods. These are explained in detail in Chapter 2.

1.4.1 Bridging theory articulation and empirical testing: research objectives

Research objective 1: To describe variation between and within countries in nurses' reports on tasks below their skill level, taking into account nurse-level information of migratory status.

In Chapter 3, we use the RN4CAST study data to describe variability across countries and hospitals within countries in nurses' reports of having performed tasks below their skill level. As noted in the previous section, RN4CAST has evidenced that this factor is of importance in theories linking nursing to outcomes. Anecdotal evidence suggest that nurses migrating from developing to developed countries may be at greater risk of practicing below their knowledge and skill levels. We therefore take a particular focus on comparing reports from domestically trained and foreign trained on having performed tasks below their skill level. Using a three-level hierarchical model that accounts for the clustering of nurses within hospitals and within countries, we estimate the overall effect (i.e. over all countries) of nurses' migratory status on performing tasks below skill level. We evaluate the consistency of this effect across countries.

Research objective 2: To explore the effect of the nursing unit, hospital, and country level variability on the relationship between dimensions of nurses' work environment and dimensions of burnout.

The RN4CAST study presents an opportunity to study knowledge of variation (cf. Deming's SoPK) at several systems levels: countries, hospitals and nursing units. Our second research objective specifically relates to recommendations for studying relations across multiple hierarchies as suggested by SEIPS 2.0 (Holden et al., 2013). This is essential to design system level-specific interventions as recommended by Vincent and collaborators (1998). In clinical terms, the question asked in Chapter 4 is 'at which level of the system certain associations are most pronounced?'. Two studies that examined the effect of nurse staffing on patient mortality found an association at the nursing unit level but not at the hospital level (Sales et al., 2008; Van den Heede, Lesaffre, et al., 2009). Both argued that nurse staffing can be very dissimilar across nursing units within the same hospital. These findings resulted from separate hospital

and nursing unit level analyses. We will apply a multilevel approach to theory and research on the impact of the organization of nursing care on nurse wellbeing. We investigate the effect of the nursing unit, hospital, and country level variability on the relationship between three dimensions of nurses' work environment and three dimensions of burnout. The outcomes of interest, i.e. the three burnout dimensions, take place at the individual level. System-level work environment dimensions are conceptualized to influence these level-one outcomes. The final model is a three-variate four-level probit model in which we modeled the work environment and burnout dimensions jointly.

Research objective 3: To empirically test previously and newly theorized mediation and moderation effects in the relationship between nursing organization of care and patient outcomes.

In Chapter 5, we ask questions that focus on explaining why and how associations are observed, as suggested by Nursing Services Delivery Theory: "Which mechanisms of the system underlie the relations between explanatory and outcome variables, and at which level?", and "Which personal or system-level factors affect associations?". We hypothesize that lower hospital patient-to-nurse ratios, more positive nursing work environments and a lower proportion of nurses performing non-nursing tasks positively affect the amount of missed care, which in turn positively affects patients' experiences with care. In addition, we hypothesize that the effects of lower patient-to-nurse ratios and better nurse work environment on missed care decrease as nurse education levels increase. This implies a multilevel moderated mediation analysis.

1.4.2 Measurement invariance evaluation: research objectives

Research objective 4: To evaluate within-country measurement invariance of nurse work environment measures, and examine ways in which individual propensities explain possible non-invariance.

Chapter 6 involves an evaluation of measurement invariance for various individual propensities simultaneously, including nurses' function (manager, frontline worker) and language (French, Dutch) among others. This evaluation takes place in the form of a two-level multiple indicators multiple causes (MIMIC) model with covariates at the individual nurse and nursing unit level.

References

Aiken, L. H., Cimiotti, J. P., Sloane, D. M., Smith, H. L., Flynn, L., & Neff, D. F. (2011). Effects of Nurse Staffing and Nurse Education on Patient Deaths in Hospitals With Different Nurse Work Environments. Medical Care, 49(12), 1047–1053.

Aiken, L. H., Clarke, S. P., Cheung, R. B., Sloane, D. M., & Silber, J. H. (2003). Educational levels of hospital nurses and surgical patient mortality. JAMA : The Journal of the American Medical Association, 290(12), 1617–23.

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. The Journal of Nursing Administration, 38, 223–229.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., Busse, R., Clarke, H., Giovannetti, P., Hunt, J., Rafferty, A. M., & Shamian, J. (2001). Nurses' reports on hospital care in five countries. Health Affairs (Project Hope), 20(3), 43–53.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA : The Journal of the American Medical Association, 288(16), 1987–93.

Aiken, L. H., Clarke, S., & Sloane, D. M. (2002). Hospital staffing, organization, and quality of care: cross-national findings. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 14(1), 5–13.

Aiken, L. H., Sermeus, W., Van den Heede, K., Sloane, D. M., Busse, R., McKee, M., Bruyneel, L., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Tishelman, C., Scott, A., Brzostek, T., Kinnunen, J., Schwendimann, R., Heinen, M., Zikos, D., Sjetne, I. S., Smith, H. L., & 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.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E., McHugh, M. D., Moreno-Casbas, M. T., Rafferty, A. M., Schwendimann, R., Scott, P. A., Tishelman, C., van Achterberg, T., & Sermeus, W. (2014). Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. Lancet, 383(9931), 1824–30.

Aiken, L. H., Sloane, D. M., Cimiotti, J. P., Clarke, S. P., Flynn, L., Seago, J. A., Spetz, J., & Smith, H. L. (2010). Implications of the California nurse staffing mandate for other states. Health Services Research, 45(4), 904–21.

Aiken, L. H., Sochalski, J., & Lake, E. T. (1997). Studying outcomes of organizational change in health services. Medical Care, 35(11 Suppl), NS6–18.

American Nurses Association. (2014a). Mandatory Overtime. Retrieved from http://www.nursingworld.org/MainMenuCategories/Policy-Advocacy/State/Legislative-Agenda-Reports/MandatoryOvertime

American Nurses Association. (2014b). Nurse staffing plans and ratios. Retrieved from http://www.nursingworld.org/MainMenuCategories/Policy-Advocacy/State/Legislative-Agenda-Reports/State-StaffingPlansRatios

Ausserhofer, D., Zander, B., Busse, R., Schubert, M., De Geest, S., Rafferty, A. M., Ball, J., Scott, A., Kinnunen, J., Heinen, M., Strømseng Sjetne, I., Moreno-Casbas, T., Kózka, M., Lindqvist, R., Diomidous, M., Bruyneel, L., Sermeus, W., Aiken, L. H., & Schwendimann, R. (2013). Prevalence, patterns and predictors of nursing care left undone in European hospitals: results from the multicountry cross-sectional RN4CAST study. BMJ Quality & Safety.

Bates, D. W., Larizgoitia, I., Prasopa-Plaizier, N., & Jha, A. K. (2009). Global priorities for patient safety research. BMJ (Clinical Research Ed.), 338(may14_1), b1775.

Brennan, T. A., Leape, L. L., Laird, N. M., Hebert, L., Localio, A. R., Lawthers, A. G., Newhouse, J. P., Weiler, P. C., & Hiatt, H. H. (1991). Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. The New England Journal of Medicine, 324(6), 370–6.

Bruyneel, L., Van den Heede, K., Diya, L., Aiken, L. H., & Sermeus, W. (2009). Predictive validity of the International Hospital Outcomes Study questionnaire: an RN4CAST pilot study. Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau, 41(2), 202–210.

Carayon, P., Schoofs Hundt, A., Karsh, B.-T., Gurses, A. P., Alvarado, C. J., Smith, M., & Flatley Brennan, P. (2006). Work system design for patient safety: the SEIPS model. Quality & Safety in Health Care, 15 Suppl 1, i50–8.

Carayon, P., Xie, A., & Kianfar, S. (2014). Human factors and ergonomics as a patient safety practice. BMJ Quality & Safety, 196–205.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases, 40(5), 373–383.

Cheung, R. B., & Aiken, L. H. (2008). Building an international nursing outcomes research agenda. Asian Nursing Research, 2(2), 69–73.

Clarke, S. P., & Aiken, L. H. (2008). An international hospital outcomes research agenda focused on nursing: lessons from a decade of collaboration. Journal of Clinical Nursing, 17(24), 3317–23.

Coetzee, S. K., Klopper, H. C., Ellis, S. M., & Aiken, L. H. (2013). A tale of two systems--nurses practice environment, well being, perceived quality of care and patient safety in private and public hospitals in South Africa: a questionnaire survey. International Journal of Nursing Studies, 50(2), 162–73.

Deming, W. E. (1993). The New Economics for Industry, Government, Education. Academy of Management Perspectives. MIT Press.

Donabedian, A. (1966). Evaluating the quality of medical care. The Milbank Quarterly, 44, 166–203.

Donabedian, A. (1988). The quality of care. How can it be assessed? JAMA : The Journal of the American Medical Association, 260(12), 1743–8.

Doran, D. I., Sidani, S., Keatings, M., & Doidge, D. (2002). An empirical test of the Nursing Role Effectiveness Model. Journal of Advanced Nursing, 38(1), 29–39.

Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. Medical Care, 36(1), 8–27.

Emanuel, L., Berwick, D. M., Conway, J., Combes, J., Hatlie, M., Lucian, L., Reason, J., Paul, S., Vincent, C., & Walton, M. (2008). What exactly is patient safety? In K. Henriksen, J. Battles, M. Keyles, & M. Grady (Eds.), Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment) (pp. 1–18). Agency fo Healthcare Research and Quality.

Estabrooks, C. A., Midodzi, W. K., Cummings, G. G., Ricker, K. L., & Giovannetti, P. (2005). The impact of hospital nursing characteristics on 30-day mortality. Nursing Research, 54(2), 74–84.

Friese, C. R., Lake, E. T., Aiken, L. H., Silber, J. H., & Sochalski, J. (2008). Hospital nurse practice environments and outcomes for surgical oncology patients. Health Services Research, 43(4), 1145–63.

Gunnarsdóttir, S., Clarke, S. P., Rafferty, A. M., & Nutbeam, D. (2009). Front-line management, staffing and nurse-doctor relationships as predictors of nurse and patient outcomes. a survey of Icelandic hospital nurses. International Journal of Nursing Studies, 46(7), 920–7.

Heinen, M. M., van Achterberg, T., Schwendimann, R., Zander, B., Matthews, A., Kózka, M., Ensio, A., Sjetne, I. S., Moreno Casbas, T., Ball, J., & Schoonhoven, L. (2013). Nurses' intention to leave their profession: a cross sectional observational study in 10 European countries. International Journal of Nursing Studies, 50(2), 174–84.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics, 56(11), 1669–86.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18(3-4), 117–44.

Hugonnet, S., Uçkay, I., & Pittet, D. (2007). Staffing level: a determinant of late-onset ventilator-associated pneumonia. Critical Care (London, England), 11(4), R80.

Institute of Medicine (U.S.) Committee on Quality of Health Care in America. (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, D.C.: National Academies Press.

Institute of Medicine (U.S.) Committee on the Adequacy of Nursing Staff in Hospitals and Nursing Homes. (1996). Nursing Staff in Hospitals and Nursing Homes: Is It Adequate?. (G. Wunderlich, F. Sloan, & C. Davis, Eds.). Washington, D.C.: National Academies Press.

Institute of Medicine (U.S.) Committee on the Work Environment for Nurses and Patient Safety. (2004). Keeping Patients Safe: Transforming the Work Environment of Nurses. (A. Page, Ed.). Washington, D.C.: National Academies Press.

International Ergonomics Association. (2014). What is ergonomics: definition and domains of ergonomics. Retrieved from http://www.iea.cc/whats/index.html

Irvine, D., Sidani, S., & Hall, L. M. (1998). Linking outcomes to nurses' roles in health care. Nursing Economics, 16(2), 58–64, 87.

Kalisch, B. J., Doumit, M., Lee, K. H., & Zein, J. El. (2013). Missed nursing care, level of staffing, and job satisfaction: Lebanon versus the United States. The Journal of Nursing Administration, 43(5), 274–9.

Kalisch, B. J., Tschannen, D., & Lee, K. H. (2011). Do staffing levels predict missed nursing care? International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 23(3), 302–8.

Katz, D., & Kahn, R. L. (1978). The Social Psychology of Organizations (Chapter 13). In The Social Psychology of Organizations (pp. 397–425).

Kelly, L. A., McHugh, M. D., & Aiken, L. H. (2012). Nurse outcomes in Magnet® and non-Magnet hospitals. The Journal of Nursing Administration, 42(10 Suppl), S44–49.

Kohn, L. T., Corrigan, J. M., & Molla, S. (1999). To Err Is Human: Building a Safer Health System. Medicine (Vol. 126, p. 312).

Kutney-Lee, A., McHugh, M. D., Sloane, D. M., Cimiotti, J. P., Flynn, L., Neff, D. F., & Aiken, L. H. (2009). Nursing: a key to patient satisfaction. Health Affairs (Millwood), 28(4), w669–77.

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in Nursing & Health, 25, 176–188.

Lake, E. T. (2007). The nursing practice environment: measurement and evidence. Medical Care Research and Review, 64, 104S–22S.

Leape, L. L., & Berwick, D. M. (2005). Five years after To Err Is Human: what have we learned? JAMA : The Journal of the American Medical Association, 293, 2384–2390.

Manojlovich, M., & Laschinger, H. (2007). The Nursing Worklife Model: extending and refining a new theory. Journal of Nursing Management, 15(3), 256–63.

Manojlovich, M., & Laschinger, H. K. S. (2008). Application of the Nursing Worklife Model to the ICU setting. Critical Care Nursing Clinics of North America, 20(4), 481–7.

McHugh, M. D., Kelly, L. A., Smith, H. L., Wu, E. S., Vanak, J. M., & Aiken, L. H. (2013). Lower mortality in magnet hospitals. Medical Care, 51(5), 382–8.

Meyer, R. M., & O'Brien-Pallas, L. L. (2010). Nursing Services Delivery Theory: an open system approach. Journal of Advanced Nursing, 66(12), 2828–38.

Mitchell, P. H., Ferketich, S., & Jennings, B. M. (1998). Quality health outcomes model. American Academy of Nursing Expert Panel on Quality Health Care. Image--the Journal of Nursing Scholarship, 30(1), 43–6.

Montalvo, I. (2007). The National Database of Nursing Quality Indicators ® (NDNQI ®). The Online Journal of Issues in Nursing, 12(3).

National Quality Forum. (2004). National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set.

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), 1715–22.

Needleman, J., Buerhaus, P., Pankratz, V. S., Leibson, C. L., Stevens, S. R., & Harris, M. (2011). Nurse staffing and inpatient hospital mortality. The New England Journal of Medicine, 364(11), 1037–45.

Rafferty, A. M., Clarke, S. P., Coles, J., Ball, J., James, P., McKee, M., & Aiken, L. H. (2007). Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records. International Journal of Nursing Studies, 44(2), 175–82.

Reason, J. (2000). Human error: models and management. BMJ (Clinical Research Ed.), 320(7237), 768–70.

Sales, A., Sharp, N. D., Li, Y.-F., Lowy, E., Greiner, G., Liu, C.-F., Alt-White, A., Rick, C., Sochalski, J., Mitchell, P. H., Rosenthal, G., Stetler, C., Cournoyer, P., & Needleman, J. (2008). The association between nursing factors and patient mortality in the Veterans Health Administration: the view from the nursing unit level. Medical Care, 46(9), 938–45.

Schubert, M., Clarke, S. P., Aiken, L. H., & de Geest, S. (2012). Associations between rationing of nursing care and inpatient mortality in Swiss hospitals. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 24(3), 230–8.

Schubert, M., Glass, T. R., Clarke, S. P., Aiken, L. H., Schaffert-Witvliet, B., Sloane, D. M., & De Geest, S. (2008). Rationing of nursing care and its relationship to patient outcomes: the Swiss extension of the International Hospital Outcomes Study. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 20(4), 227–37.

Schubert, M., Glass, T. R., Clarke, S. P., Schaffert-Witvliet, B., & De Geest, S. (2007). Validation of the Basel Extent of Rationing of Nursing Care instrument. Nursing Research, 56(6), 416–24.

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Busse, R., Lindqvist, R., Scott, A. P., Bruyneel, L., Brzostek, T., Kinnunen, J., Schubert, M., Schoonhoven, L., & Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nursing, 10, 6.

Silber, J. H., Rosenbaum, P. R., & Ross, R. N. (1995). Comparing the Contributions of Groups of Predictors: Which Outcomes Vary with Hospital Rather than Patient Characteristics? Journal of the American Statistical Association, 90(429), 7–18.

Sochalski, J., & Aiken, L. H. (1999). Accounting for variation in hospital outcomes: a cross-national study. Health Affairs (Project Hope), 18(3), 256–9.

Sochalski, J., Aiken, L. H., & Fagin, C. M. (1997). Hospital restructuring in the United States, Canada, and Western Europe: an outcomes research agenda. Medical Care, 35(10 Suppl), OS13–25.

Stelfox, H. T., Palmisani, S., Scurlock, C., Orav, E. J., & Bates, D. W. (2006). The "To Err is Human" report and the patient safety literature. Quality & Safety in Health Care, 15(3), 174–8.

Tervo-Heikkinen, T., Kvist, T., Partanen, P., Vehviläinen-Julkunen, K., & Aalto, P. (2008). Patient satisfaction as a positive nursing outcome. Journal of Nursing Care Quality, 23(1), 58–65.

Tervo-Heikkinen, T., Partanen, P., Aalto, P., & Vehviläinen-Julkunen, K. (2008). Nurses' work environment and nursing outcomes: a survey study among Finnish university hospital registered nurses. International Journal of Nursing Practice, 14(5), 357–65.

Thomas, E. J., Studdert, D. M., Burstin, H. R., Orav, E. J., Zeena, T., Williams, E. J., Howard, K. M., Weiler, P. C., & Brennan, T. A. (2000). Incidence and types of adverse events and negligent care in Utah and Colorado. Medical Care, 38(3), 261–71.

Tourangeau, A. E., Doran, D. M., McGillis Hall, L., O'Brien Pallas, L., Pringle, D., Tu, J. V, & Cranley, L. A. (2007). Impact of hospital nursing care on 30-day mortality for acute medical patients. Journal of Advanced Nursing, 57(1), 32–44.

Vahey, D. C., Aiken, L. H., Sloane, D. M., Clarke, S. P., & Vargas, D. (2004). Nurse burnout and patient satisfaction. Medical Care, 42(2 Suppl), II57–66.

Van Bogaert, P., Clarke, S., Vermeyen, K., Meulemans, H., & Van de Heyning, P. (2009). Practice environments and their associations with nurse-reported outcomes in Belgian hospitals: development and preliminary validation of a Dutch adaptation of the Revised Nursing Work Index. International Journal of Nursing Studies, 46(1), 54–64.

Van den Heede, K., Lesaffre, E., Diya, L., Vleugels, A., Clarke, S. P., Aiken, L. H., & Sermeus, W. (2009). The relationship between inpatient cardiac surgery mortality and nurse numbers and educational level: analysis of administrative data. International Journal of Nursing Studies, 46(6), 796–803.

Van den Heede, K., Sermeus, W., Diya, L., Clarke, S. P., Lesaffre, E., Vleugels, A., & Aiken, L. H. (2009). Nurse staffing and patient outcomes in Belgian acute hospitals: cross-sectional analysis of administrative data. International Journal of Nursing Studies, 46(7), 928–39.

Vincent, C., Taylor-Adams, S., & Stanhope, N. (1998). Framework for analysing risk and safety in clinical medicine. BMJ (Clinical Research Ed.), 316(7138), 1154–7.

Wachter, R. M. (2012). Understanding Patient Safety (Second edi.). New York: McGraw-Hill Professional.

You, L., Aiken, L. H., Sloane, D. M., Liu, K., He, G., Hu, Y., Jiang, X., Li, X., Li, X., Liu, H., Shang, S., Kutney-Lee, A., & Sermeus, W. (2013). Hospital nursing, care quality, and patient satisfaction: cross-sectional surveys of nurses and patients in hospitals in China and Europe. International Journal of Nursing Studies, 50(2), 154–61.

2 STUDY DESIGN AND METHODOLOGICAL PRINCIPLES

The aim of this Chapter is to create a good understanding of the study data and statistical techniques that will be used to answer the objectives of this PhD study. The first section presents an overview of the RN4CAST study design. The second section gives a concise overview of the various statistical techniques that are applied in a multilevel analytic framework in this PhD study.

2.1 RN4CAST study design

RN4CAST was funded under the European Commission's Seventh Framework Programme (FP7) from 2009 to 2011. The study was initiated with the idea that effective nursing workforce strategies could enhance the performance of health care organizations and health (Sermeus et al., 2011). With firm evidence at hand from the U.S., RN4CAST used a pool of well-known and extensively-validated tools, supplemented with recent measures that are important for evaluating nurses' roles in patient care. The core battery of survey items that were found to be important in earlier research (Aiken, Clarke, Sloane, et al., 2002; Aiken et al., 2001, 2003, 2008; Kutney-Lee et al., 2009) served as starting point for RN4CAST.

The study was led by the Katholieke Universiteit Leuven (Belgium) and the University of Pennsylvania School of Nursing (U.S.). Protocol harmonization, instrument development, and data collection were carried out during the first half of the funding period. The second half was devoted to statistical analyses and communication of research findings. New study findings continue to be disseminated even though FP7 funding has come to an end.

2.1.1 Setting and sample

The RN4CAST research consortium proposed a cross-sectional study design to evaluate the impact of nursing resources on patient outcomes in hospitals. RN4CAST protocol harmonization is described in detail by Sermeus and colleagues (2011). The twelve participating European countries (Belgium, England, Finland, Germany, Greece, Ireland, the Netherlands, Norway, Poland, Spain, Sweden, and Switzerland) reflect the diversity of health systems in Europe, ensuring a rich perspective of nursing workforce issues from all angles and countries. Three partners outside Europe (China, Botswana and South Africa) provide additional perspectives. In each participating country in Europe at least 30 general acute hospitals were sampled as primary sampling units, although most countries have more and the Netherlands has fewer.

Next, in each of the selected hospitals, general medical and surgical nursing units (at least 2 per hospital) were randomly selected. As described in the next section, nurse and patient data were collected on these nursing units. The type of study data that emerged from RN4CAST are thus of a complex four-level nature as imposed by the naturally occurring dependencies as well as the study design: patients and nurses (level 1) are clustered in nursing units (level 2) within hospitals (level 3) within countries (level 4).

2.1.2 Instruments and measures

Data were obtained via nurse, patient and organizational surveys and via routinely collected hospital discharge data. The core variables are displayed in Table 2.1 and will be clarified in the next sections. All data sources can be linked via common identifiers at the country, hospital and nursing unit level. Each grantee organization in the participating countries received ethical approval at the institutional level to do nurse surveys, patient surveys and analyze administrative data for patient outcomes. Country level approval to acquire and analyze patient outcomes data was also obtained.

Table 2.1 Core variables collected in the RN4CAST study

Variable	Measure	Source
Patient outcomes		
Mortality	30-day in-hospital mortality	Patient discharge
Satisfaction	Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)	Patient survey
Nurse outcomes		
Burnout	Maslach Burnout Inventory	Nurse survey
Job Satisfaction	Single item	Nurse survey
Intent to Leave	Single item	Nurse survey
Nursing system delivery strategies		
Nurse Staffing	Average number of patients cared for during last shift	Nurse survey
Nurse Practice Environment	Practice Environment Scale of the Nursing Work Index (PES-NWI)	Nurse survey
Nurse education	Proportion of bachelor-prepared nurses	Nurse survey
Skill mix	Nurses to total nursing staff ratio	Nurse survey
Tasks below skill level	Investigator-developed	Nurse survey
Process of nursing care		
Necessary nursing tasks left undone	Investigator-developed	Nurse survey
Risk-adjustment		
Comorbidity	Based on Charlson Comorbidity Index	Patient discharge
Patient chacteristics	Age, sex, type of admission	Patient discharge
Patient chacteristics	Educational level, self-rated health	Patient survey
Nurse characteristics	Age, sex, years of experience, degree, migratory status	Nurse survey
Hospital characteristics	Bed size, technology, teaching status	Organizational survey

2.1.2.1 Nurse survey

The nurse survey (self-administered paper questionnaires) included all staff nurses involved in direct patient care activities. Nurses were defined in each country as those meeting the European Union definition of trained and licensed nurses according to directive 2005/35/EC. The nurse survey has two main purposes. The first is to measure, within and across countries, characteristics of the hospital nurse workforce, nurses' future employment intentions, and of nurses' perspectives on quantity and quality of care. The second is to allow the creation of systems level measures of staffing and working conditions for nurses through aggregation of responses from nurses working in each institution (or nursing unit).

Nurse education is created from the survey at the institution (or subspecialty) level by calculating the percentage of nurses with a bachelor degree or higher.

Measuring the complexity of organizational support for nursing care required the development of a rigorous instrument. A large number of instruments are available (Lake, 2007), but the Practice Environment Scale of the Nursing Work Index (PES-NWI) (Lake, 2002), is most frequently used worldwide. It is endorsed by the U.S. National Quality Forum (2004) in its national voluntary consensus standards for nursing-sensitive care. It was recently identified as one of the strongest instruments available to conduct international comparisons of organizational work factors among hospital nurses due to its robust content, construct, discriminant, and concurrent validity (Warshawsky & Havens, 2011). This instrument is derived from early U.S. research into Magnet Hospitals. These are accredited hospitals for being successful in attracting and retaining nurses (Kramer & Hafner, 1989; McClure, Poulin, Sovie, & Wandelt, 1983). This research produced the 65-item Nursing Work Index (NWI). For each item, nurses could indicate whether they found it important to their job satisfaction, whether they found it important to provide quality patient care, and whether it is present in their current job situation. The Revised Nursing Work Index (NWI-R) (Aiken & Patrician, 2000) comprised 57 items and only asked nurses for each item whether it is present in their current job. Four subscales were proposed: autonomy, control over the practice setting, nurse-physician relationships, and organizational support. The PES-NWI was derived from the NWI-R and includes 32 items for five subscales: nurse participation in hospital affairs (9 items), nursing foundations

for quality care (8 items), nurse manager ability, leadership, and support of nurses (4 items), staffing and resource adequacy (4 items), and collegial nurse-physician relations (7 items) (Lake, 2002). The nurse-physician relationship subscale is thus the only subscale that was proposed for both the NWI-R and PES-NWI.

Nurse staffing is created from nurses' reports on the number of patients cared for during the last shift, the total number and type of nursing personnel and the total number of patients. Average patient to nurse ratios are calculated by dividing the number of nurses by the number of patients on the unit. Skill mix is the proportion of nurses to all nursing personnel.

Nurses indicated for nine tasks below their skill level whether they had performed the following tasks never, sometimes, or often during their last shift: routine phlebotomy/blood draw for tests, transporting of patients within hospitals, performing non-nursing care, performing non-nursing services not available on off-hours, delivering and retrieving food trays, answering phones/clerical duties, arranging discharge referrals and transportation, obtaining supplies or equipment, and cleaning patient rooms and equipment.

Nurses' reports on necessary nursing tasks left undone are used to get an idea of the process of nursing care. These include the following: adequate patient surveillance, skin care, oral hygiene, pain management, treatments and procedures, administering medication on time, frequently changing the patient's position, comfort/talk with patients, educating patients and family, preparing patients and families for discharge, developing or updating nursing care plans/care pathways, planning care, adequately documenting patient care.

Nurse wellbeing was assessed by studying job satisfaction, intention-to-leave the hospital or the nursing profession, and burnout. Job satisfaction was measured with a single item measure with four response categories. For intention-to-leave, nurses were asked whether they intend to leave their current position within a year. Burnout was measured with the Maslach Burnout Inventory (MBI). This instrument measures three related components of work-related burnout: emotional exhaustion; depersonalization;

and lack of personal accomplishment; the authors do not recommend combining the three subscales into a single measure (Maslach & Jackson, 1981).

Detailed information on demographic characteristics of the nurse respondents that were obtained from the nurse survey included highest educational attainment, gender, age, years of experience, and migratory status.

The core battery of questions was translated into 10 primary languages (Dutch, German, Greek, French, Italian, Finnish, Norwegian, Swedish, Spanish, Polish) using the translation-back translation method. A translation manager was appointed to ascertain high standards of instrument translation that reduce item bias. Construct bias was reduced by assessing dissimilarity of constructs in the investigated countries through the application of content validity indexing procedures. In each country the quality of the translated instruments was assessed by a panel of 9-12 bilingual experts employing Content Validity Indexing (CVI) resulting in Scale-CVI scores between .67 and .95 for the nurse survey (Squires et al., 2013).

In Europe, a total of 33,659 nurses participated to the survey. Response rates by country varied from 39% in Finland to 97% in Poland. Table 2.2 also displays the number of participating hospitals and nursing units in each country.

Table 2.2 Number of participating hospitals, nursing units and nurses to the RN4CAST nurse survey

Country	Number of hospitals	Number of nursing units	Number of nurses	Response rate
Belgium	67	272	3,186	72.1 %
England	46	413	2,918	39.3 %
Finland	32	126	1,131	46.2 %
Germany	49	199	1,508	47.0 %
Greece	24	65	367	54.4 %
Ireland	30	112	1,406	56.4 %
Netherlands	28	131	2,217	68.7 %
Norway	35	238	3,752	56.9 %
Poland	30	119	2,605	97.3 %
Spain	33	281	2,804	84.0 %
Sweden	79	79	10,133	69.3 %
Switzerland	35	134	1,632	72.2 %
Total	488	2169	33,659	63.0%

All research objectives of this PhD study draw on these nurse survey data, except for Chapter 6 which involves a new data collection that was completed in 2013. The principles of this data collection

were identical to the RN4CAST study, but whereas RN4CAST only included staff nurses, this follow-up study also invited nurse managers to participate. Data were collected in one Belgian hospital entity with 118 patient care units. On 87 nursing units, French and Dutch speaking nursing unit managers (n=87, 75.0% response rate) and staff nurses (n=821, 70.8% response rate) completed survey questionnaires identical to those used in RN4CAST. It will be evaluated in Chapter 6 if differences in staff nurses' and managers' ratings of the nursing work environment, measured using the PES-NWI, represent true group differences, or are the consequence of non-invariant measures.

2.1.2.2 Patient survey

In eight out of twelve countries, a patient survey was conducted. Participating countries are Belgium, Finland, Germany, Greece, Ireland, Poland, Spain, and Switzerland. In some countries all the selected hospitals are included in the patient survey while in other countries the patient survey is only conducted in a selection of hospitals. A one-day census approach is used to select patients of the selected nursing units in the participating hospitals.

All patients were asked to complete the 27-item Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). Only patients judged by nurses to be incapable of completing the survey even with help were excluded. The HCAHPS is reported as a set of ten measures (6 summary measures, 2 single items and 2 global ratings) related to: communication with nurses and doctors, responsiveness of hospital staff, pain management, communication about medicines, discharge information, cleanliness and quietness of the hospital environment, overall rating of the hospital and willingness to recommend the hospital to friends and family. Demographic questions related to patients' education levels and self-rated health.

The core battery of questions was translated into 7 primary languages (Dutch, Finnish, French, German, Greek, Italian and Polish) again using the translation-back translation method and CVI techniques, resulting in Scale-CVI scores between .67 and .99 (Squires et al., 2012)

A total of 10,866 patients participated to the survey. The patient survey response rate for all European countries is over 70% (Table 2.3).

Table 2.3 Number of participating hospitals, nursing units and patients to the RN4CAST patient survey

Country	Number of hospitals	Number of nursing units	Number of patients	Response rate
Belgium	60	244	2,623	67.78%
Finland	32	128	1,947	51.89%
Germany	13	-[a]	262	67.18%
Greece	24	65	847	63.49%
Ireland	10	38	285	92.83%
Poland	30	119	4,136	93.66%
Spain	15	104	470	-[b]
Switzerland	34	127	997	68.38%
Total	218	825	10,866	71.36%

[a] Surveys were distributed at the hospital level, [b] The number of distributed surveys was not tracked

Patient survey data are used in combination with nurse survey data in Chapter 5, where it is evaluated which variables explain the association between nursing system delivery strategies and patient experiences with care.

2.1.2.3 Organizational survey

These data allow controlling the analyses for institutional differences. Hospital structural characteristics that are measured include teaching status (post graduate medical residents), bed size and technology level (services for open-heart surgery, organ transplantation, or both).

2.1.2.4 Patient discharge data

The nine countries and number of hospitals and patients for which discharge data are available are shown in Table 2.4. RN4CAST undertook an extensive process of mapping patient outcomes data across countries by major diagnoses, 17 co-morbidities (Charlson et al., 1987), and procedures in the case of surgery. The data below present the selection of general, orthopedic, and vascular surgery patients at least 50 years of age for whom 30-day in-hospital mortality (death within 30 days of hospital admission) is known. Other risk adjustment covariates collected in this study include age, sex and type of admission. Patient discharge data are not used in the PhD study.

Table 2.4 Number of hospitals and patients for which discharge data are available

Country	Number of hospitals	Number of patients
Belgium	95	88,078
England	30	78,045
Finland	25	27,867
Ireland	27	19,822
Netherlands	22	31,216
Norway	28	35,195
Spain	16	21,520
Sweden	62	80,800
Switzerland	31	40,187
Total	300	422,730

2.2 Methodological principles

All research objectives of this PhD study aim to exploit the multilevel RN4CAST study design. An important condition to empirically test multilevel research questions effectively is to have a comprehensive theory that is aligned with an appropriate statistical analysis (Corley & Gioia, 2011; Klein, Dansereau, & Hall, 1994). The absence of such alignment is a general concern in nursing health services research (Griffiths, 2009; Levy, Landerman, & Davis, 2011a; Mark, Hughes, & Jones, 2004; Mark, 2006; Julie Sochalski, 2004). Contrary to the level of theory, the statistical level of analysis in this field has been largely restricted to act as if the phenomena and relations occur at one level only. This seems to be a limitation in patient safety research in general. In his critical review of the systems approach within patient safety research, Waterson (2009) was surprised by the lack of studies performing multiple level analyses. He saw as one possible explanation the fact that these types of studies require the use of specialist statistical techniques, and called for patient safety researchers to urgently overcome this issue. A multiple levels systems perspective indeed brings with it some methodological challenges. Multiple levels induce correlation of subjects within levels. That is, explanatory or outcome variables of subjects within the same cluster of a certain level of the system can be expected to be correlated. This can be the result of interaction processes and dynamics, a shared collective climate, and exposure to homogeneous contextual constraints and common organizational features, events and processes (Klein et al., 1994). Neglecting in the analysis the correlation that the four-level RN4CAST data structure induces can be problematic since the fundamental independence assumption underlying many commonly used statistical techniques would be violated. Multilevel models on the other hand acknowledge this clustered structure.

The multilevel modeling techniques that are used in this PhD study are briefly explained below. Complete specification of the models (e.g. assumptions about normality) is given in the respective chapters.

2.2.1 A basic multilevel model

This section is largely based on the notions advanced by Goldstein and colleagues (Goldstein, Browne, & Rasbash, 2002; Goldstein & McDonald, 1988; 1986, 1991).

Equation 2.1 shows a single-level linear regression model, which represents a relationship between an explanatory variable *x* and an outcome *y*:

$$y_n = \beta_0 + \beta_1 x_n + e_n, \tag{2.1}$$

where *n* is the index for nurses. β is a regression parameter. β_0 is the intercept, which equals the mean of *y* when *x* equals zero. β_1 is the slope, which is the coefficient for the explanatory variable *x*, showing how the outcome variable *y* changes with *x*. β_0 and β_1 represent the fixed part of the model. e_n is the random part and is also called the residual. When the model is correctly specified its mean equals and one observes no patterning of the residuals in a *x*-*y* plot of the residuals versus *x*. Often it is assumed that there is constant variability, i.e. e_n has the same variance for different values of *x* and *y*. It is also often assumed that the residual has a normal distribution, and hence this holds also for the outcome.

Next, suppose that nurses are clustered within hospitals. This is presented as follows:

$$y_{nh} = \beta_{0h} + \beta_1 x_{nh} + e_{nh}, \tag{2.2}$$

where *h* is the index for hospitals. Let y_{nh} be a normally distributed response, representing the *n*th nurse within the *h*th hospital. There now is a differential effect associated with the hospital. In other words, for any given value of *x*, the predicted level of *y* differs between the hospitals. This is called level 2 variation. Here, the regression coefficient remains constant. Because of this differential effect, intercept β_0 in Equation 2.2 is indexed and can be looked at as follows:

$$\beta_{0h} = \beta_0 + u_{0h}, \tag{2.3}$$

where the group-dependent intercept β_{0h} is split into an average intercept β_0 and the group dependent deviation u_{0h}.This intercept variation is assumed to be random. This equation is referred to as the macro-model of a random intercepts model. It specifies the between-hospital part of the model. Equation 2.2 is the micro-model, specifying the within-hospital part of the model, i.e. the nurse level. Substituting the macro-model into the micro-model yields:

$$y_{nh} = \beta_0 + \beta_1 x_{nh} + u_{0h} + e_{nh}, \tag{2.4}$$

where β_0 and β_1 represent the fixed part, u_{0h} is the random part at level 2, and e_{nh} is the random part at level 1.

Now suppose that the slopes are also allowed to be group-dependent. Equation 2.5 presents the macro-model of random slopes.

$$\beta_{1h} = \beta_1 + u_{1h}, \tag{2.5}$$

where the group-dependent slope β_{1h} is split into an average slope β_1 and the group dependent deviation u_{1h}. Substituting this macro-model into Equation 2.4 yields:

$$y_{nh} = \beta_0 + \beta_1 x_{nh} + u_{0h} + u_{1h} x_{nh} + e_{nh}, \tag{2.6}$$

where variation in *y* for nurses *n* in hospitals *h* is a function of *x*, a differential effect associated with hospitals, and a differential effect for *x* in relation to hospitals. β_0 and β_1 represent the fixed part , u_{0h} and u_{1h} present the random part at level 2, and e_{nh} is the random part at level 1.

2.2.2 A three-level random intercept model

In Chapter 3, we study the effect of individual nurses' migratory status on their reports on tasks below their skill level performed during their last shift, taking into account the clustering of nurses within hospitals within countries deploying a three-level random intercept model. This gives the following equation for the macro level:

$$\beta_{0hc} = \beta_0 + u_{0c} + u_{0hc} + e_{nhc}, \tag{2.7}$$

where *n* is the index for nurse (level 1), *h* is the index for hospital (level 2) and *c* is the index for country (level 3). u_{0c} is the random effect at the country level, an allowed-to-vary departure from the grand mean. u_{0hc} is the random effect at the hospital level, a departure from the country effect. e_{nhc} is the random effect at the nurse level, a departure from the hospital effect within a country.

2.2.3 A four-level model with covariates at multiple levels

In Chapter 4, the model is extended by formulating hypotheses on the effects of explanatory variables at multiple levels simultaneously. Thus looking into the question whether certain associations (interventions) might be strongest (most effective) at either the country, hospital or nursing unit level. For a model with one explanatory variable included at multiple levels, this yields:

$$y_{nuhc} = \beta_0 + \beta_1 x_{uhc} + \beta_2 x_{hc} u_{0j} + \beta_3 x_c + u_{0c} + u_{0hc} + u_{0uhc} + e_{nuhc,} \quad (2.8)$$

where *n* is the index for nurse (level 1), *u* is the index for nursing unit (level 2), *h* is the index for hospital (level 3) and *c* is the index for country (level 4). u_{0c}, u_{0hc} and u_{0uhc} can be interpreted similarly to the explanation given for Equation 2.7. Complete specification and multivariate extensions to this model are described in Chapter 4.

2.2.4 Multilevel mediation, moderation, and moderated mediation.

In Chapter 5, our research objectives relate to answering questions on the mechanisms underlying relations ('why'-question) and the conditions under which investments are likely to be most effective ('how'-question). In technical terms, these are called statistical tests of mediation and moderation respectively. Mediators illuminate the mechanisms that explain why associations can be observed. Moderators are contextual factors that define the conditions under which such associations can be predicted (Baron & Kenny, 1986; James & Brett, 1984). Although such effects have long been hypothesized, statistical tests have remained notably absent in nursing research, both in single-level and multilevel analyses. Building on the single-level linear regression model (Equation 2.1), the equations for a single-level mediation model equal:

$$y_n = \beta_0 + \beta_c x_n + e_n$$
$$y_n = \beta_0 + \beta_{c'} x_n + \beta_b m_n + e_n, \quad (2.9)$$

where $\hat{\beta}_c - \hat{\beta}_{c'}$ equals the mediated effect as the extent to which the mediator m_n accounts for the relationship between the initial explanatory variable x_n and outcome variable y_n.

Alternatively, keeping the first part of Equation 2.9 as is, the mediated effect can be estimated from:

$$m_n = \beta_0 + \beta_a x_n + e_n,$$

and taking the product $\hat{\beta}_a \hat{\beta}_b$, where $\hat{\beta}_a \hat{\beta}_b = \hat{\beta}_c - \hat{\beta}_{c'}$.

The equation for a single-level moderation model equals:

$$y_n = \beta_0 + \beta_1 x_n + \beta_2 z_n + \beta_3 x_n z_n + e_n, \quad (2.10)$$

where β_1 is the coefficient for the explanatory variable *x*, β_2 is the coefficient for the moderator *x*, β_3 is the coefficient for the interaction term xz. β_1 shows how the outcome *y* changes with *x* when *z* equals zero. β_2 shows how *y* changes with *z* when *x* equals zero. β_3 shows whether there is moderation of the relationship between *x* and *y*.

Statistical techniques and practical examples of (multilevel) mediation, moderation and moderated mediation models have recently been extensively discussed in the organizational sciences (Bauer, Preacher, & Gil, 2006; Edwards & Lambert, 2007; Fairchild & Mackinnon, 2010; Krull & MacKinnon, 2001; Preacher, Rucker, & Hayes, 2007; Z. Zhang, Zyphur, & Preacher, 2009). Chapter 5 will describe multilevel moderated mediation in full detail. In this introductory chapter, multilevel mediation models are described. Krull and MacKinnon (2001) labeled the various structures that multilevel mediation models can take. Suppose for nurses *n* clustered in hospitals *h*: 1) a 1→1→1 model, in which the initial covariate (x_{nh}), the mediator (m_{nh}) and the outcome (y_{nh}) are all analyzed at the lowest level; 2) a 2→1→1 model, in which the initial covariate (x_h) is operationalized at a higher level, i.e. the hospital level; and 3) a 2→2→1 model, in which both the initial covariate (x_h) and the mediator (m_h) are at a higher level. In

Chapter 5 of this PhD study, a 2→2→1 model is used. Reformulating the single-level mediation model in Equation 2.9 to a 2→2→1 model yields the following micro-models and macro-models:

$$\begin{aligned} y_{nh} &= \beta_{0h} + e_{nh} & \beta_{0h} &= \beta_0 + \beta_c x_h + u_{oh} \\ y_{nh} &= \beta_{0h} + e_{nh} & \beta_{0h} &= \beta_0 + \beta_{c'} x_h + \beta_b m_h + u_{oh} \end{aligned} \tag{2.11}$$

The same logic applies to reformulating Equation 2.10:

$$m_h = \beta_0 + \beta_a x_h + e_h, \tag{2.12}$$

Using simulation models on survey research examples with clustered data, Krull and MacKinnon (2001) showed that single-level mediation models underestimate the standard error of the mediated effect. The standard errors from the multilevel mediation model were on the contrary generally accurate.

To clarify the 2→2→1 model, we give two concrete examples. First, in their study on transformational leadership and impact on team performance and individuals' job satisfaction, Braun and colleagues (2013) hypothesized a 2→2→1 model in which transformational leadership influences trust in the team, which in turn affects job satisfaction. Transformational leadership and trust in the team were measured at the individual level using multidimensional survey instruments, but were aggregated to the team level. Trust in the team did not mediate the relationship transformational leadership and job satisfaction. Second, Meunier et al. (2013) hypothesized that shared family risks impact a child's behavioral outcomes, and that this effect is mediated by differential parenting. Behavioral outcomes included aggression, attention, emotional problems and social relationships. All four outcomes were measured using multiple item research scales. Differential parenting was both self-reported as well as observed by the researchers. A cumulative risk index was used to asses family risk, and included the mothers' educational level, maternal depression, maternal history of physical and sexual abuse, organization of safety, and family type. Up to four children per family were included in a sample of 397 families. The association between child level variables (outcomes) and family level variables (family risks (initial covariate) and differential parenting (mediator)) was modeled using a 2→2→1 multilevel mediation analysis. Findings confirmed the hypothesis.

2.2.5 Multilevel structural equation models

Structural equation modeling (SEM) is 'a statistical method that includes the estimation of models with regressions among continuous latent variables or factors' (Muthén & Muthén, 2012). Latent variables 'generally correspond to hypothetical structures or factors, which are explanatory variables presumed to reflect a continuum that is not directly observable'. Observed variables on the other hand are data that are directly collected and entered into the data file. Observed variables are also called manifest variables (Kline, 2011). Mediation and moderation analyses are often viewed as a simple case of SEM where all variables are manifest variables. The general notation of SEM (Muthén & Asparouhov, 2008) expresses a measurement model (Equation 2.13) and a structural model (Equation 2.14) as explained by Preacher, Zyhpur and Zhang (2010):

$$\boldsymbol{Y}_n = \boldsymbol{\nu} + \Lambda\boldsymbol{\eta}_n + \mathbf{K}\mathbf{X}_{\boldsymbol{n}} + \boldsymbol{\varepsilon}_n, \tag{2.13}$$

where *n* is the index for nurses. $\boldsymbol{Y}_n$ is a p-dimensional vector of observed variables. $\boldsymbol{\nu}$ is a *p*-dimensional vector of variable intercepts. $\boldsymbol{\varepsilon}_n$ is a *p*-dimensional vector of error terms. Λ is a $p \times m$ loading matrix, where *m* is the number of random effects (latent variables). $\boldsymbol{\eta}_n$ is an $m \times 1$ vector of random effects. $\mathbf{K}$ is $p \times q$ matrix of slopes for the *q* exogenous covariates in $\mathbf{X}_{\boldsymbol{n}}$.

$$\boldsymbol{\eta}_n = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_n + \Gamma\mathbf{X}_{\boldsymbol{n}} + \boldsymbol{\xi}_n, \tag{2.14}$$

where $\boldsymbol{\alpha}$ is an $m \times 1$ vector of intercept terms. $\mathbf{B}$ is an $m \times m$ matrix of structural regression parameters. Γ is $m \times q$ matrix of slope parameters for exogenous covariates, and $\boldsymbol{\xi}_n$ is an *m*-dimensional vector of latent variable regression residuals.

Among survey researchers, there is a growing appreciation for multilevel SEM (Asparouhov & Muthén, 2012). Preacher, Zyhpur and Zhang (2010) described how SEM combines with multilevel models. In this example, variables and parameter matrices of Equations 2.13 and 2.14 are indexed with *h,* indicating the clustering of nurses in hospitals. In addition, the multilevel part is taken into account in the level 2 structural model:

$$\boldsymbol{\eta}_h = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\eta}_h + \boldsymbol{\gamma}\mathbf{X}_{\boldsymbol{h}} + \boldsymbol{\xi}_h \tag{2.15}$$

2.2.6 Evaluating measurement invariance

In this PhD study, we address the relevance of (multilevel) SEM models in survey research by studying measurement invariance. Measurement invariance pertains to the extent to which respondents across groups perceive and interpret the content of survey instrument items in the same way (Horn & McArdle, 1992). Measurement non-invariance could result in erroneous inferences, possibly serious enough to result in incorrect interpretation of the findings. Lack of measurement invariance is also referred to as differential item functioning (DIF). Not being cautious towards DIF could, therefore, lead to incorrect research conclusions and implications. Measurement invariance is predominantly studied in a structural equation modeling framework. By applying progressively more stringent cross-group equality constraints, researchers can study various hierarchical degrees of measurement invariance (Meredith, 1993). Three types of invariance are commonly examined by researchers in their analyses: configural, metric, and scalar invariance. Configural invariance pertains to identical factor structures across groups, but does not require invariant parameter estimates across groups. That type of invariance only indicates that the same concepts are measured across groups, yet scores are not comparable. Metric invariance refers to the equality of factor loadings across groups; thus, mean-corrected scores can subsequently be compared across groups. Scalar invariance, however, requires equality of intercepts (thresholds for ordered categorical indicator scores) across groups and allows for full score cross-group comparisons. By testing the equality of parameters of the measurement models for the respective comparison groups, researchers can study the comparability of measurements. Two analytical approaches are generally applied to address the issue of measurement in a structural equation modeling framework: multiple indicators multiple causes modeling (MIMIC) and multiple group confirmatory factor analysis (MGCFA). Both allow survey researchers to jointly study measurement invariance. Finch and French (2011) showed the importance of accounting for the hierarchical nature when applying a MIMIC model. From their simulation study, the authors concluded that in the case of data sampling using a hierarchical framework, researchers should consider this multilevel structure and also include variables that explain measurement non-invariance at both the within and between level. Actual applications of measurement invariance

evaluation in a multilevel analytical framework are scarce. Chapters 6 and 7 present a study of measurement invariance in a multilevel context using both MIMIC and MGCFA modeling techniques.

2.2.7 Bayesian statistics

Important to denote is that Bayesian instead of frequentist methods are used in three of four thematic chapters (Chapters 4 to 6). An introductory note on concepts of Bayesian analysis is given below. It is largely based on B.O. Muthèn & Asparouhov's (2012) groundbreaking work on Bayesian structural equation modeling.

Conventional frequentist statisticians view parameters as being fixed. Bayesian statisticians on the other hand view parameters as variables that can have a prior distribution. This is because, often before a study is conducted, researchers already have some prior notion of the value of the parameter. Ideally, this prior notion is based on accumulating evidence resulting from prior studies. Priors thus allow researchers to bring in information in the model not only from the data, but from a prior notion. The combination of data with a prior results in a posterior, which gives the Bayesian estimate. Important to note is that priors can be noninformative (also called diffuse) or informative. Non-informative priors have a large prior variance. Therefore the maximum likelihood estimate from a frequentist approach and the Bayesian estimate will be very similar. The stronger the prior information, the narrower the variance and the more likely it is that the Bayesian and frequentist approach produce different results. Confirmatory factor analysis, which belongs to the family of structural equation models, can be seen as classic example of using priors. This technique allows researchers to summarize many manifest items into a small number of latent variables. Every manifest item only loads to the factor or factors it was hypothesized to measure, and has zero loadings on the other factors. The idea of zero loadings on other factors represents a strong prior, with mean zero and variance zero. This assumption might not be realistic, and priors can be used to specify loadings on other factors by including some variation (Muthén & Asparouhov, 2012). The use of informative priors in confirmatory factor analysis will be demonstrated in Chapter 6, in which measurement invariance across nurse managers and staff nurses is evaluated. Estimation and model fit evaluation are repeatedly explained in Chapters 4 to 6.

The philosophy of a prior distribution originates from Thomas Bayes' theorem (1763):

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \tag{2.16}$$

where P(A) and P(B) are the probabilities of events A and B respectively. P(B|A) is the probability of B given A, and P(A|B) is the probability of A given B. The theorem is derived from probability theory in which $P(A,B) = P(A|B)P(B) = P(B|A)P(A)$. In modeling terminology, replacing A by data and B by parameter values, this yields (Muthén & Asparouhov, 2012):

$$[parameters|data] = \frac{[data|parameters] \times [parameters]}{data} \tag{2.17}$$

Given that $[data|parameters] = likelihood$ and $parameters = prior$, this yields:

$$[parameters|data] = \frac{likelihood \times prior}{data} \tag{2.18}$$

Finally, replacing $[parameters|data]$ with $posterior$, yields:

$$posterior \propto likelihood \times prior, \tag{2.19}$$

where $\propto$ means proportional to (Muthén & Asparouhov, 2012).

Criticasters of Bayesian analysis see subjectivity in determining the prior as one of the major drawbacks of this approach. In Chapter 6, where informative priors are used, we will discuss the impact of varying the prior findings on our study findings, as suggested by Muthén & Asparouhov (2012).

References

Aiken, L. H., Clarke, S. P., Cheung, R. B., Sloane, D. M., & Silber, J. H. (2003). Educational levels of hospital nurses and surgical patient mortality. *JAMA : The Journal of the American Medical Association*, *290*(12), 1617–23.

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. *The Journal of Nursing Administration*, *38*, 223–229.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., Busse, R., Clarke, H., … Shamian, J. (2001). Nurses' reports on hospital care in five countries. *Health Affairs (Project Hope)*, *20*(3), 43–53.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA : The Journal of the American Medical Association*, *288*(16), 1987–93.

Aiken, L. H., & Patrician, P. A. (2000). Measuring organizational traits of hospitals: the Revised Nursing Work Index. Nursing Research, 49(3), 146–53.

Asparouhov, T., & Muthén, B. O. (2012). Multiple group multilevel analysis.

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), 1173–82.

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), 142–63.

Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transaction, 53, 370–418.

Braun, S., Peus, C., Weisweiler, S., & Frey, D. (2013). Transformational leadership, job satisfaction, and team performance: A multilevel mediation model of trust. The Leadership Quarterly, 24(1), 270–283.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases, 40(5), 373–383.

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory buildking: what constitutes a theoretical contribution? Academy of Management Review, 36(1), 12–32.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. Psychological Methods, 12(1), 1–22.

Fairchild, A. J., & Mackinnon, D. P. (2010). A general model for testing mediation and moderation effects. Prevention Science, 10(2), 87–99.

Finch, W. H., & French, B. F. (2011). Estimation of MIMIC Model Parameters with Multilevel Data. Structural Equation Modeling: A Multidisciplinary Journal, 18(2), 229–252.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73, 43–56.

Goldstein, H. (1991). Multilevel modelling of survey data. The Statistician, 40(2), 235–244.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Multilevel modelling of medical data. Statistics in Medicine, 21(21), 3291–315.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. Psychometrika, 53, 455–467.

Griffiths, P. (2009). RN+RN=better care? What do we know about the association between the number of nurses and patient outcomes? International Journal of Nursing Studies, 46(10), 1289–90.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18(3-4), 117–44.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. Journal of Applied Psychology, 69(2), 307–321.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. Academy of Management Review, 19, 195–229.

Kline, R. B. (2011). Principles and practice of structural equation modeling. Structural Equation Modeling (Vol. 156, p. 427).

Kramer, M., & Hafner, L. P. (1989). Shared values: impact on staff nurse job satisfaction and perceived productivity. Nursing Research, 38(3), 172–7.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel Modeling of Individual and Group Level Mediated Effects. Multivariate Behavioral Research, 36(2), 249–277.

Kutney-Lee, A., McHugh, M. D., Sloane, D. M., Cimiotti, J. P., Flynn, L., Neff, D. F., & Aiken, L. H. (2009). Nursing: a key to patient satisfaction. Health Affairs (Millwood), 28(4), w669–77.

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in Nursing & Health, 25, 176–188.

Lake, E. T. (2007). The nursing practice environment: measurement and evidence. Medical Care Research and Review, 64, 104S–22S.

Levy, J. A., Landerman, L. R., & Davis, L. L. (2011). Advances in mediation analysis can facilitate nursing research. Nursing Research, 60(5), 333–9.

Mark, B. A. (2006). Methodological issues in nurse staffing research. Western Journal of Nursing Research, 28(6), 694–709.

Mark, B. A., Hughes, L. C., & Jones, C. B. (2004). The role of theory in improving patient safety and quality health care. Nursing Outlook, 52(1), 11–6.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. Journal of Organizational Behavior, 2(2), 99–113.

McClure, M. L., Poulin, M. A., Sovie, M. D., & Wandelt, M. A. (1983). Magnet hospitals: Attraction and retention of professional nurses. Kansas City: American Nurses Association.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika.

Meunier, J. C., Boyle, M., O'Connor, T. G., & Jenkins, J. M. (2013). Multilevel mediation: cumulative contextual risk, maternal differential treatment, and children's behavior within families. Child Development, 84(5), 1594–615.

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Modlenberghs (Eds.), Longitudinal data analysis (pp. 143–165). Chapman & Hall/CRC.

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. Psychological Methods, 17(3), 313–35.

Muthén, L. K., & Muthén, B. O. (2012). Mplus User's Guide. Seventh Edition. Losa Angeles, CA: Muthén & Muthén.

National Quality Forum. (2004). National Voluntary Consensus Standards for Nursing-Sensitive Care: An Initial Performance Measure Set.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. Multivariate Behavioral Research, 42(1), 185–227.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. Psychological Methods, 15(3), 209–33.

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Busse, R., Lindqvist, R., Scott, A. P., Bruyneel, L., Brzostek, T., Kinnunen, J., Schubert, M., Schoonhoven, L., & Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nursing, 10, 6.

Sochalski, J. (2004). Is more better?: the relationship between nurse staffing and the quality of nursing care in hospitals. Medical Care, 42(2 Suppl), II67–73.

Squires, A., Aiken, L. H., van den Heede, K., Sermeus, W., Bruyneel, L., Lindqvist, R., Schoonhoven, L., Stromseng, I., Busse, R., Brzostek, T., Ensio, A., Moreno-Casbas, M., Rafferty, A. M., Schubert, M., Zikos, D., & Matthews, A. (2013). A systematic survey instrument translation process for multi-country, comparative health workforce studies. International Journal of Nursing Studies, 50(2), 264–73.

Squires, A., Bruyneel, L., Aiken, L. H., Van den Heede, K., Brzostek, T., Busse, R., Ensio, A., Schubert, M., Zikos, D., & Sermeus, W. (2012). Cross-cultural evaluation of the relevance of the HCAHPS survey in five European countries. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 24(5), 470–5.

Warshawsky, N. E., & Havens, D. S. (2011). Global use of the Practice Environment Scale of the Nursing Work Index. Nursing Research, 60(1), 17–31.

Waterson, P. (2009). A critical review of the systems approach within patient safety research. Ergonomics, 52(10), 1185–95.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing Multilevel Mediation Using Hierarchical Linear Models - Problems and Solutions. Organizational Research Methods, 12(4), 695–719.

3 A MULTI-COUNTRY PERSPECTIVE ON NURSES' TASKS BELOW THEIR SKILL LEVEL: REPORTS FROM DOMESTICALLY TRAINED NURSES AND FOREIGN TRAINED NURSES FROM DEVELOPING COUNTRIES

This chapter is published as:

Luk Bruyneel, Baoyue Li, Linda Aiken, Koen Van den Heede and Walter Sermeus. 2013. A multi-country perspective on nurses' tasks below their skill level: reports from domestically trained nurses and foreign trained nurses from developing countries. International Journal of Nursing Studies 50(2):202-209

and

Luk Bruyneel, Baoyue Li, Linda Aiken, Koen Van den Heede and Walter Sermeus. 2013. A multi-country perspective on nurses' tasks below their skill level: reports from domestically trained nurses and foreign trained nurses from developing countries. In James Buchan, Matthias Wismar, Irene A. Glinos and Jeni Bremner. 2014. Health professional mobility in a changing Europe. World Health Organization

Abstract

Background: Several researchers concluded that the use of nurses' time and energy is often not optimized. Given widespread migration of nurses from developing to developed countries, it is important for human resource planning to know whether nursing education in developing countries is associated with more exaggerated patterns of inefficiency.

Objectives: First, to describe nurses' reports on tasks below their skill level. Second, to examine the association between nurses' migratory status (domestically trained nurse or foreign trained nurse from a developing country) and reports on these tasks.

Methods: The Registered Nurse Forecasting Study used a cross-sectional quantitative research design to gather data from 33,731 nurses in 486 hospitals in 12 European countries. Here, nurse-reported information on migratory status and tasks below their skill level performed during their last shift was used. Random effects models estimated the effect of nurses' migratory status on reports of these tasks.

Results: 832 nurses were trained in a developing country (2.5% of total sample). Across countries, a high proportion of both domestically trained and foreign trained nurses from developing countries reported having performed tasks below their skill level during their last shift. After adjusting for nurses' type of last shift worked, years of experience, and level of education, there remained a pronounced overall effect of being a foreign trained nurse from a developing country and an increase in reports of tasks below skill level performed during the last shift.

Conclusion: The findings suggest that there remains much room for improvement to optimize the use of nurses' time and energy. Special attention should be given to raising the professional level of practice of foreign trained nurses from developing countries. Further research is needed to understand the influence of professional practice standards, skill levels of foreign trained nurses from developing countries and values attached to these tasks resulting from previous work experiences in their home countries. This will allow us to better understand the conditions under which foreign trained nurses from developing countries can optimally contribute to professional nursing practice in developed country contexts.

3.1 Background

System-level interventions like increasing nurse staffing and creating superior work environments have been associated with improved patient safety outcomes and a higher degree of nurse wellbeing (Aiken et al., 2012; Kelly et al., 2012). Also central to the efficient structuring of nurses' work is optimizing the use of their time and effort. When asked about their last shift however, nurses across three countries (U.S., Canada, Germany) consistently reported high percentages of non-nursing tasks performed, including transporting of patients, delivering or retrieving of food trays, and performing of housekeeping activities. At the same time, they reported many nursing tasks that were necessary but left undone because they lacked the time to complete them (Aiken et al., 2001). Al-Kandari and Thomas (2009) used the list of non-nursing tasks from the study of Aiken and colleagues among 780 Kuwaiti nurses. Increased non-nursing task workload was positively correlated to incompletion of nursing activities. Two recent time-and motion studies found that nurses spent considerable amounts of time in non-nursing activities. A 36-hospital time-and-motion study found that activities considered by nurses to be 'waste' (waiting, looking, retrieving, and delivering) consumed 6.6% of reported time per 10-h shift (Hendrich, Chow, Skierczynski, & Lu, 2008). Another time-and-motion study showed that nurses spent 9.0% of their time during their last shift on non-nursing tasks, including replenishing charts and forms, tidying up rooms, making beds, answering phones, searching for people, gathering linen, and answering call bells (Desjardins, Cardinal, Belzile, & McCusker, 2008).

The employment of internationally trained nurses may suggest a shortage of nurses at the institution or national level. Thus it is particularly important to optimize the full scope of professional nursing practice in institutions that employ nurses educated in other countries. Studies have shown that migrant nurses sometimes experience discrimination by means of lower wage and less upward mobility, and may be employed as nursing aids rather than as nurses, which negatively impacts their wellbeing (Center for Health Workforce Studies, University of Albany, 2008; International Organization for Migration, 2010; Kline, 2003). Other research suggests that nurses trained abroad aspire to the same professional nursing practice standards common to their country of current employment (Flynn & Aiken, 2002). In light of the increasing international mobility of nurses, Humphries et al. (2009) finds the

evaluation of how migrant nurses' skills are utilized a prerequisite to incorporating nurse migration into workforce planning.

The twelve-country Registered Nurse Forecasting (RN4CAST) study measured and linked organizational features of nurses' work places to nurse wellbeing and patient outcomes to challenge assumptions underpinning previous nurse workforce planning efforts (Sermeus et al., 2011). The aim of this study is to determine whether there is a difference between domestically trained and foreign trained nurses from developing countries in nurses' reports on tasks below their skill level performed during their last shift. We consider the implications of our findings for human resources management.

3.2 Methods

3.2.1 Study design

The RN4CAST study favored a rigorous quantitative multi-country cross-sectional design on the basis of research methods used in a five-nation study of critical issues in nurse staffing and the impact on patient care (Aiken et al., 2001). Data were gathered from four data sources (nurse, patient and hospital profile surveys and routinely collected hospital discharge data). The design of the RN4CAST-study is described in detail by Sermeus et al. (2011). This analysis used nurse-reported information on migratory status and tasks below skill level performed during their last shift.

3.2.2 Ethical approval

Depending on national legislation, the study protocol was approved by either central ethical committees (e.g. nation or university) or local ethical committees (e.g. hospitals). All nurses received an information letter explaining the design of the study.

3.2.3 Study sample

A total of 486 hospitals were sampled as primary sampling units in twelve European countries, with at least 30 hospitals per country. In two countries, the selected hospitals represent all of the relevant institutions in the country (Ireland, Norway). In Belgium, Germany, the Netherlands, Switzerland, England and Spain, a stratified random selection (geographical location within the countries, hospital size, and

hospital type) was done. Additionally, the Belgian and German research teams also gave the opportunity for hospitals to participate on a voluntary basis. In Finland, Poland and Greece, hospitals were selected via purposive sampling (i.e. geographical spread, hospital size, hospital type). In Sweden, all nurses were approached via the Swedish Nursing Association, which covers about 85% of all nurses working in Sweden. Nurses were then asked to identify the hospital in which they work. In each of the selected hospitals at least two general medical and surgical nursing units were randomly selected from a master list of nursing units. All staff nurses involved in direct patient care activities served as informants on organization of nursing care, nurse wellbeing, patient safety and quality of care. Nurses were defined in each country as those meeting the European Union definition of trained and licensed nurses according to directive 2005/35/EC. The sample consists of 33,731 nurses (62% response rate) from Belgium (n = 3,186), England (n = 2,990), Finland (n = 1,131), Germany (n = 1,508), Greece (n = 367), Ireland (n = 1,406), the Netherlands (n = 2,217), Norway (n = 3,752), Poland (n = 2,605), Spain (n = 2,804), Sweden (n = 10,133), and Switzerland (n = 1,632).

3.2.4 Study measures

A self-administered questionnaire was distributed. Nurses were asked to indicate whether they had received their training in the country they are currently working in and if not, in which country they did receive their training. Based on the World Economic Outlook classification of countries (International Monetary Fund, 2010), nurses were categorized into domestically trained, foreign trained in a country with an emerging or developing economy (further referred to as foreign trained in a developing country), or foreign trained in a country with an advanced economy (further referred to as foreign trained in a developed country). The IMF list of emerging and developing economies (150 out of 184 countries) includes countries from all over the world. Some recent entrants to the European Union for example have remained classified as emerging economies (e.g. Latvia, Poland).

Within a series of questions about their last shift, nurses were asked to report on a list of tasks below their skill level whether they had performed these tasks never, sometimes, or often during their last shift. The following nine tasks were presented to nurses: routine phlebotomy/blood draw for tests,

transporting of patients within hospitals, performing non-nursing care, performing non-nursing services not available on off-hours, delivering and retrieving food trays, answering phones/clerical duties, arranging discharge referrals and transportation, obtaining supplies or equipment, and cleaning patient rooms and equipment.

Three types of variables were used to control for confounders: nurses' type of last shift worked (morning, evening, night), number of years worked as a nurse and level of education (bachelor degree or not).

3.2.5 Statistical analysis

We first described for each country the share of foreign trained nurses and the share of nurses from developing and developed countries. We provide detailed data on the country of origin. We also assess whether there are statistically significant differences between domestically trained nurses and foreign trained nurses from developing countries in reporting type of last shift worked, number of years worked, and level of education. Second, we described nurses' reports on the list of nine tasks performed during their last shift. Third, we compared reports on tasks performed by domestically trained nurses and foreign trained nurses from developing countries. For analytic purposes, we dichotomized nurses' responses as 'never performed' and 'sometimes/often performed'. A heat map (Sneath, 1957) was used to graphically compare these reports, with a system of color-coding where a dark grey square indicates that a higher proportion of foreign trained nurses from developing countries reported this task compared to domestically trained nurses (and light grey square vice versa). A composite measure of tasks performed during nurses' last shift (min = 0, max = 9) was calculated for each individual nurse by taking the sum of the nine dichotomized nursing tasks. This composite measure had a binomial distribution. The overall effect (i.e. over all countries) of nurses' migratory status on this composite measure was estimated using a two-level logistic random effects regression. The country effect was modeled as a fixed effect. The hospitals were treated as a random effect. We calculated the intraclass correlation coefficient at the hospital level as an indication of the degree of homogeneity. The analysis was adjusted for nurses' type of last shift worked, number of years worked as a nurse and level of education. We analyzed the consistency

of the overall effect by specifying interaction effects between the countries under study and migratory status. We also constructed a series of similar two-level random effects regression models to analyze the overall effect of migratory status on each task separately. Despite all efforts to get random effects models with interaction effects to converge, this proved to be hard for four out of nine tasks because of computational problems. The statistical software package issued a non-convergence warning, which indicates that the model may not fit the data well. Descriptive findings for these tasks showed repetitive high proportions of both domestically trained nurses and foreign trained nurses from developing countries indicating they had performed these tasks during their last shift. We repeated our analysis comparing nurses' reports on tasks never/sometimes performed and often performed, and found similar findings. We also compared the difference in tasks reported between domestically trained nurses and foreign trained nurses from a developed country and found no statistically significant differences. The data analysis for this paper was generated using SAS/STAT software, Version 9.3 of the SAS System for Windows. Copyright 2011 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

3.3 Findings

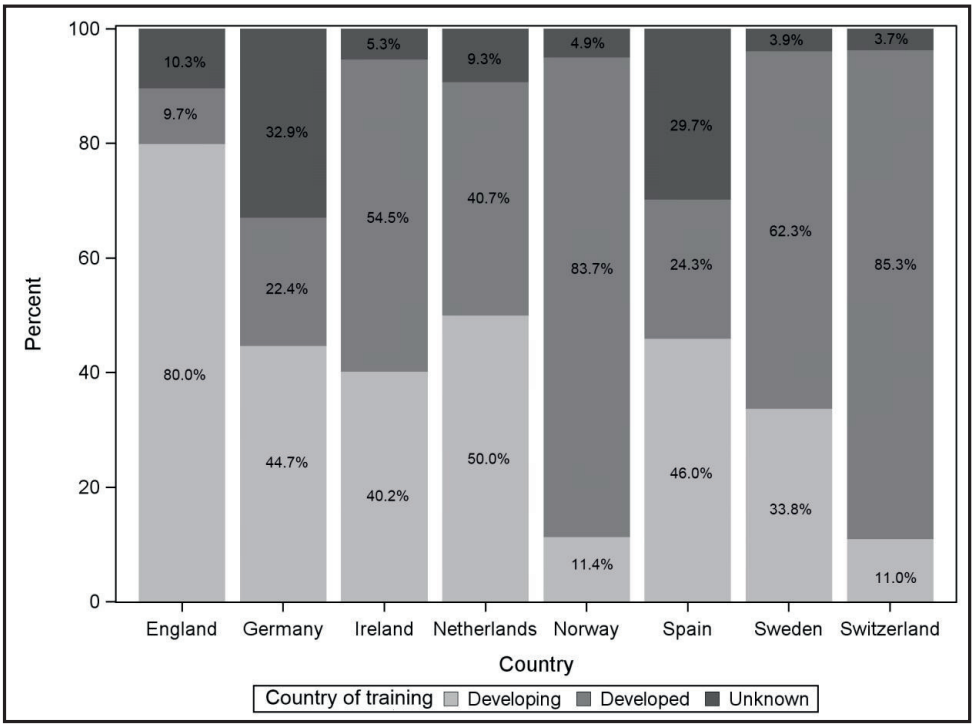
3.3.1 Foreign trained nurses

There were 2107 nurses (6.2% of the total sample) who indicated that they were trained in a different country from the one where they were currently employed, of which 832 were trained in a developing country (2.5% of total sample). There was large variation in the share of foreign-trained nurses between countries: Ireland (38.6%), Switzerland (22.1%), England (16.7%), Norway (5.5%), Germany (5.1%), Greece (5.1%), Belgium (3.1%), the Netherlands (2.4%), Sweden (2.3%), Spain (1.3%) and Finland (.9%). In Poland, all nurses that participated in the study were domestically trained nurses and in Greece there were no foreign-trained nurses from developing countries. The share of foreign-trained nurses varied considerably between hospitals in the top three countries with foreign-trained nurses, ranging from 16 to 56% (Ireland), 4 to 50% (Switzerland) and 1 to 52% (England). Countries with low numbers of foreign-trained nurses from developing countries (Finland, Greece, Poland) or high missing values on

country of training (Belgium) were dropped from further analysis, which resulted in a total of 813 foreign-trained nurses from developing countries remaining for further analysis. Figure 3.1 presents the large variation in the share of nurses from developing countries employed in the sample of eight remaining European countries. The percentage of foreign trained nurses trained in developing countries varied from 11% in Switzerland to as high as 80% in England.

In many countries, a large part of the share of foreign-trained nurses could be explained by mobility between neighboring countries or countries in the region: 112 of 354 foreign-trained nurses in Switzerland (31.6%) were trained in Germany, 107 (30.2%) in France and 41 (11.6%) in Italy. Nurses trained in developing countries now working in Switzerland included nurses from India (7; 2.0%), the Philippines (4; 1.1%) and Bosnia and Herzegovina (3; .85%), among others. In Sweden, 62 of the 231 foreign-trained nurses (26.8%) had obtained their training in Finland and 27 (11.7%) in Germany.

Figure 3.1 Foreign trained nurses: share of foreign trained nurses from developing and developed countries



The share of foreign-trained nurses from developing countries was ethnically very diverse, with most nurses trained in Bosnia and Herzegovina (15; 6.5%). In Spain, a different image emerged, with a large share of nurses trained in South American countries, mainly in Peru (8; 21.6%). Norway's largest group of foreign-trained nurses came from Sweden (53 of 231; 26.1%) with 31 (15.3%) from Australia and 31 (15.3%) from Denmark. Nurses from developing countries (200 in total) came from the Philippines (5; 2.5%), Lithuania (4; 2.0%) and Bosnia and Herzegovina (3; 1.5%), among others. In the Netherlands, after Belgium (7 of 54; 13.0%) and Germany (5; 9.3%), there was a substantial number of nurses from the former Dutch colonies of Suriname (12; 22.2%) and Indonesia (8; 14.8%). Of the 76 foreign-trained nurses in Germany, 13 (17.1%) came from Poland and 4 (5.3%) from Kazakhstan. In England, the main source countries were the Philippines (153 of 494; 31.0%) and India (117; 23.7%) but also nurses from sub-Saharan Africa (Ghana, Kenya, Nigeria, South Africa, Uganda, Zambia and Zimbabwe) accounted for a large proportion (78; 15.8%). As in England, the use of overseas recruiters is widespread in Ireland. Contrary to England, however, Ireland's share of nurses from developing countries was almost completely accounted for by nurses from India (111 of 531; 20.9%) and the Philippines (92; 17.3%) only. The share of European foreign-trained nurses in Ireland (53.2% of total) was almost exclusively made up of nurses who had received their training in the United Kingdom (51.5% of total).

In all eight countries, foreign-trained nurses from developing countries had more years of experience in working as a nurse than domestically trained nurses. These differences were statistically significant across all countries. Statistically significant differences were found for the level of education in England and Ireland, where the share of foreign-trained nurses from developing countries reporting that they had obtained a bachelor level degree in their home country was higher than the share for domestically trained nurses.

3.3.2 Nurses' reports on tasks performed during their last shift.

Across countries, a high proportion of nurses reported having sometimes or often performed tasks below their skill level during their last shift. Most reported tasks (country-weighted average) were answering telephones/clerical duties (97.4%), performing non-nursing care (90.1%) and obtaining supplies

or equipment (71.2%). There was large variability between countries in nurses' reports. For example in Spain, only 16.8% reported having cleaned patient rooms and equipment while in England this was 90% (Table 3.1).

Table 3.1 Nurses' reports of tasks below their skill level performed during their last shift

		Delivering and retrieving food trays	Performing non-nursing care	Arranging discharge referrals	Routine phlebotomy/blood draw for tests	Transport of patients within the hospital	Cleaning patient rooms and equipment	Filling in for non-nursing services not available on off-hours	Obtaining supplies or equipment	Answering phones, clerical duties
Belgium	Overall (n=3038)	83.8%	96.8%	76.9%	85.8%	69.9%	82.6%	47.6%	71.6%	97.9%
	Domestic (n=3021)	83.7%	96.9%	76.9%	85.8%	69.9%	82.7%	47.5%	71.5%	97.9%
	Developing (n=17)	100%	94.1%	76.5%	88.2%	64.7%	70.6%	58.8%	88.2%	100%
Switzerland	Overall (n=1274)	76.7%	97.2%	59.8%	74.3%	59.3%	57.0%	58.6%	65.8%	96.9%
	Domestic (n=1246)	76.6%	97.3%	59.7%	74.1%	58.9%	57.2%	58.6%	65.5%	96.9%
	Developing (n=28)	78.6%	92.9%	64.3%	85.2%	77.8%	46.4%	59.3%	80.0%	96.3%
Germany	Overall (n=1448)	82.4%	98.0%	74.1%	41.7%	70.9%	63.9%	65.8%	85.4%	98.7%
	Domestic (n=1414)	82.3%	97.9%	74.0%	41.5%	71.2%	63.8%	65.7%	85.1%	98.6%
	Developing (n=34)	85.3%	100%	76.5%	50.0%	55.9%	67.6%	70.6%	97.1%	100%
Spain	Overall (n=2746)	44.1%	91.0%	57.7%	86.2%	45.2%	16.8%	22.5%	72.8%	98.5%
	Domestic (n=2729)	44.0%	91.0%	57.5%	86.1%	45.2%	16.7%	22.5%	72.7%	98.4%
	Developing (n=17)	68.8%	88.2%	76.5%	94.1%	52.9%	37.5%	23.5%	87.5%	100%
Finland	Overall (n=1070)	63.3%	87.2%	41.6%	12.8%	31.7%	56.4%	72.1%	38.6%	97.8%
	Domestic (n=1068)	63.3%	87.2%	41.6%	12.8%	31.5%	56.5%	72.1%	38.6%	97.8%
	Developing (n=2)	50.0%	100%	.0%	.0%	100%	.0%	50.0%	50.0%	100%
Greece	Overall (n=335)[a]	37.7%	77.2%	79.4%	93.8%	63.6%	64.9%	65.5%	86.2%	94.8%
Ireland	Overall (n=1061)	64.2%	95.2%	80.7%	28.5%	67.5%	81.6%	69.2%	84.3%	99.1%
	Domestic (n=847)	58.7%	94.9%	79.0%	26.3%	64.1%	78.9%	70.0%	85.1%	99.1%
	Developing (n=214)	86.3%	96.7%	87.5%	37.1%	81.0%	92.8%	65.9%	81.1%	99.5%
Netherlands	Overall (n=2180)	57.4%	93.2%	76.2%	29.8%	68.5%	63.2%	40.7%	49.8%	98.7%
	Domestic (n=2153)	57.1%	93.1%	76.0%	29.7%	68.3%	63.0%	40.5%	49.3%	98.7%
	Developing (n=27)	85.2%	96.3%	85.2%	42.3%	81.5%	77.8%	59.3%	84.6%	96.3%
Norway	Overall (n=3516)	78.8%	71.5%	64.3%	39.9%	42.6%	68.1%	59.1%	63.9%	97.4%
	Domestic (n=3493)	78.8%	71.5%	64.2%	39.7	42.6%	68.1%	59.0%	63.7%	97.4%
	Developing (n=23)	82.6%	78.3%	69.6%	69.6%	34.8%	73.9%	65.2%	87.0%	100%
Poland	Overall (n=2593)[a]	75.0%	94.1%	59.0%	97.7%	90.4%	85.7%	62.6%	70.0%	97.9%
Sweden	Overall (n=9913)	64.5%	84.2%	58.1%	79.5%	53.4%	69.5%	37.7%	81.3%	94.6%
	Domestic (n=9837)	64.4%	84.1%	58.0%	79.4%	53.3%	69.4%	37.6%	81.2%	94.6%
	Developing (n=76)	75.7%	98.6%	63.0%	91.9%	63.0%	82.4%	49.3%	89.2%	94.7%
England	Overall (n=2866)	66.7%	96.0%	83.1%	54.4%	60.7%	90.0%	63.2%	85.5%	99.7%
	Domestic (n=2472)	63.6%	96.0%	82.1%	51.8%	57.9%	89.0%	62.1%	85.2%	99.7%
	Developing (n=394)	86.5%	95.6%	89.1%	71.0%	78.3%	96.4%	70.3%	87.8%	99.7%

[a] All domestic trained

Nurses' responses were dichotomized as 'never performed' and 'sometimes/often performed'. Migratory status was based on the World Economic Outlook classification of countries (International Monetary Fund, 2010).

3.3.3 Comparison of reports from domestically trained nurses and nurses trained in developing countries

The heat map (Figure 3.2) shows that in 62 out of 72 cases, higher percentages of nurses from developing countries reported they performed the nine tasks compared with domestically trained nurses (Table 3.1 for detailed findings). Findings were consistent between hospitals and for nurses from the same developing country working in the different countries under study here. For example, 25 English trusts had a total of 153 Philippines employed. In 24 out of 25 trusts, Philippine-trained nurses more often reported that they had delivered and retrieved food trays during their last shift compared with domestically trained nurses. This was also the case in 19 out of 20 Irish hospitals where Philippine-trained nurses were working.

The intraclass correlation coefficient for the nine items varied from .08 to .35, and was .21 for the composite measure, justifying the need for specifying a multilevel model. Table 3.2 shows that, after adjusting for last shift worked, years of experience and level of education, there remained a pronounced overall effect of being a foreign-trained nurse from a developing country and having an increase in reports of tasks performed during the last shift. This overall effect was found for the model testing the association between nurses' migratory status and the composite measure of tasks performed during the last shift. The interaction effect for this analysis was not significant. The series of models to analyze the overall effect of migratory status on each task separately showed that for eight out of nine tasks there was an overall effect of being a foreign-trained nurse from a developing country and an increase in reporting those tasks. Being a foreign-trained nurse from a developing country was a significant predictor of all five tasks for which an interaction effect was specified. The interaction effect was non-significant for three tasks (arranging discharge referrals, routine phlebotomy/blood draw for tests, cleaning patient rooms and equipment). For 'delivering and retrieving food trays' and 'obtaining supplies or equipment', the interaction effect was significant. For three out of four tasks for which no interaction effect could be specified, being a foreign-trained nurse from a developing country was a significant predictor (performing non-nursing care, transport of patients within the hospital, filling in for non-nursing services not available at off-hours). Migratory status failed to predict the task of 'answering telephones, clerical duties', for

which in each country at least 90% of both domestically trained nurses and foreign-trained nurses reported they had performed this task during their last shift.

Figure 3.2 Nurses' reports of tasks below their skill level performed during their last shift. Heat map comparing reports from domestically trained nurses and nurses trained in a developing country

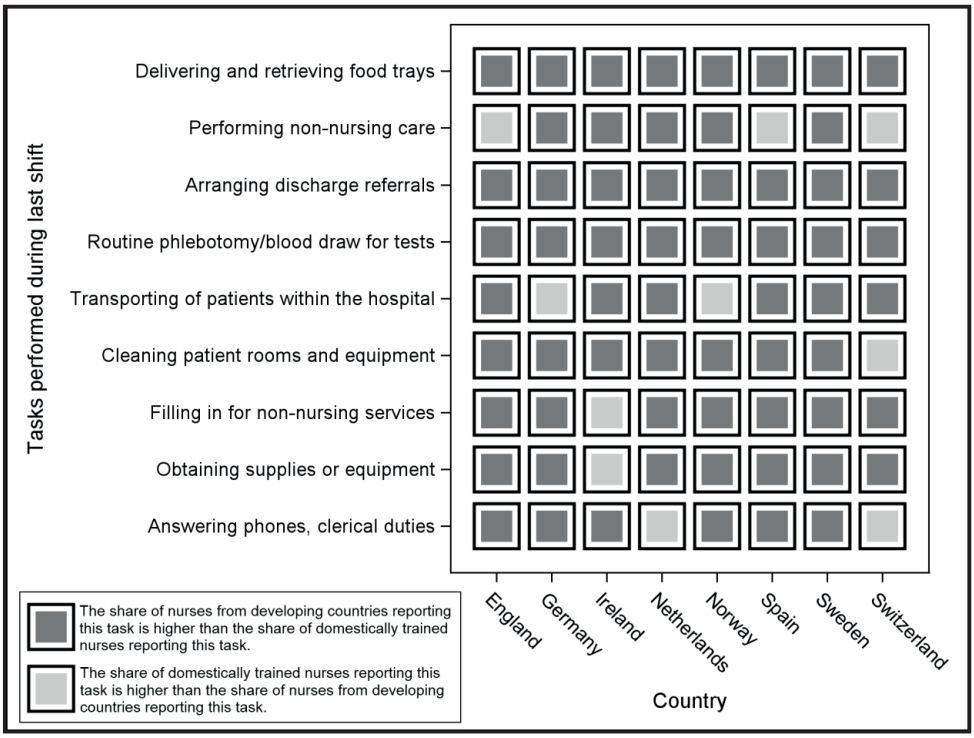


Table 3.2 Logistic random effects models estimating the overall effect of nurses' migratory status (trained in a developing country versus domestically trained) across eight countries on task below skill level performed during nurses' last shift

Tasks performed during nurses' last shift [a]	Estimate	Odds Ratio (95% CI)	p-value
Composite measure of nine nursing tasks	.74	2.10 (1.68-2.61)	<.0001
Delivering and retrieving food trays [b]	1.65	5.21 (4.04-6.72)	<.0001
Performing non-nursing care [c]	.53	1.70 (1.13-2.56)	.014
Arranging discharge referrals [b]	.89	2.44 (1.92-3.08)	<.0001
Routine phlebotomy/blood draw for tests [b]	.90	2.46 (1.91-3.17)	<.0001
Transport of patients within the hospital [c]	.73	2.08 (1.71-2.52)	<.0001
Cleaning patient rooms and equipment [b]	.64	1.90 (1.44-2.50)	<.0001
Filling in for non-nursing services not available on off-hours [c]	.19	1.21 (.99-1.47)	.048
Obtaining supplies or equipment [b]	.30	1.35 (1.03-1.78)	.033
Answering phones, clerical duties [c]	.53	1.70 (.70-4.10)	.235

[a] Model adjusted for reported last shift worked (morning, evening, night); number of years worked as a nurse and degree obtained (bachelor degree or not); trained in a developing country versus domestically trained based on the World Economic Outlook classification of countries (International Monetary Fund, 2010); eight countries covered were England, Germany, Ireland, the Netherlands, Norway, Spain, Sweden, Switzerland; [b] Interaction effect specified; [c] No interaction effect specified because of computational problems.

3.4 Discussion

This study documented high proportions of nurses across twelve countries indicating they had performed tasks below their skill level during their last shift. These findings support the previous studies of Aiken et al. (2001), Desjardins et al. (2008) and Hendrich et al. (2008) in which nurses reported much time spend on non-nursing tasks or much time wasted during their last shift.

Findings also revealed that, while a high share of all nurses reported having performed tasks below their skill level during their last shift, being a foreign trained nurse from a developing country was a significant predictor of performing tasks below skill level. These findings resulted from a two-level logistic random effects regression model testing the overall effect of migratory status on a composite measure of tasks performed, and persisted for a series of two-level random effects regression models to analyze the overall effect of migratory status on each task separately. The consistency in results across countries and hospitals makes these findings compelling.

In 2010, the World Health Assembly adopted the WHO Global Code of Practice on the International Recruitment of Health Personnel (World Health Organization, 2010). The ambition of this first code global in scope is for WHO Member States to refrain from the active recruitment of health personnel from developing countries facing critical shortages of health workers. The code also emphasizes the importance of equal treatment for migrant health workers and the domestically trained health workforce.

The RN4CAST data provided an opportunity to contribute to our understanding of this limited topic of research. The mix of countries participating in this study reflects the diversity of health systems in Europe, ensuring a rich perspective of nursing workforce issues from all angles. Robust statistical techniques were used to analyze the differences among domestically trained nurses and foreign-trained nurses from developing countries. Several limitations, however, warrant consideration. The proportion of foreign trained nurses from developing or from developed countries corresponded closely to that observed by the Organisation for Economic Cooperation and Development. Several limitations, however, warrant consideration. First, the measure of migratory status used may not have captured adequately the nationality of the nurses since only the country of training was known. Second, the proportion of foreign

trained nurses to the total sample was rather small. From the twelve countries under study, we had to drop from our analysis three countries with low numbers of foreign trained nurses from developing countries (Finland, Greece, Poland) and one country with high missing values on the variable of country of training. Third, the list of items on tasks below nurses' skill level was investigator developed (Aiken et al., 2001). However, it has shown to have predictive validity to the incompletion of nursing activities (Al-Kandari & Thomas, 2009). Also, other authors used similar items to describe waste (Hendrich et al., 2008) or non-nursing tasks (Desjardins et al., 2008). An early work measurement study from Connor (1961) already identified activities such as housekeeping and dietary tray delivery as non-nursing activities. It are exactly these tasks that could be delegated to non-nursing personnel (O'Brien-Pallas et al., 2004). It is however conceivable that some tasks surveyed here in certain situations of care were indeed nursing tasks. Also, it is possible that not all tasks below nurses' skill level may have been captured adequately, since the response was limited to a list of nine tasks. Last, in this multi-country European context, the context in which nurses performed these tasks can be very diverse. The influence of professional practice standards, skill levels of foreign-trained nurses from developing countries and values attached to these tasks resulting from previous work experiences in their home countries was unknown. It was not known, for example, whether foreign-trained nurses from developing countries were more likely than domestically trained nurses to be assigned to perform tasks below their skill level or whether foreign-trained nurses were more task oriented and brought the customs and roles of nursing from their developing country backgrounds into developed countries, and thus were more prone to voluntarily take on tasks below their skill level. The differences found between reports from domestically trained nurses and foreign-trained nurses were, however, not attributable to a lower level of education or fewer years of experience. To the contrary, in each country the foreign-trained nurses from developing countries had significantly more experience in working as a nurse than the domestically trained nurses. However, it was not known how long they had been working as a nurse in their destination country.

3.5 Conclusion

The findings suggest that there remains much room for improvement to optimize the use of nurses' time and energy. Human resources management should give more attention to professional socialization

and life-long learning for nurses to improve their priority setting and time management as well as ensuing that non-nursing resources are designated to carry out tasks that do not require the unique training of professional nurses. Nurses from developing countries may be particularly in need of continuing education on professional nurse roles and responsibilities in complex healthcare settings. Further research is needed to understand the influence of professional practice standards, skill levels of foreign trained nurses from developing countries and values attached to these tasks resulting from previous work experiences in their home countries. This will allow us to better understand the conditions under which foreign trained nurses from developing countries performed these tasks and to support improved structuring their work.

References

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., Busse, R., Clarke, H., … Shamian, J. (2001). Nurses' reports on hospital care in five countries. Health Affairs (Project Hope), 20(3), 43–53.

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.

Al-Kandari, F., & Thomas, D. (2009). Factors contributing to nursing task incompletion as perceived by nurses working in Kuwait general hospitals. Journal of Clinical Nursing, 18(24), 3430–40.

Center for Health Workforce Studies, University of Albany. (2008). The hospital nursing workforce in New York: Findings from a survey of hospital registered nurses.

Connor, R. J. (1961). A work sampling study of variations in nursing work load. Hospitals, 35, 40–1.

Desjardins, F., Cardinal, L., Belzile, E., & McCusker, J. (2008). Reorganizing nursing work on surgical units: a time-and-motion study. Nursing Leadership (Toronto, Ont.), 21(3), 26–38.

Flynn, L., & Aiken, L. H. (2002). Does international nurse recruitment influence practice values in U.S. hospitals? Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau, 34(1), 67–73.

Hendrich, A., Chow, M. P., Skierczynski, B. A., & Lu, Z. (2008). A 36-hospital time and motion study: how do medical-surgical nurses spend their time? The Permanente Journal, 12(3), 25–34.

Humphries, N., Brugha, R., & McGee, H. (2009). Career Progression of Migrant Nurses in Ireland: Nurse Migration Project Policy Brief 5. Epidemiology and Public Health Medicine Reports. Retrieved from http://epubs.rcsi.ie/ephmrep/5

International Monetary Fund. (2010). World Economic Outlook: Rebalancing Growth. Washington, DC. Retrieved from https://www.imf.org/external/pubs/ft/weo/2010/01/

International Organization for Migration. (2010). IOM Migration Research Series No. 41. The Role of Migrant Care Workers in Ageing Societies; Report on Research Findings in the United Kingdom, Ireland, Canada and the United States.

Kelly, L. A., McHugh, M. D., & Aiken, L. H. (2012). Nurse outcomes in Magnet® and non-Magnet hospitals. The Journal of Nursing Administration, 42(10 Suppl), S44–49.

Kline, D. S. (2003). Push and pull factors in international nurse migration. Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau, 35(2), 107–11.

O'Brien-Pallas, L., Thomson, D., McGillis Hall, L., Pink, G., Kerr, M., Wang, S., … Meyers, R. (2004). Evidence-Based Standards for Measuring Nurse Staffing and Performance. Retrieved from http://www.hrhresourcecenter.org/node/1459

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., … Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nursing, 10, 6.

Sneath, P. H. A. (1957). The Application of Computers to Taxonomy. Journal of General Microbiology, 17(1), 201–226.

World Health Organization. (2010). Global Code of Practice on the International Recruitment of Health Personnel. World Health Organization. Retrieved from http://www.who.int/hrh/migration/code/full_text/en/

4 GROUP-LEVEL IMPACT OF WORK ENVIRONMENT DIMENSIONS ON BURNOUT EXPERIENCES AMONG NURSES: A MULTIVARIATE MULTILEVEL PROBIT MODEL

This chapter is published as:

Baoyue Li, Luk Bruyneel, Walter Sermeus, Koen Van den Heede, Kenan Matawie, Linda Aiken, and Emmanuel Lesaffre. 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

Abstract

Background: Nurses' work environments are associated with burnout experiences among nurses. The RN4CAST project provides data on these constructs within a four-level structure (nurse, nursing unit, hospital, and country), implying more complicated multilevel analysis strategies than have been used in previous efforts studying this relationship.

Objectives: First, to explore and investigate the effect of the nursing unit, hospital, and country level variability on the relationship between dimensions of nurses' work environment and dimensions of burnout. Second, to explore the significance of the nursing unit, hospital, and country level variability among the burnout dimensions.

Methods: Data from the RN4CAST project were available from a cross-sectional survey among 23,446 nurses in 2087 nursing units in 352 hospitals in 11 countries. Nurses reported on their work environment (managerial support for nursing, doctor-nurse collegial relations, and promotion of care quality) and burnout experiences (emotional exhaustion, depersonalization and personal accomplishment). Ecological measures of the nurse work environment dimensions were specified and combined with individual-level outcomes within a series of multilevel statistical models. The final model was a multivariate multilevel probit model in which the work environment and burnout dimensions are modeled jointly.

Results: Doctor-nurse collegial relations affected all burnout dimensions, but at the unit level only. For the dimension of promotion of care quality, the effect of the ecological exposure on burnout was pronounced at both the nursing unit and the hospital level for all three burnout dimensions. Findings for the dimensions of managerial support for nursing were ambiguous.

Conclusion: Nurse work environment dynamics are related to nurses' burnout experiences at both the nursing unit and the hospital level. This implies that both hospital-wide and unit-specific interventions should be considered to achieve excellent work environments. The correlation structure among the three burnout outcomes varies across countries, but is stable between hospitals within countries and between nursing units within hospitals.

4.1 Background

The context in which nursing outcomes research is undertaken is often multilevel in nature. Multilevel complexity can be caused by naturally occurring dependencies (e.g. nurses in hospitals) or imposed-by-design dependencies (multistage sampling). These complex structures imply an explicit multilevel analysis to take into account the correlated nature of the data. Several studies introduced the conceptual and statistical background in multilevel analysis for nursing research and portrayed examples of the application of two-level techniques for meta-analysis (Wu, 1997), confirmatory factor analysis (Gajewski, Boyle, Miller, Oberhelman, & Dunton, 2010) and regression analysis (Adewale et al., 2007; Cho, 2003; Park & Lake, 2005). They detailed how features of these two-level techniques overcome the fallacies of conventional single level models in the analysis of clustered data. These studies vastly contributed to illustrating the basics in multilevel modeling for patient and organizational outcomes research.

The current article takes the application of multilevel regression techniques in nursing research a step further by analyzing the association between nurses' work environment and burnout in a four-level data set (country, hospital, nursing unit, and nurse) resulting from the Registered Nurse Forecasting (RN4CAST) project. This multi-country nurse workforce study has provided a unique data set on organizational features of nursing care and measures of nurse wellbeing, patient satisfaction with care, and quality and safety of patient care (Sermeus et al., 2011). Understanding research problems in this data structure dictates more complicated multilevel analysis strategies than have been used in previous efforts. Nurse work environment and burnout, the constructs of interest studied in the current article, have been well researched previously. Large-scale studies have shown that nurses working in both post-industrial (Aiken et al., 2001; Hasselhorn, Tackenberg, & Müller, 2003) and developing countries (Poghosyan, Clarke, Finlayson, & Aiken, 2010) are susceptible to burnout. Burnout in turn impacts patient satisfaction with nursing care (Vahey et al., 2004) and plays a mediating role in nurses' reports on quality of care and adverse events, job dissatisfaction and turnover intentions (Leiter & Maslach, 2009; Spence Laschinger & Leiter, 2006; Van Bogaert, Meulemans, Clarke, Vermeyen, & Van de Heyning, 2009). The consequences of burnout thus potentially negatively affect nurses, patients, organizations and health

systems in general. Of interest is that the large majority of nurse researchers studying burnout have mainly focused on the emotional exhaustion dimension of the syndrome. This dimension refers to feelings of being overextended and depleted of one's emotional and physical resources, and has indeed been described by the world leading researchers in the field of burnout as the key aspect of burnout (Maslach & Jackson, 1981). However, they have also repeatedly emphasized the significance of the three-dimensional burnout model in that it "clearly places the individual strain experience within the social context of the workplace and involves the person's conception of both self and others" (Maslach, 1993). Measuring emotional exhaustion only "fails to capture the critical aspects of the relationships that people have with their work" (Maslach & Leiter, 2008). We therefore study in this article all three burnout dimensions: emotional exhaustion, depersonalization and personal accomplishment. Depersonalization refers to negative, callous, or excessively detached responses to various aspects of the job. Feelings of incompetence and a lack of achievement and productivity in work are captured by the personal accomplishment dimension (Maslach & Leiter, 2008). A point of departure for our line of research presented here is the well-documented evidence on the causes of burnout. Studies across countries worldwide found that modifiable dimensions of nurses' work environment and workload predict burnout rates among nurses (Bruyneel et al., 2009; Kelly et al., 2012; Nantsupawat et al., 2011). Such dimensions include staffing and resource adequacy, managerial support for nursing, nurse participation in hospital affairs, doctor–nurse collegial relations, and promotion of care quality. Nurses' reports on their work environment and burnout experience, that are both multidimensional constructs, provide an excellent opportunity to introduce nurse researchers to advanced multilevel regression analyses.

The aim of this study is two-fold. First, to explore and investigate the effect of the nursing unit, hospital, and country level variability on the relationship between dimensions of nurses' work environment and dimensions of burnout. Second, to explore the significance of the nursing unit, hospital, and country level variability among the burnout dimensions.

4.2 Methods

4.2.1 Data sources

The data used in this study come from the RN4CAST project, a three year (2009–2011) nurse workforce study funded by the Seventh Framework Programme of the European Union. For the RN4CAST project, research teams from across 12 countries used a multilevel observational design to determine how system-level features in the organization of nursing care (work environment, education, and workload) impact individual measures of nurse wellbeing (burnout, job satisfaction, and turnover) and patient safety outcomes and care satisfaction. The design of the RN4CAST project is described in detail by Sermeus et al. (2011). The relevant data for the current analysis include nurses' ratings of their work environment and reports on burnout experiences.

4.2.2 Ethical considerations

In all but one country, depending on national legislation, the study protocol was approved by either central ethical committees (e.g. nation or university) or local ethical committees (e.g. hospitals). In the Netherlands no ethical approval was required.

4.2.3 Study sample

A four-level hierarchical structure is the form of the sampling strategy used in the RN4CAST project. The study encompasses data from 33,731 nurses (level 1) in 2,089 nursing units (level 2) in 486 hospitals (level 3) in 12 countries (level 4). The participating countries are Belgium, England, Finland, Germany, Greece, Ireland, the Netherlands, Norway, Poland, Spain, Sweden, and Switzerland. A minimum of 30 hospitals participated in each country. In most of the countries, the selected hospitals either represented all hospitals in the country (Ireland and Norway) or were random samples of all general (nonspecialized) hospitals. In Finland, Poland and Greece, the purposive sampling was used which was based on the geographical spread, hospital size and hospital type. At least two general medical and surgical nursing units for each hospital were randomly selected, of which all nurses involved in direct patient care activities were invited to participate in the study. A different sampling strategy was used in Sweden that

nurses were selected via the Swedish Nursing Association, and the hospitals in which they work were then identified. The overall response rate of 62% compares favorably with rates seen in other nursing outcomes research studies and was for most countries consistently high across nursing units and hospitals. England (38.6%), Finland (46.2%) and Germany (41.6%) had lower response rates. Swedish data were excluded as no unit identifiers were available from the Swedish sampling design. The collected data have the characteristics of a strict hierarchical structure. First, lower level units are nested within one and only one unit at the next higher level. Second, lower level units present repeated samples of higher level units. Third, there was successive sampling from each level of the hierarchical population. Fourth, as can be expected, the sample size within higher level units was imbalanced, albeit there were a sufficient number of respondents for analysis in the sampled units.

4.2.4 Study measures

The nurse work environment was measured using the Practice Environment Scale of the Nursing Work Index (PES-NWI), an internationally validated organizational measure (Warshawsky & Havens, 2011) that reflects the multidimensionality of nurses' work environment. The PES-NWI operationalizes five dimensions that facilitate or constrain nursing practice. Nurses therefore score statements about the work environment on a four point Likert scale ('Totally agree' = 4, 'Agree' = 3, 'Not agree' = 2, 'Totally not agree' = 1). The five dimensions, example item and number of items of each dimension are: managerial support for nursing ('A nurse manager who is a good manager and leader'; 4 items), nurse participation in hospital affairs ('Career development/clinical ladder opportunity'; 8 items), doctor–nurse collegial relations ('Physicians respect nurses as professionals'; 7 items), staffing and resource adequacy ('Enough registered nurses on staff to provide quality patient care'; 4 items), and promotion of care quality ('Working with nurses who are clinically competent'; 9 items). The Pearson coefficient correlation matrix showed relatively higher correlations between the dimensions of staffing and resource adequacy and nurse participation in hospital affairs and other dimensions. This has the potential to adversely affect regression estimates. The multicollinearity checking showed a potential problem for the dimension nurse participation in hospital affairs. In addition, another paper by Kutney-Lee et al. (2009) used the same three environment dimensions. We therefore did not include these two dimensions for further analyses.

The response variable for this analysis is the multidimensional burnout phenomenon. Burnout is a syndrome of emotional exhaustion, depersonalization and reduced personal accomplishment that can occur among individuals who do 'people work' of some kind (Maslach & Jackson, 1981). We evaluated burnout using the 22-item Maslach Burnout Inventory (MBI) that has been extensively used to capture the three dimensions of burnout. On a seven point Likert scale, nurses assessed the frequency (ranging from never to every day) of burnout experiences.

Degrees of burnout are calculated separately for the dimensions of emotional exhaustion (9 items), depersonalization (5 items) and reduced personal accomplishment (8 items) by using the numerical cut-off points listed on Maslach and Jackson's (1986) scoring key. This key categorizes respondents into low, average and high ranges of experienced burnout for each dimension. We dichotomize respondents as experiencing high burnout or not, thus requiring a binary modeling. A probit regression was chosen in this study. About 10% of the nurses, at least one data value was missing on either the work environment or the burnout items. For the work environment dimensions, any missing data values were completed with the mean of the nonmissing data values. For the dimensions of burnout, missing data values were imputed using the multinomial distribution of frequencies per hospital. That is, each missing was replaced by a random value from the observed multinomial distribution in each hospital. After applying the missing data imputation strategies in R (version 2.13.0), the final data set contained 23,446 nurses coming from 2,087 nursing units, within 352 hospitals and 11 countries.

4.2.5 Statistical analysis

4.2.5.1 Intraclass correlation coefficient

The basic statistical prerequisite for the appropriate application of multilevel analyses includes clustered data with a positive intraclass correlation coefficient (ICC). A (true) positive ICC violates the independent observations assumption of ordinary least squares estimation, resulting in downwardly biased standard error estimates, overly large test statistics, and inflated type I error rates (Krull & MacKinnon, 2001). We estimated the ICC to get an idea of the degree of variation in the burnout dimensions that were explained at each level. For the four-level model, the ICCs are defined as follows:

$$ICC_c = \frac{\sigma_c^2}{\sigma_{all}^2}, ICC_h = \frac{\sigma_c^2 + \sigma_h^2}{\sigma_{all}^2}, ICC_u = \frac{\sigma_c^2 + \sigma_h^2 + \sigma_u^2}{\sigma_{all}^2}, \quad (4.1)$$

with σ_{all}^2 the stum of all variances, i.e. $\sigma_{all}^2 = \sigma_c^2 + \sigma_h^2 + \sigma_u^2 + \sigma_n^2$. The subscripts *c*, *h*, *u*, *n* represent country, hospital, nursing unit and nurse levels, respectively. For the probit model, the lowest level variance, i.e. σ_n^2, is set to be one using the latent variable strategy. Regarding the interpretation of the intra-class correlation, we note that ICC_c is the correlation of two nurses' responses within the same country (different hospitals and nursing units), and ICC_h is the correlation of two nurses' responses within the same hospital (same country but different nursing units), while ICC_u is the correlation of two nurses' responses within the same nursing unit (same country and hospital). The higher the ICC scores, the higher the degree of homogeneity among nurses clusters. In order to get the partitioned proportion of the total variation into each level, we subtracted the higher level ICC from the lower level ICC, that is:

$$p_c = \frac{\sigma_c^2}{\sigma_{all}^2}, p_h = \frac{\sigma_h^2}{\sigma_{all}^2}, p_u = \frac{\sigma_u^2}{\sigma_{all}^2}, p_n = \frac{\sigma_n^2}{\sigma_{all}^2}. \quad (4.2)$$

4.2.5.2 Model specification

The outcomes of interest, i.e. the three burnout subscales, take place at the individual level. These are called level-one outcomes. The work environment dimensions were conceptualized to influence these level-one outcomes at higher organizational levels. In multilevel jargon, such variables are called ecological variables. We will continue to use this term and the term work environment dimension throughout this paper. Contrary to previous nurse workforce studies, we will avoid naming these variables as environmental variables. Ecological variables were calculated as the average of the item responses of nurses within nursing units, hospitals and countries. In a multilevel context, the effect of a covariate can be decomposed into effects on different levels, which is recommended by Neuhaus and Kalbfleisch (1998). This decomposition allows us to learn the difference of the relationship at each level, which renders the modeling more flexible. We now rewrite each of the ecological variables as follows:

$$\bar{X}_u = (\bar{X}_u - \bar{X}_h) + (\bar{X}_h - \bar{X}_c) + \bar{X}_c \quad (4.3)$$

where $\bar{X}$ is the aggregated average value of one of the work environment dimensions and the subscripts *u*, *h*, *c* represent the nursing unit, hospital and country levels, respectively. This representation partitions

the nursing unit level covariate into a sum of three parts: the nursing unit level deviation from the hospital level mean, the hospital level deviation from the country level mean and the country level mean. The rationale is that, by partitioning the covariates we can estimate the pure effect of the covariate at each level. For example, by subtracting the hospital level mean from the nursing unit level variable, we keep only the unit level effect thereby removing the higher level effects.

4.2.5.3 Step-by-step multilevel modeling approach

We propose a step-by-step approach toward building a model for multilevel regression analysis of the relationship between a multidimensional covariate and a multidimensional dichotomous outcome. To relate the work environment dimensions to burnout dimensions, we first build a series of nine univariate simple multilevel probit models. Here, we consider the impact of each work environment dimension on each burnout dimension separately. Second, we build a series of three univariate multiple multilevel probit models. Here, we consider the joint impact of the work environment dimensions on each of the burnout dimensions separately. The extension to the multiple model only involves adding more covariates at each level. The probit model assumes that there is an underlying latent variable Z that follows a normal distribution with standard deviation one, conditional on all the fixed effects. This latent variable expresses the true feeling of the nurse and is assumed that for $Z > 0$, burnout is expressed on a manifest scale indicated by $Y = 1$, otherwise zero. In the univariate simple multilevel probit models, we have one ecological covariate decomposed as in Equation 4.3, augmented with a random intercept for each higher level. The random effects follow a normal distribution with mean zero and a specific variance. No random slopes were included into the model following exploratory analyses (using Akaike's information criterion (Akaike, 1974), indicating that the relationship between the work environment and burnout differs in different levels, but stays constant within each level. We used the same settings for the random effects in the univariate multiple multilevel probit models. The final outcome of our step-by-step approach is a multivariate multilevel probit model. The combination of the three univariate multiple multilevel probit models results in a three-variate four-level probit model, which could also be called, in general, the multilevel multivariate probit model (MVP). In this modeling, a common factor was introduced to construct the correlations among the three burnout dimensions (cf. three-variate). Similar to the

univariate models, all three covariates are partitioned into three parts (unit, hospital, and country level pure effects). There are three random intercepts vectors corresponding to the three higher levels for each outcome, as well as the three random factor loadings, which imply a varying correlation structure.

4.2.5.4 Computational aspects

We used the R lme4 package (Bates & Sarkar, 2007) to fit the univariate simple and multiple multilevel probit models. However, the multivariate multilevel probit model is beyond the scope of this package and we are not aware of any frequentist software that can handle this model. For this reason, we used the popular WinBUGS package. This software is based on the Bayesian paradigm and uses Markov Chain Monte Carlo sampling techniques to arrive at the parameter estimates. A Bayesian analysis needs prior distributions for all its parameters. We have used here the following priors. For the regression coefficients we have taken a normal distribution with mean zero and a large variance. The factor loadings (ls) were given a (multivariate) normal distribution with hyper-parameters, i.e. the variance (matrix), which has a vague conjugate inverse Wishart distribution. The random intercepts at each level followed the same priors as the factor loadings. For the posterior statistics, we calculated the posterior mean, median, standard error, and the 95% equal tail credibility interval. This credibility interval is the Bayesian equivalent of the classical 95% confidence interval, which indicates a significant nonzero estimate if the interval does not include zero, and a non-significant estimate if the interval includes zero. The hierarchical centering strategy of Gelfand, Sahu & Carlin (1995) was applied to improve the convergence of the MCMC iterations. Three chains were initialized with different starting values. We obtained posterior means and 95% credible intervals based on 10,000 iterations after having removed a burn-in part of 20,000 iterations. The Brooks–Gelman–Rubin diagnostic plot (Brooks & Gelman, 1998), which tests the within- and between-chains variation, was used to check the convergence of all parameters. The WinBUGS program is available from the first author.

4.3 Results

4.3.1 General description

The mean estimates of emotional exhaustion and depersonalization burnout rates in Greece are higher than for the other countries, while Poland has the highest rate of reduced personal accomplishment. Greece shows the widest interquartile ranges for all three dimensions of burnout, while the Netherlands shows the narrowest. The burnout rates thus vary greatly across Greece hospitals but are stable across Dutch hospitals. Swiss nurses' ratings of their work environment are the highest for all three dimensions. Greek and Polish nurses' ratings of their work environment are lowest. The full descriptive findings were published previously by Aiken et al. (2012).

4.3.2 Intra-class correlation coefficients

Table 4.1 shows the proportion of total variance that could be explained at each level for the three burnout dimensions and environment dimensions. The country level explained about 22% of the variation in emotional exhaustion, 13% in depersonalization and 6% in personal accomplishment. The hospital level explains the least variation. Less than 5% in the variation of all three outcomes can be explained at the hospital level. The nursing unit level contributes about 10%, 6% and 2% for the three outcomes, respectively. These multilevel variances decomposition indicate the modeling for multilevel analyses. For the variances decomposition of the three environment dimensions, the different proportions at each level suggest different ranges of the environment variations. The hospital level variation for each environment dimension is the smallest among the three levels, respectively. This will be further discussed in the discussion part.

Table 4.1 Proportion of total variance explained of the three burnout dimensions and environment dimensions at the country, hospital, nursing unit and individual

		Country	Hospital	Nursing unit	Nurse
Outcome	Emotional exhaustion	22.4	3.8	9.5	64.4
	Depersonalization	13.0	3.2	6.3	77.5
	Personal Accomplishment	6.0	2.3	2.3	89.4
Covariates	Managerial support for nursing	7.3	3.1	89.6	-
	Doctor-nurse collegial relations	14.8	2.0	83.2	-
	Promotion of care quality	15.0	4.6	80.4	-

4.3.3 Relationship between the work environment and burnout

All nine univariate simple multilevel models gave significant negative effect estimates for the ecological variables at almost all levels (Table 4.2). The effect is most pronounced for emotional exhaustion, while personal accomplishment shows the weakest effect. An exemplary graph of a univariate simple multilevel model is given in Figure 4.1. This figure displays a clear negative trend between the country level ecological variable of managerial support of nurses and emotional exhaustion. The negative regression line is the adjusted line that takes into account the number of nurses in each country, which is represented by the area of the circle. Greece has the smallest sample size and appears to be an outlier. Table 4.2 displays the results for the three univariate multiple multilevel probit models. There is a pronounced ecological effect of the nursing unit level variability in the relationship between the work environment dimensions of doctor–nurse collegial relations and promotion of care quality and all three burnout dimensions. The effect of the nursing unit level variability for the dimension of managerial support of nursing is only present for emotional exhaustion. At the hospital level, the latter effect is present for both emotional exhaustion and depersonalization. Doctor–nurse collegial relations have no effect on either burnout dimension at the hospital level. Promotion of care quality is significantly related to all three burnout dimensions at the hospital level. At the country level, we found only an effect for doctor–nurse collegial relations on personal accomplishment. This effect was absent in the three-variate four-level probit model (Table 4.3). The other fixed effects and the standard deviations in the final model are similar to those of univariate multiple multilevel models. The 95% equal tail credibility interval is the Bayesian equivalent of the classical 95% confidence interval. That is, the estimate is significantly larger/smaller than zero if the interval does not include zero, and not significant if the interval includes zero.

Table 4.2 Univariate simple and multiple probit model estimates

Outcomes	Covariates	Levels	Univariate simple models			Univariate multiple models		
			EST	SE	p-value	EST	SE	p-value
Emotional exhaustion	Managerial support for nursing	Nursing unit	-.623	.039	<.0001	-.270	.050	<.001
		Hospital	-.672	.079	<.001	-.540	.107	<.001
		Country	-2.330	.708	.001	-2.604	1.341	.052
	Doctor-nurse collegial relations	Nursing unit	-.483	.050	<.001	-.108	.052	.037
		Hospital	-.496	.119	<.001	.073	.133	.582
		Country	-1.114	.729	.126	.634	.877	.469
	Promotion of care quality	Nursing unit	-1.109	.061	<.001	-.789	.078	<.001
		Hospital	-.684	.097	<.001	-.298	.125	.017
		Country	-1.559	.661	.018	-.421	.743	.571
Depersonalization	Managerial support for nursing	Nursing unit	-.406	.037	<.001	-.083	.048	.082
		Hospital	-.489	.071	<.001	-.270	.097	.005
		Country	-1.634	.494	.001	-1.839	.940	.050
	Doctor-nurse collegial relations	Nursing unit	-.419	.047	<.001	-.172	.050	.001
		Hospital	-.493	.104	<.001	-.085	.120	.480
		Country	-.793	.504	.116	.429	.615	.485
	Promotion of care quality	Nursing unit	-.838	.058	<.001	-.671	.075	<.001
		Hospital	-.597	.084	<.001	-.354	.113	.002
		Country	-1.081	.467	.021	-.273	.521	.600
Personal accomplishment	Managerial support for nursing	Nursing unit	-.257	.032	<.001	-.031	.041	.453
		Hospital	-.302	.059	<.001	-.087	.080	.277
		Country	-1.008	.331	.002	-.279	.557	.616
	Doctor-nurse collegial relations	Nursing unit	-.343	.039	<.001	-.200	.043	<.001
		Hospital	-.350	.085	<.001	-.087	.100	.387
		Country	-.900	.228	<.001	-.727	.364	.046
	Promotion of care quality	Nursing unit	-.537	.049	<.001	-.407	.065	<.001
		Hospital	-.437	.069	<.001	.336	.093	<.001
		Country	-.448	.341	.189	-.025	.310	.939

Figure 4.1 Relationship between emotional exhaustion and managerial support for nursing at the country level

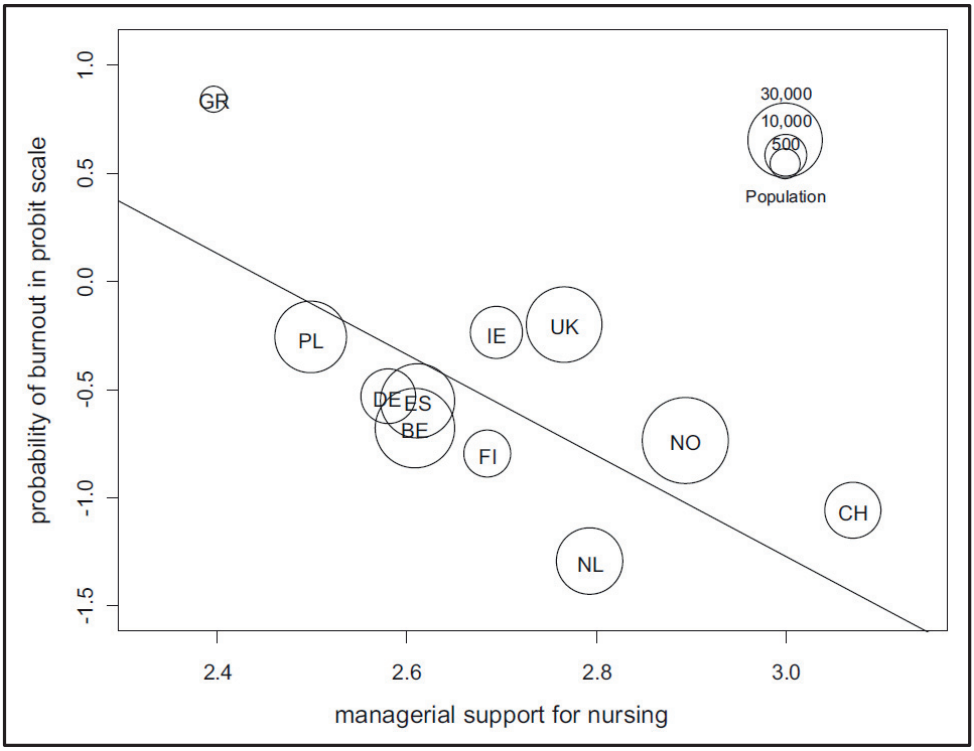


Table 4.3 Bayesian multivariate multilevel probit model estimates

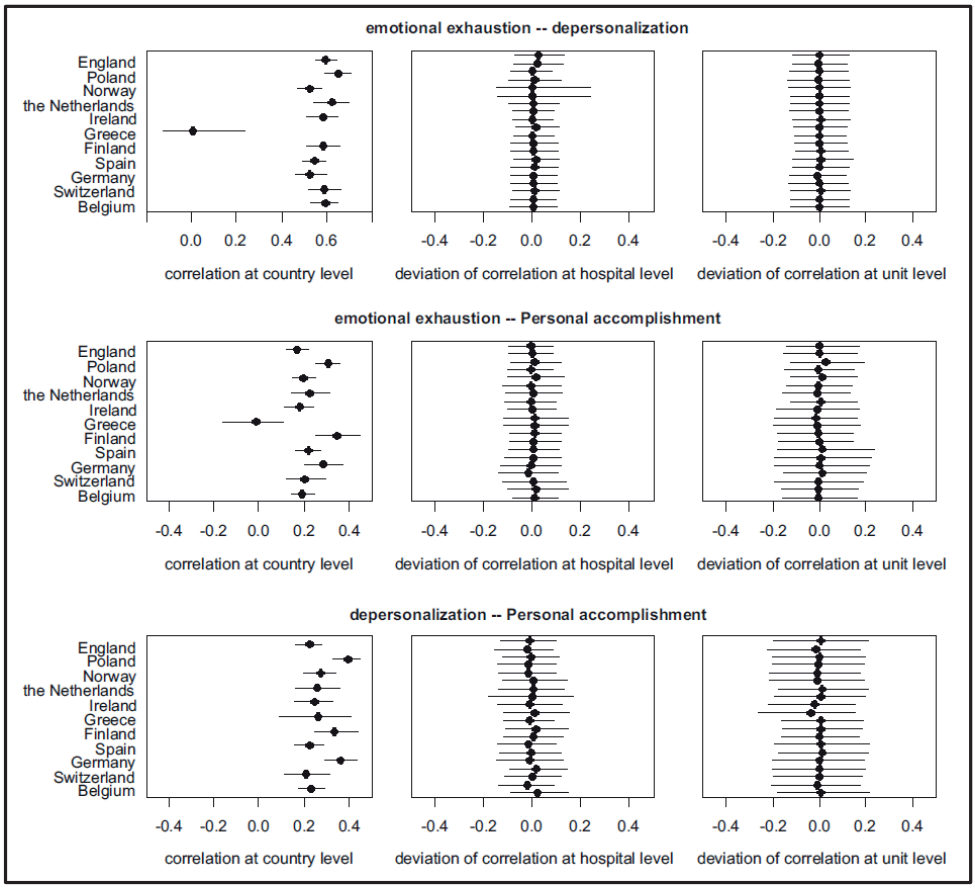
Outcomes	Covariates	Levels	Mean	SE	2.5%	Median	97.5%
Emotional exhaustion	Managerial support for nursing	Nursing unit	-.277	.050	-.373	-.278	-.178
		Hospital	-.532	.115	-.754	-.531	-.310
		Country	-2.572	2.283	-7.138	-2.571	1.935
	Doctor-nurse collegial relations	Nursing unit	-.108	.054	-.214	-.108	-.002
		Hospital	.056	.142	-.219	.057	.335
		Country	.628	1.477	-2.295	.619	3.587
	Promotion of care quality	Nursing unit	-.783	.079	-.935	-.783	-.628
		Hospital	-.284	.135	-.545	-.286	-.019
		Country	-.451	1.293	-3.043	-.440	2.074
Depersonalization	Managerial support for nursing	Nursing unit	-.094	.049	-.188	-.094	.002
		Hospital	-.265	.107	-.478	-.264	-.059
		Country	-1.856	1.892	-5.730	-1.822	1.769
	Doctor-nurse collegial relations	Nursing unit	-.168	.050	-.265	-.167	-.071
		Hospital	-.105	.133	-.362	-.105	.154
		Country	.442	1.237	-1.949	.425	2.930
	Promotion of care quality	Nursing unit	-.667	.077	-.818	-.665	-.516
		Hospital	-.344	.126	-.588	-.346	-.093
		Country	-.297	1.066	-2.356	-.304	1.832
Personal accomplishment	Managerial support for nursing	Nursing unit	-.031	.043	-.114	-.031	.053
		Hospital	-.101	.093	-.281	-.101	.081
		Country	-.236	1.581	-3.351	-.244	2.953
	Doctor-nurse collegial relations	Nursing unit	-.204	.045	-.293	-.204	-.116
		Hospital	-.084	.114	-.310	-.083	.137
		Country	-.755	1.034	-2.799	-.760	1.291
	Promotion of care quality	Nursing unit	-.410	.038	-.545	-.410	-.277
		Hospital	-.319	.107	-.531	-.319	-.113
		Country	-.042	.883	-1.841	-.036	1.737

4.3.4 Relationship among the burnout responses

Fig. 4.2 displays the partitioned level-specific correlation structures among the three burnout dimensions. The dots are the posterior means and the lines are the 95% confidence interval in Bayesian way (for the country level or hospital level, this confidence interval is actually the median interval within the country or hospital). A sample of 20 is randomly selected at hospital and nursing unit levels respectively to make the figure readable. At the country level (first column in Figure 4.2), the correlations varied vastly, with some significant differences between countries. At the hospital level (second column in Figure 4.2), all remaining correlation structures stayed close to zero after removing the country level correlations. Similar findings are seen for the nursing unit level after removing the country and hospital level correlations (third column in Figure 4.2). This indicates that the correlation structure among the three outcomes was quite different between countries, but stayed stable between hospitals within

countries and between nursing units within hospitals. Greece again performed very differently from the other countries.

Figure 4.2 Partitioned level-specific correlation structures among the three burnout outcomes. (For the hospital and nursing unit level deviation of correlation (the 2nd and 3rd columns), a sample of 20 is randomly selected, respectively)



4.4 Discussion

In this paper, we investigated the relationship between nurse work environment dimensions (managerial support for nursing, doctor–nurse collegial relations, and promotion of care quality) and burnout dimensions (emotional exhaustion, depersonalization, and personal accomplishment) using an advanced multilevel approach. We aimed to explore and investigate the effect of the nursing unit, hospital, and country level variability on the relationship between dimensions of nurses' work

environment and dimensions of burnout. We also explored the significance of the nursing unit, hospital, and country level variability among the burnout dimensions. We first specified ecological measures of the nurse work environment dimensions at the three organizational levels (nursing unit, hospital, and country). The effect of the covariate was decomposed into effects on different levels. This so called partitioning strategy allowed us to specify the pure effect of the covariate at each level. We then combined these ecological measures with individual-level burnout experiences within a series of multilevel statistical models that would allow us to model the complex contextuality and heterogeneity. Our approach toward building a model for multilevel regression analyses of the relationship between such multidimensional covariate and multidimensional dichotomous outcome took three steps. We first fitted univariate simple multilevel probit models where we considered the impact of each work environment dimension on each burnout dimension separately. Second, we fitted univariate multiple multilevel probit models where we considered the joint impact of the work environment dimensions on each of the burnout dimensions separately. Last, we fitted a multivariate multilevel probit model where we considered the joint impact of the work environment dimensions on the three burnout dimensions. Not surprisingly, our results showed a negative relationship between work environment dimensions and burnout experiences among nurses. However, by maintaining in our advanced analyses the social context in which the data were collected, we added some interesting findings to what was already known about this relationship. By using the partition strategy and modeling the burnout and work environment dimensions jointly, we now have a more detailed view of their relationship. The final model showed no country-level effect for either work environment dimension on any of the three burnout dimensions. Doctor–nurse collegial relations affected all burnout dimensions at the nursing unit level only. For the dimension of promotion of care quality, the effect of the ecological exposure on burnout was pronounced at both the nursing unit and the hospital level for all three burnout dimensions. The magnitude of this effect was consistently stronger at the nursing unit. Findings for the dimension of managerial support for nursing were ambiguous. The effect on emotional exhaustion was less pronounced at the nursing unit level than at the hospital level. An effect on depersonalization was only present at the hospital level. There was no effect on personal accomplishment at either level. In developing the PES-NWI, Lake (2002)

had already identified that substantive domains of the subscales ranged from the broad hospital context to the immediate nursing unit context, leading her to conclude that the subscales exhibit multilevel range in hospital contexts. The varying magnitude in effects found at different levels in this study pleas for the use of a multilevel analysis in future studies. The results should, however, be interpreted with caution. Previous efforts from our research group have shown that aggregating features of nursing care of all nursing units at the hospital level might obscure the hospital level effects on outcome measures (Van den Heede, Sermeus, et al., 2009). For this study, that means that even though there is no hospital effect for some work environment dimensions on certain burnout dimensions, human resources management should not conclude that interventions at the hospital level are by definition not effective. Statistical support for this is given by the large variance of work environment and burnout dimensions at the unit than at the hospital level, as seen from the intraclass correlations (Table 4.1). The results indicate that all three work environment dimensions deserve the attention of human resources management to secure better outcomes. The pronounced effects of the dimensions of promotion of care quality and managerial support for nurses at both the nursing unit and the hospital level point to a need for leaders from bedside to boardroom to further develop managerial skills and share goals for achieving positive health care environments. Front-line, middle and upper managers need to move toward an integrated vision on promotion of care quality in tune with the workforce. As shown by our empirical findings, at the unit level, nurses should partner up with physicians. The multivariate multilevel probit model allowed for a flexible hierarchical correlation structure. We found a positive correlation matrix among the three burnout variables. This varied across countries, but was stable across hospitals and nursing units within a specific country. There is a large body of literature that has described the relationship between burnout and work environment. Although some of them used different measure instruments, they came up with similar findings. Melchior et al. (1997) analyzed the relationship between burnout and work environment at nurse level and nursing unit level separately. They found significant relationships at each level. However, their modeling is not very efficient (separate analysis for each level) and suffers from a small sample size at the nursing unit level. Van Bogaert et al. (2009) explored the nursing unit level relationship between work environment dimensions and burnout using a 2-level linear mixed model for each of the three

burnout dimensions separately. They found significant relationships for all environment coefficients. O'Mahony (2011) studied the relationship between work environment and two burnout dimensions (emotional exhaustion and depersonalization), and found a significant correlation through simple linear regressions. Liu et al. (2012) analyzed data from South China using a logistic regression model and concluded that improving the unit level work environment from poor to better leads to a moderate (33% decrease in job-related emotional exhaustion. All studies mentioned used an RN4CAST-like operationalization of the work environment and burnout. As can be inferred in the results part, our 3-variate 4-level probit model detected similar findings as previous works at the nursing unit level. However, we provide more detailed information, also at higher levels, i.e. hospital and country levels. Stepping back from what this article adds, it is not free from statistical, practical and conceptual limitations. First, we encountered multicollinearity between dimensions of the work environment. This multicollinearity could be due to specific items of different dimensions correlating highly, rather than the whole dimensions. A confirmatory factor analysis (CFA) is needed to study the factor structure underlying these items. The four-level RN4CAST data structure would require a more complicated multilevel CFA to detect the potentially different factor structure in each level. Such approach requires the application of new goodness-of-fit tests for verifying the statistical assumptions made at the different levels of the hierarchy. This analysis was beyond the scope of this study. Second, the Bayesian multivariate multilevel probit model included no fixed effects in the correlation structure among the three burnout dimensions, although adding covariates is theoretically possible. In practice, such models with both fixed and random effects in correlation structure causes rather slow convergence and needs to be improved further.

A third possible limitation is that country was treated as a random effect throughout this paper. However, since country is not chosen at random (in contrast to hospital and nursing unit), we could have assumed that it has a fixed effect, involving an index variable for each country. In studies that involve multiple levels, researchers should be cautious of four types of fallacies that potentially arise when the methods fail to fit the conceptual model (Diez-Roux, 1998). Ecological fallacies arise when drawing inferences at the individual level based on group-level data. Atomistic fallacies occur when drawing inferences at the group level based on individual level data. Ecological and atomistic fallacies are both

types of inferential fallacies that can be overcome by ensuring that the data collected match the level at which inferences are to be made, as was accomplished by the design of the RN4CAST project. The psychological fallacy would arise when ignoring the relevant group-level covariates in a study of individual level outcomes. In this article we have considered the nursing unit, hospital and country level variability in the relationship under study. Fourth, the sociological fallacy would arise when ignoring the role of individual level factors in a study of groups. This brings us to a fourth potential limitation of the study. We have shown that social contexts shape burnout experiences among nurses by including group-level variables. By not including possible confounding individual level variables like nurses' age and gender, it might appear that we have perpetuated the idea that burnout experiences are absolutely socially determined rather than leaving room for individual determinants. Combining group-level and individual-level covariates in the proposed models is methodologically challenging. It would be meaningful in future papers to analyze the joint impact of social context and individual characteristics on burnout experiences.

As described in the study sample section, the RN4CAST project accommodated within the framework of a strict hierarchy. A fifth potential limitation is that it is plausible that, although participating nurses strictly worked in the sampled nursing unit and hospital, both covariates and outcomes may be conditioned through social processes operating between nursing units in hospitals. Last, excluding Sweden from the final analysis may be considered misleading and inefficient. We therefore did a sensitivity analysis to see the influence of the Swedish data on the model estimates. These tests consist of three models which are: the hospital level univariate random effects model with country as the random effects, the country level univariate regression and the nurse leve univariate random effects model with country and hospital as the random effects. All these models were ran using both datasets with and without Swedish data to detect differing estimates. The difference was minimal, and the estimates from the two data sets for all the three models were close, both for fixed and random effects.

4.5 Conclusions

Nurse work environment dynamics are related to nurses' burnout experiences at both the nursing unit and the hospital level. The correlation structure among the three burnout outcomes varies across

countries, but is stable between hospitals within countries and between nursing units within hospitals. The findings provide a motivation for nurses and physicians within nursing units to partner up and for nurse leaders from bedside to boardroom to further develop their managerial skills. There is a clear need toward an integrated vision on promotion of care quality in tune with the workforce. The results also imply that, in evaluating health care organizations, researchers should sample the different levels of the organization under study and maintain this structure in analyzing the data.

References

Adewale, A. J., Hayduk, L., Estabrooks, C. A., Cummings, G. G., Midodzi, W. K., & Derksen, L. (2007). Understanding hierarchical linear models: applications in nursing research. Nursing Research, 56(4 Suppl), S40–6.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., Busse, R., Clarke, H., … Shamian, J. (2001). Nurses' reports on hospital care in five countries. Health Affairs (Project Hope), 20(3), 43–53.

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.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

Bates, D., & Sarkar, D. (2007). lme4: Linear Mixed-Effects Models Using s4 Classes.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7(4), 434–455.

Bruyneel, L., Van den Heede, K., Diya, L., Aiken, L. H., & Sermeus, W. (2009). Predictive validity of the International Hospital Outcomes Study questionnaire: an RN4CAST pilot study. Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau, 41(2), 202–210.

Cho, S. H. (2003). Using multilevel analysis in patient and organizational outcomes research. Nursing Research, 52(1), 61–65.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. American Journal of Public Health.

Gajewski, B. J., Boyle, D. K., Miller, P. A., Oberhelman, F., & Dunton, N. (2010). A multilevel confirmatory factor analysis of the Practice Environment Scale: a case study. Nursing Research, 59(2), 147–53.

Gelfand, A. E., Sahu, S. K., & Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. Biometrika, 82(3), 479–488.

Hasselhorn, H.-M., Tackenberg, P., & Müller, B. H. (2003). Working conditions and intent to leave the profession among nursing staff in Europe.

Kelly, L. A., McHugh, M. D., & Aiken, L. H. (2012). Nurse outcomes in Magnet® and non-Magnet hospitals. The Journal of Nursing Administration, 42(10 Suppl), S44–49.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel Modeling of Individual and Group Level Mediated Effects. Multivariate Behavioral Research, 36(2), 249–277.

Kutney-Lee, A., McHugh, M. D., Sloane, D. M., Cimiotti, J. P., Flynn, L., Neff, D. F., & Aiken, L. H. (2009). Nursing: a key to patient satisfaction. Health Affairs (Millwood), 28(4), w669–77.

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in Nursing & Health, 25, 176–188.

Leiter, M. P., & Maslach, C. (2009). Nurse turnover: the mediating role of burnout. Journal of Nursing Management, 17(3), 331–9.

Liu, K., You, L.-M., Chen, S.-X., Hao, Y.-T., Zhu, X.-W., Zhang, L.-F., & Aiken, L. H. (2012). The relationship between hospital work environment and nurse outcomes in Guangdong, China: a nurse questionnaire survey. Journal of Clinical Nursing, 21(9-10), 1476–85.

Maslach, C. (1993). Burnout: a multidimensional perspective. In Professional burnout: Recent developments in theory and research (pp. 191–218).

Maslach, C., & Jackson, S. (1986). Maslach burnout inventory. Consulting Psychologists Press.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. Journal of Organizational Behavior, 2(2), 99–113.

Maslach, C., & Leiter, M. P. (2008). Early predictors of job burnout and engagement. The Journal of Applied Psychology, 93(3), 498–512.

Melchior, M. E. W., Van Den Berg, A. A., Halfens, R., Huyer Abu-Saad, H., Philipsen, H., & Gassman, P. (1997). Burnout and the work environment of nurses in psychiatric long-stay care settings. Social Psychiatry and Psychiatric Epidemiology, 32(3), 158–164.

Nantsupawat, A., Srisuphan, W., Kunaviktikul, W., Wichaikhum, O.-A., Aungsuroch, Y., & Aiken, L. H. (2011). Impact of nurse work environment and staffing on hospital nurse and quality of care in Thailand. Journal of Nursing Scholarship : An Official Publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau, 43(4), 426–32.

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data. Biometrics, 54(2), 638.

O'Mahony, N. (2011). Nurse burnout and the working environment. Emergency Nurse : The Journal of the RCN Accident and Emergency Nursing Association, 19(5), 30–7.

Park, S., & Lake, E. T. (2005). Multilevel Modeling of a Clustered Continuous Outcome. Nursing Research, 54(6), 406–413.

Poghosyan, L., Clarke, S. P., Finlayson, M., & Aiken, L. H. (2010). Nurse burnout and quality of care: cross-national investigation in six countries. Research in Nursing & Health, 33(4), 288–98.

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., … Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nursing, 10, 6.

Spence Laschinger, H. K., & Leiter, M. P. (2006). The Impact of Nursing Work Environments on Patient Safety Outcomes. JONA: The Journal of Nursing Administration, 36(5), 259–267.

Vahey, D. C., Aiken, L. H., Sloane, D. M., Clarke, S. P., & Vargas, D. (2004). Nurse burnout and patient satisfaction. Medical Care, 42(2 Suppl), II57–66.

Van Bogaert, P., Meulemans, H., Clarke, S., Vermeyen, K., & Van de Heyning, P. (2009). Hospital nurse practice environment, burnout, job outcomes and quality of care: test of a structural equation model. Journal of Advanced Nursing, 65(10), 2175–2185.

Van den Heede, K., Sermeus, W., Diya, L., Clarke, S. P., Lesaffre, E., Vleugels, A., & Aiken, L. H. (2009). Nurse staffing and patient outcomes in Belgian acute hospitals: cross-sectional analysis of administrative data. International Journal of Nursing Studies, 46(7), 928–39.

Warshawsky, N. E., & Havens, D. S. (2011). Global use of the Practice Environment Scale of the Nursing Work Index. Nursing Research, 60(1), 17–31.

Wu, Y. W. B. (1997). An Application of Hierarchical Linear Models to Meta-Analysis in Nursing Research. Nursing Research, 46(5), 295–298.

Appendix

A. Univariate simple multilevel probit model

Let Z_{nuhc} be one of the three latent normally distributed responses, representing the *n*th nurse, within the *u*th nursing unit, the *h*th hospital and the *c*th country. This variable expresses the true burnout feeling of the nurse. It is assumed that for $Z_{\text{nuhc}} > 0$, burnout is expressed on a manifest scale indicated by $Y_{\text{nuhc}} = 1$, otherwise 0.

$$\begin{aligned} &Z_{nuhc} = \beta_0 + \beta_1 X_{uhc} + \beta_2 X_{hc} + \beta_3 X_c + b_{0c} + b_{0hc} + b_{0uhc} + \varepsilon_{nuhc} \\ &b_{0c} \sim N(0,\sigma_1^2), b_{0hc} \sim N(0,\sigma_2^2), b_{0uhc} \sim N(0,\sigma_3^2), \varepsilon_{nuhc} \sim N(0,1) \\ &n = 1,2,...,N_u, u = 1,2,...,N_h, h = 1,2,...,N_c, c = 1,2,...,11 \end{aligned} \quad (1)$$

B. Multivariate multilevel probit model

For a better understanding, we first introduce the classical multivariate probit model (MVP) (single-level structure). The classical MVP has been widely studied by many researchers with different theories and solutions, see e.g. Bock and Gibbons (1996), Chib and Greenberg (1998), Lawrence et al. (Lawrence, Bingham, Liu, & Nair, 2008). Here we adopt Bock and Gibbons' solution of factor modeling which is formally defined as follows:

$$\begin{aligned} &Z_i = \beta_0 + \beta_1 X_i + \lambda_0 F_i + \varepsilon_i \\ &F_i \sim N(0,1), \varepsilon_i \sim N(0,\Sigma_\varepsilon), i = 1,2,...,N \end{aligned} \quad (2)$$

where in our three-variate case, Z_i is a 3×1 vector of latent continuous responses at the *i*th observation, with the same definition of latent variable as in the univariate case. The vectors of regression coefficients β_0 and β_1, and the vector of factor loadings λ_0 are of length 3, corresponding to the three outcomes. The common factor F_i serves to model the correlations of the three outcomes in combination with the covariance structure of the residuals ε_i, Σ_ε. We assumed that this covariance matrix is diagonal (errors independent). The covariance matrix of the random part, i.e. $\text{cov}(\lambda_0 F_i + \varepsilon_i) = \lambda_0 \lambda_0^T + \Sigma_\varepsilon$, is called

the factor analytic representation of the covariance which can reproduce any 3×3 covariance matrix by an appropriate choice of λ_0 and Σ_ε. Note that here the covariance matrix equals the correlation matrix because the variances are assumed to be one. This model assumes that the correlation matrix is the same across all countries, hospitals and nursing units, which might be not a realistic assumption. In order to vary the correlation across countries one could replace $\lambda_0 F_i$ by $(\lambda_0 + \lambda_c)F_i$, whereby λ_c changes with country. In this way there are as many correlation matrices as countries. We can further extend this expression to let the correlation vary also with hospital and nursing unit resulting in a term $(\lambda_0 + \lambda_c + \lambda_{hc} + \lambda_{uhc})F_{nuhc}$. In this way we have a different correlation matrix for each nursing unit. Because of the large number of nursing units (2,087) we have assumed that for λ_c, λ_{hc} and λ_{uhc}, each has a normal distribution with mean zero and a variance to be estimated, which reduces drastically the number of parameters to estimate but also expresses that we do expect that the correlations across nursing units, hospitals and countries do not vary wildly.

We then implemented the four-level structure into the MVP model. This is defined as follows (similar notation as in the univariate multilevel probit model in appendix A):

$$\begin{aligned} &Z_{nuhc} = \beta_0 + \beta_1 X_{uhc} + \beta_2 X_{hc} + \beta_3 X_c + b_{0c} + b_{0hc} + b_{0uhc} + (\lambda_0 + \lambda_c + \lambda_{hc} + \lambda_{uhc})F_{nuhc} + \\ &b_{0c} \sim N_3(0,\Sigma_1), b_{0hc} \sim N_3(0,\Sigma_2), b_{0uhc} \sim N_3(0,\Sigma_3) \\ &\lambda_c \sim N(0,\Sigma_4), \lambda_{hc} \sim N(0,\Sigma_5), \lambda_{uhc} \sim N(0,\Sigma_6), F_{nuhc} \sim N(0,1), \varepsilon_{nuhc} \sim N(0,\Sigma_6) \\ &n = 1,2,...,N_u, u = 1,2,...,N_h, h = 1,2,...,N_c, c = 1,2,...,11 \end{aligned} \tag{3}$$

In this model, all observed and latent variables have a multilevel structure with multiple subscripts defined in the same way as before. The covariates are partitioned into three parts, as was done in the univariate model. There are three random intercepts vectors corresponding to the three higher levels for each outcome, as well as the three random factor loadings. As the factor is introduced to model the correlations among the three outcomes, the varying factor loadings imply a varying correlation structure. Restrictions are needed to render the model identifiable, which are of the same type as above.

Appendix references

Bock, R. D., & Gibbons, R. D. (1996). High-dimensional multivariate probit analysis. Biometrics, 52(4), 1183–94.

Chib, S. (1998). Analysis of multivariate probit models. Biometrika, 85(2), 347–361.

Lawrence, E., Bingham, D., Liu, C., & Nair, V. N. (2008). Bayesian Inference for Multivariate Ordinal Data Using Parameter Expansion. Technometrics, 50(2), 182–191.

5 ORGANIZATION OF HOSPITAL NURSING, PROVISION OF NURSING CARE, AND PATIENT ASSESSMENTS OF CARE IN EUROPE

This chapter is submitted for publication:

Luk Bruyneel, Baoyue Li, Dietmar Ausserhofer, Emmanuel Lesaffre, Irina Dumitrescu, Herbert Smith, Douglas Sloane, Linda Aiken and Walter Sermeus. Organization of hospital nursing, provision of nursing care, and patient assessments of care in Europe.

Abstract

This study integrates previously isolated findings of nursing outcomes research into an explanatory framework in which care left undone and nurse education levels are of key importance. A moderated mediation analysis of survey data from 11,549 patients and 10,733 nurses in 217 hospitals in eight European countries shows that patient care experience is better in hospitals with better nurse staffing and a more favorable work environment in which less clinical care is left undone. Clinical care left undone is a mediator in this relationship. It is left undone less frequently in hospitals with better nurse staffing and more favorable nurse work environments, and in which nurses perform less overtime and are more experienced. Higher proportions of nurses with a bachelor's degree reduce the effect of worse nurse staffing on more clinical care left undone.

5.1 Introduction

Hospital work environments supportive of professional nursing practice are critically important in providing safe, high quality patient care (Aiken et al., 2012, 2014; Kane, Shamliyan, Mueller, Duval, & Wilt, 2007; McHugh et al., 2013). High-quality care inheres in the interaction of nurses with patients. Hospital work environments that are supportive of nursing practice are those in which well-trained nurses have the autonomy and time to exercise maximally their professional competences in service to patient care. Although the patient safety movement has emphasized systems for avoiding errors of commission in the delivery of health care (e.g., marking the wrong eye for surgery), low quality nursing care also inheres in the omission of beneficial care (Kalisch, Landstrom, & Hinshaw, 2009). Thus studies in England (Ball, Murrells, Rafferty, Morrow, & Griffiths, 2014) and the U.S. (Sochalski, 2004) have shown that units evaluated by nurses as having lower quality of care are often those in which required patient care tasks are being missed or omitted.

Kalisch (2006) has delineated nine specific aspects of regularly missed nursing care—ambulation, turning, delayed or missed feedings, patient teaching, discharge planning, emotional support, hygiene, intake and output documentation, and surveillance—and observes that nursing staff typically attribute their omission to factors that are indicative of organizational deficiencies, including too few staff, poor use of existing staff resources, delays in nursing intervention, poor teamwork, and ineffective delegation, habit and denial (Kalisch, Doumit, et al., 2013; Kalisch, Tschannen, & Lee, 2011).

In this paper, we elaborate the role of the provision of nursing care—specifically, the amplitude of nursing care tasks that are left undone—on the relationship between the nursing work environment and care as perceived by patients. This has two aspects.

First, we measure the extent to which nursing care that is left undone *mediates* the relationship between the organization of hospital nursing and patient assessments of the quality of their in-hospital care. The direct association between patient assessments and two key elements of the organization of hospital nursing, patient-to-nurse ratios and the character of the nurse work environment, is well-attested (Aiken et al., 2012; Kutney-Lee et al., 2009; Vahey et al., 2004; You et al., 2013). The recent

Registered Nurse Forecasting (RN4CAST) study (Sermeus et al., 2011), conducted in 12 European countries, demonstrates that although necessary care activities are frequently omitted in the nursing care process, this happens less often in hospitals with more favorable work environments, lower patient-to-nurse ratios, lower proportions of nurses performing non-nursing tasks (Ausserhofer et al., 2013) and lower proportions of nurses working overtime (Griffiths et al., 2014). The study by Griffiths et al. also shows that overtime is associated with nurses' reports of poor or failing patient safety and poor or fair quality of care. Researchers in the U.S. have found that working overtime is associated with an increased likelihood of making errors, an effect that is stronger with longer shifts (Rogers, Hwang, Scott, Aiken, & Dinges, 2004). Patient satisfaction decreases with extended shifts, and nurses are also at a higher risk of burnout (Stimpfel, Sloane, & Aiken, 2012). Patients recognize when needed care is omitted (Kalisch, McLaughlin, & Dabney, 2012), and patients' reports of missed care correlate with adverse events and converge with nurses' reports of missed care (Kalisch, Xie, & Dabney, 2013). We show that relationships similar to those observed in the U.S.—from nursing organization and the extent of work left undone to patient care assessment—obtain in the European data as well. We then decompose the association between nursing organization and patient care assessment into the direct relationship and the portion that is mediated by the provision of nursing care. In particular, we establish the dimensions of missed nursing care that figure most prominently in the path to lower patient care assessments.

Second, we specify and measure how the importance of tasks left undone, as an intervening variable, is *moderated* by the effects of the educational level of hospital staff nurses. Previous analyses of RN4CAST study data have shown a significant association between better educated nurses and lower patient mortality (Aiken et al., 2014), a result also found in comparable data for the U.S. (Aiken et al., 2011). However, these data did not show a direct effect of education on patient experiences with care (Aiken et al., 2012). This may be because organizations with a better educated nurse workforce will increase the productivity, in the sense of allocating tasks more efficiently across the available labor supply. We know that similar interactions have been observed elsewhere—for example, that increases in staffing (quantity labor factors) are less productive in deleterious nursing work environments than in good work environments (Aiken et al., 2011). Our analytic model thus emphasizes the *synergy* between nursing

education and nurse staffing. Nurses educated to at least a bachelor's degree may be better able to mitigate the negative impacts of constraints posed on nurse staffing within a hospital (Griffiths, Maben, & Murrells, 2011). Nurse staffing levels will interact with nurse education in their effect on care left undone and patient evaluations of care. It is in this sense that nursing education is a moderating factor (Baron & Kenny, 1986; James & Brett, 1984).

5.2 New contribution

Process variables explaining why the organization of nursing care impacts care outcomes have long been theorized, but their effects have not been empirically tested through the application of appropriate statistical techniques. Advances in multilevel mediation and moderation analysis allow us to better represent statistically the evolving theory. Using quality measures that are the basis for public reporting of patients' assessments of their hospital experience and for hospital value based purchasing in the U.S., we now provide evidence on the interrelationship between the structure, the process and patient outcomes of nursing care. We also provide decision-makers with evidence of the joint effect of nurse workload and education levels on the completion of necessary nursing care.

5.3 Method

5.3.1 Study design

We analyze the cross-sectional RN4CAST study data from countries in which surveys of both nurses and patients on general surgical and internal medicine units were collected in 2009-2010. As Table 5.1 shows, data are available for 11,549 patients and 10,733 nurses in 217 hospitals in eight countries (Belgium, Finland, Germany, Greece, Ireland, Poland, Spain and Switzerland). Patients rated the quality of their care. Nurses provided information on nursing care in their hospital, their wellbeing, their last shift and staff characteristics. For both the nurse and patient surveys, a rigorous translation procedure and cultural adaptation of items was undertaken (Squires et al., 2012, 2013). The full data collection protocol is described by Sermeus et al. (2011).

Table 5.1. Hospitals, patients and nurses sampled in eight European countries

Country	Number of hospitals	Number of nurses	Number of patients
Belgium	60	2,866	2,623
Finland	32	1,131	1,947
Germany	12	504	244
Greece	24	367	847
Ireland	10	486	285
Poland	30	2,605	4,136
Spain	15	1,181	470
Switzerland	34	1,593	997
Total	217	10,733	11,549

5.3.2 Measures

5.3.2.1 Patient experiences with hospital care (outcome variables)

Patients' overall ratings of the hospital and their willingness to recommend the hospital to friends and family, two global items derived from the U.S. Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey (Darby, Hays, & Kletke, 2005), are used as measures of patient experiences with care. Both measures are re-scaled to reflect the empirical densities of responses and the modal response categories used by patients. Ratings of the hospital from 0 (worst) to 10 (best) were dichotomized to contrast ratings of 0-8 versus 9-10. For recommending the hospital to family and friends, we contrasted respondents who would definitely recommend the hospital with respondents who had reservations about doing so (*'probably yes'*, *'probably no'*, *'definitely no'*) (Aiken et al., 2012; Jha, Orav, Zheng, & Epstein, 2008).

5.3.2.2 Organization of nursing care (main explanatory variables)

Nurse staffing is calculated as the mean number of patients assigned to nurses on their last shift. The nurse work environment is defined by organizational characteristics that facilitate or constrain professional nursing practice, using the 32-item Practice Environment Scale of the Nursing Work Index (PES-NWI). This measure consists of five subscales of items rated on a 4-point scale, with responses ranging from *'strongly disagree'* to *'strongly agree'* (Lake, 2002). The mean of four subscales is used to calculate a composite nursing work environment score (Aiken et al., 2012). The four subscales reflect 'collegial nurse-physician relations', 'nurse participation in hospital affairs', 'nursing foundations for

quality of care' and 'nurse manager ability, leadership, and support of nurses'. A fifth subscale, which measures 'staffing and resource adequacy,' was not included as it overlaps empirically with the direct measure of nurse staffing we employ in our models. The concept of non-nursing tasks performed is measured from nurses' responses about the extent (*'never'*, *'sometimes'*, or *'often'*) to which they had performed non-nursing tasks during their last shift. Responses are categorized as *'often'* versus *'never'* or *'sometimes'*. Overtime was measured from nurses' answers to whether or not they had worked beyond their contracted hours on the last shift they worked. We also employ measures of nurses' type of employment (full-time, part-time) and years of experience as a nurse.

5.3.2.3 Nursing care left undone (mediating variables)

Our construct of nursing care left undone is based on nurses' reports of tasks that were left undone on their last shift due to lack of time, from a list of 13 nursing activities. Exploratory factor analysis was used to examine the dimensions of care left undone. Two dimensions emerged (comparative fit index=.98; Tucker-Lewis Index=.95; root mean square error of approximation=.050; standardized root mean square residual=.035). The first factor contains seven items (adequate patient surveillance, skin care, oral hygiene, pain management, treatments and procedures, administering medication on time, frequently changing the patient's position) that reflect clinical nursing care activities left undone. The second factor consists of five items that reflect planning and communication activities left undone (comfort/talk with patients, educating patients and family, preparing patients and families for discharge, developing or updating nursing care plans/care pathways, planning care). The item 'adequately documenting patient care' did not have a significant loading on any of these factors and was therefore excluded. We evaluated measurement invariance of this factor solution across the eight countries, applying progressively more stringent constraints in multiple group confirmatory factor analysis (Chen, 2007; Meredith, 1993). We found evidence of configural (invariant factor loading pattern), metric (invariant factor loadings) and scalar invariance (invariant factor intercepts) of the factor solution (Horn & McArdle, 1992). This implies that scores can be compared across countries.

5.3.2.4 Nurse education levels (moderating variable)

Nurse education was measured by the percentage of nurses in the hospital that had obtained at least a bachelor's degree. This information was available from the nurse survey.

5.3.2.5 Hospital structural characteristics (confounding variables)

Data on hospital characteristics were provided by the hospital management using their own institutional record systems. We distinguished hospitals with respect to size (number of beds), teaching status (teaching hospital or non-teaching hospital) and technology level (with open heart surgery, organ transplantation or both defining high technology hospitals).

5.3.3 Statistical analysis

Nurse and patient survey data and administrative data on hospital structural characteristics are linked through common identifiers at the country and hospital level. For patients, the original data structure with patients nested within hospitals within countries is preserved, and outcomes are observable at the individual level. These are linked with aggregated (hospital-level) measures of the main explanatory, mediating, moderating and confounding variables derived from the nurse survey and the administrative data. For the outcome measures of patients ratings of their hospital and willingness to recommend it, percentages of missing values were low: respectively, 4.3% (n=498) and 4.5% (n=555). The rates of missing data values from the patient survey are similar to rates in the nurse survey data. Since missing values were evenly distributed across hospitals and since these data were aggregated to the hospital level, this was not of concern in the statistical analyses.

To frame the mediation analysis (Baron & Kenny, 1986; MacKinnon, Fairchild, & Fritz, 2007) we estimate the associations between (a) the main explanatory variables and the outcomes; (b) the main explanatory variables and the hypothesized mediating variables; and (c) the hypothesized mediating variables and the outcomes while also including the main explanatory variables. Statistical techniques and practical examples of (multilevel) mediation and moderated mediation models have recently been extensively discussed in organizational sciences. This study uses the framework proposed by Preacher,

Rucker and Hayes (2007). To account for the clustering of patients within hospitals within countries, our computational model includes three levels. We apply a multilevel (random intercepts) mediation model to analyze whether the two dimensions of care left undone mediate the relationship between hospital-level patient-to-nurse ratios and nursing work environments and patient experiences with care. We then specify a multilevel moderated mediation model to evaluate whether the indirect effect of patient-to-nurse ratios on patient experiences with care through the amount of care left undone remains similar across hospitals with different nurse education levels. The direct effect of care left undone on patient outcomes remains constant. The variables involved in the interaction analysis are group-mean centered, meaning that they deviate around the country mean (Enders & Tofighi, 2007). In all models, we control for hospital size, teaching status and technology level.

Mplus Version 7.2 was used for this study (Muthén & Muthén, 2012). The three-level models were estimated using Bayesian methods, which are imposed by the statistical software for estimating this type of models. This is the first report using multilevel moderated mediation models to estimate the relationship between nursing organization of care and patient outcomes. This technique, including model fit evaluation, is discussed in more detail in the Appendix. Our method of model fit evaluation was used previously in the context of Bayesian analysis in nursing outcomes research (Bruyneel et al., 2014).

5.4 Results

5.4.1 Descriptive findings

Table 5.2 displays descriptive statistics for the main explanatory variables, hypothesized mediators and patient assessments of care. The means and ranges are estimated from hospital-level data and show, for example, that nurses in the 60 hospitals in Belgium reported an average of 1.79 clinical care activities left undone, and that the average ranges from .66 to 3.32 across the 60 hospitals. Most of these descriptive findings have been extensively discussed in previous work (Aiken et al., 2012, 2014; Aiken, Sloane, Bruyneel, Van den Heede, & Sermeus, 2013; Ausserhofer et al., 2013; Griffiths et al., 2014), but the variation in nurse-reported care left undone requires additional comment. In all countries but Greece, the average number of planning and communication activities left undone is higher than the average

number of clinical care activities left undone, despite being measured by two fewer items. On average, Greek nurses' reports on clinical care left undone by far exceed those of other countries, while German nurses report the highest average number of planning and communication activities left undone. Relative to the other countries in this study, Switzerland and Finland are on average rather low on both dimensions of care left undone. Like the other explanatory variables and outcomes included in this analysis, both dimensions of care left undone vary substantially across countries and across hospitals in each country.

5.4.2 Effects of organization of nursing care to patient experiences with care and clinical care left undone

Table 5.3 displays the findings from the regression analysis preceding the mediation analysis. The first step shows that patients report better experiences with care in hospitals with more favorable nursing work environments and lower patient-to-nurse ratios. Performing non-nursing tasks, years of experience, type of employment and performing overtime did not relate to patient assessments of care. Step two shows that more favorable work environments, lower patient-to-nurse ratios and performing less overtime significantly relate to fewer clinical nursing care tasks left undone and fewer planning and communication activities left undone. Fewer clinical care tasks left undone is also associated with more years of experience, while fewer planning and communication activities left undone related to nurses performing fewer non-nursing tasks. There was no interaction between any of these variables that explained missed nursing care. Step three shows that clinical care left undone is associated with patient ratings of their hospitals and their willingness to them, while planning and communication activities left undone is not. Only significant effects will be included in the mediation analysis. Thus, while clinical care activities left undone will be hypothesized to mediate the effect of more favorable work environments and lower patient-to-nurse ratios on better patient experiences with care, type of employment will not be included in the model since it does not relate to patient assessments of care or to the hypothesized mediators of care left undone.

Table 5.2. Measures derived from the nurse and patient surveys, by country (estimated using hospital-level data)

	Nurse survey							Patient survey	
Country	Mean Patient-to-nurse ratio	Mean Nurse work environment score	% Nurses with bachelor's degree	% Nurses often performing tasks below skill level	% Nurses working beyond contracted hours	Mean Number of clinical care activities left undone	Mean Number of planning/communication activities left undone	% Patients rating the hospital 9 or 10	% Patients definitely recommending the hospital
Belgium	10.88 (6.17-16.24)	2.58 (2.30-3.11)	53% (26%-86%)	60.71% (35.90%-84.09%)	37.70% (0%-80.56%)	1.79 (.66-3.32)	1.99 (.36-2.94)	46.50% (24.56%-82.76%)	58.31% (19.23%-80.77%)
Finland	8.33 (5.26-15.64)	2.66 (2.43-3.05)	53% (29%-77%)	21.43% (6.82%-45.45%)	17.89% (3.03%-37.50%)	1.31 (.15-2.92)	1.43 (.60-2.73)	58.70% (25.81%-77.23%)	65.22% (35.71%-8.00%)
Germany	12.95 (7.46-16.26)	2.67 (2.45-2.91)	0%	60.47% (40.00%-75.00%)	34.35% (25.00%-51.72%)	1.76 (.83-2.38)	2.55 (1.83-3.48)	47.42% (23.53%-71.43%)	65.23% (17.65%-85.71%)
Greece	10.13 (6.27-15.84)	2.35 (1.87-2.74)	26% (0%-100%)	27.39% (0%-55.56%)	30.52% (0%-72.73%)	3.20 (1.27-4.29)	2.20 (.75-3.10)	41.96% (15.00%-76.47%)	56.64% (20.97%-84.62%)
Ireland	6.41 (5.35-7.34)	2.66 (2.28-3.14)	70% (46%-85%)	51.08% (41.86%-68.18%)	47.06% (28.07%-57.89%)	1.42 (1.00-2.15)	2.27 (1.50-3.13)	60.18% (36.00%-80.00%)	74.10% (60.00%-96.67%)
Poland	10.50 (7.16-14.89)	2.44 (2.13-2.83)	22% (5%-45%)	43.86% (21.50%-65.06%)	11.78% (2.35%-23.68%)	1.29 (.62-1.95)	2.09 (1.22-2.93)	54.17% (33.57%-76.32%)	56.61% (34.43%-69.75%)
Spain	12.60 (9.50-15.51)	2.40 (2.06-2.67)	100%	16.89% (4.76%-28.76%)	18.21% (4.92%-42.86%)	1.39 (.83-1.92)	2.08 (1.51-2.73)	34.87% (0%-68.18%)	54.11% (18.52%-83.33%)
Switzerland	7.85 (4.64-12.63)	2.93 (2.41-3.34)	11% (0%-40%)	54.05% (32.61%-94.12%)	40.78% (17.86%-72.73%)	1.02 (.45-1.88)	1.58 (.67-2.35)	61.20% (36.11%-100%)	76.94% (40.00%-100%)

Note: The numbers in parentheses indicate the range in the means and percentages across the hospitals in each country.

Table 5.3. Coefficients from the regression analyses preceding the mediation analysis

	Step 1: Association between the main explanatory variables and the two outcomes						Step 2: Association between the main explanatory variables and the hypothesized mediating variables					
	Patients recommending the hospital (patient level outcome)			Patients rating the hospital (patient level outcome)			Clinical care left undone (hypothesized hospital level mediator)			Planning/communication left undone (hypothesized hospital level mediator)		
	Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI	
Organization of nursing care (hospital level main covariates)												
Nursing work environment	.324*	.081	.508	.261*	.032	.482	-1.636*	-2.044	-1.248	-.904*	-1.164	-.636
Patient-to-nurse ratio	-.040*	-.062	-.015	-.042*	-.064	-.020	.039*	.002	.072	.051*	.027	.077
Non-nursing tasks	-.149	-.540	.249	-.164	-.588	.239	.084	-.410	.564	.647*	.239	1.024
Years of experience	-.004	-.021	.014	.000	-.015	.015	-.030*	-.058	-.003	-.011	-.031	.008
Type of employment	-.333	-.724	.105	-.166	-.595	.231	.294	-.273	.889	.263	-.160	.584
Overtime	.005	-.378	.386	.168	-.197	.527	1.332*	.578	1.984	.582*	.010	1.075
Step 3: Associations between the two hypothesized mediating variables and the two outcomes, from models including the main explanatory variables												
Clinical care left undone	-.194*	-.309	-.075	-.144*	-.255	-.050						
Planning/communication left undone	-.013	-.148	.113	-.038	-.161	.080						

5.4.3 Mediation effects

Table 5.4 shows the regression coefficients estimated in the mediation analyses. The amount of clinical care left undone partially mediates the effects of patient-to-nurse ratios and work environment on patients recommending the hospital. It also partially mediates the effect of patient-to-nurse ratios on patient ratings of their hospitals, and fully mediates the effect of work environments on patient ratings of their hospitals. This can be seen from the non-significant association between work environment and patient ratings of their hospitals.

5.4.4 Moderated mediation effects

Table 5.5 provides the estimated regression coefficients from the moderated mediation models for the two patient outcomes, which extends the mediation analyses by introducing nurse education level as a moderating variable. Clinical care left undone mediates the effect of nurse staffing levels on both patient outcomes differently depending on the proportion of nurses trained to a bachelor's degree. The significant interaction effect in the upper part of Table 5.5 indicates that the effect of having higher hospital patient-to-nurse ratios on clinical care left undone is smaller in hospitals with higher proportions of nurses with at least a bachelor's degree. These findings of moderated mediation are illustrated in the path diagram in Figure 5.1. This graphical representation follows the example of first stage moderated mediation described by Edwards and Lambert (2007). The lower part of Table 5.5 shows that the extent to which clinical care left undone mediates the effect of nurse staffing is conditional on nurse education levels. The conditional indirect effect is statistically significant only for hospitals with lower than average nurse education levels. The mediation effect with regards to nursing work environment found previously was included in the moderated mediation analysis and remained stable. As described in the Appendix, model fit was acceptable.

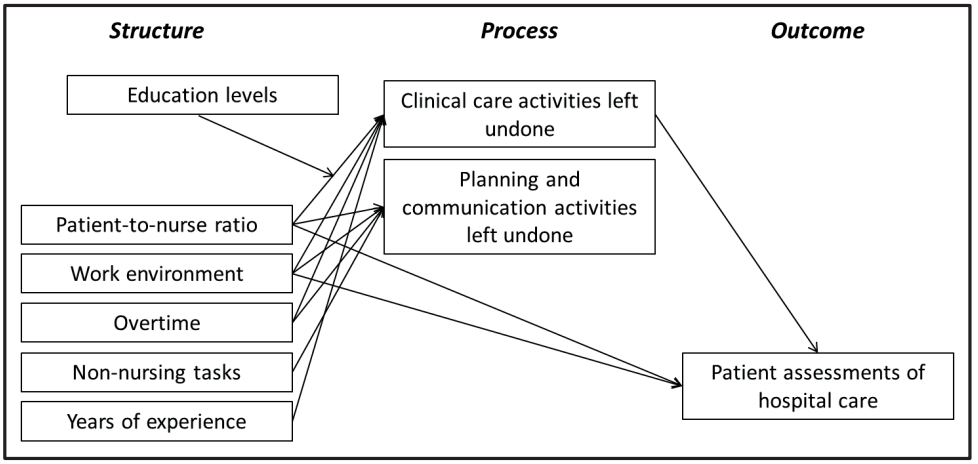
Table 5.4. Findings for the 8-country mediation analyses estimating the indirect effect of nursing work environments and patient-to-nurse ratios on patient experiences with care through clinical care left undone

	Patients recommending the hospital (patient level outcome)									Patients rating the hospital (patient level outcome)								
	Clinical care left undone (hospital level mediator)			Planning/communication left undone (hospital level outcome)			Patient experience (patient level outcome)			Clinical care left undone (hospital level mediator)			Planning/communication left undone (hospital level outcome)			Patient experience (patient level outcome)		
	Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI	
Organization of nursing care (hospital level main covariates)																		
Nursing work environment	-1.740*	-2.127	-1.369	-.904*	-1.164	-.667	.246*	.152	.341	-1.736*	-2.109	-1.389	-.890*	-1.142	-.636	.151	-.126	.318
Patient-to-nurse ratio	.037*	.006	.073	.051*	.030	.075	-.027*	-.055	-.002	.037*	.003	.071	.051*	.027	.076	-.033*	-.055	-.010
Non-nursing tasks	—	—	—	.402*	-.002	.734	—	—	—	—	—	—	.387*	.042	.725	—	—	—
Years of experience	-.031*	-.057	-.008	—	—	—	—	—	—	-.031*	-.057	-.006	—	—	—	—	—	—
Overtime	1.165*	.613	1.694	.579*	.175	.948	—	—	—	1.169*	.614	1.687	.561*	.166	.953	—	—	—
Care left undone (hospital level mediator)																		
Clinical care left undone							-.163*	-.269	-.070							-.168*	-.257	-.088
Indirect effects																		
Work environment on patient experience, through clinical care left undone							.280*	.112	.509							.289*	.147	.470
Staffing on patient experience, through clinical care left undone							-.006*	-.014	-.001							-.006*	-.014	.000

Table 5.5. Findings for the 8-country moderated mediation analyses estimating the effect of nurse education on the relationship between patient-to-nurse ratios on patient experiences with care through clinical care left undone

	Patients recommending the hospital (patient level outcome)									Patients rating the hospital (patient level outcome)								
	Clinical care left undone (hospital level mediator)			Planning/communication left undone (hospital level outcome)			Patient experience (patient level outcome)			Clinical care left undone (hospital level mediator)			Planning/communication left undone (hospital level outcome)			Patient experience (patient level outcome)		
	Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI		Est.	95% CI	
Organization of nursing care (hospital level main covariates)																		
Nursing work environment	-1.658*	-2.024	-1.309	-.893*	-1.154	-.621	.087*	-.096	.329	-1.661*	-2.026	-1.285	-.893*	-1.142	-.639	.205*	.029	.358
Patient-to-nurse ratio (P/N)	.042*	.007	.076	.051*	.027	.075	-.029*	-.053	-.005	.042*	.007	.075	.051*	.028	.075	-.033*	-.055	-.011
Nurse education (NE)	.142	-.501	.767	—	—	—	.067	-.301	.424	.130	-.510	.761	—	—	—	-.056	-.387	.284
P/N * NE Interaction	-.416*	-.669	-.160	—	—	—	-.081	-.240	.083	-.418*	-.687	-.161	—	—	—	.034	-.113	.179
Non-nursing tasks	—	—	—	.390*	.050	.733	—	—	—	—	—	—	.386*	.049	.733	—	—	—
Years of experience	-.034*	-.060	-.008	—	—	—	—	—	—	-.034*	-.060	-.009	—	—	—	—	—	—
Overtime	1.110*	.555	.1669	.566*	.192	.951	—	—	—	1.102*	.584	1.630	.567*	.173	.963	—	—	—
Care left undone (hospital level mediator)																		
Clinical care left undone							-.205*	-.308	-.103							-.154*	-.242	-.069
Indirect effects																		
Work environment on patient experience through clinical care left undone							.337*	.166	.536							.253*	.113	.417
Conditional indirect effects at range of minimum to maximum deviation around group mean																		
Effect of education on the indirect effect of staffing on patient experience through clinical care left undone																		
-.30							-.032*	-.062	-.012							-.025*	-.049	-.009
-.20							-.024*	-.046	-.010							-.019*	-.036	-.007
-.10							-.016*	-.031	-.006							-.012*	-.024	-.004
0							-.008*	-.018	-.001							-.006*	-.014	-.001
.10							.000	-.009	.009							.000	-.007	.007
.20							.008	-.004	.023							.006	-.003	.018
.30							.016	.000	.039							.012	.000	.025

Figure 5.1. Path diagram illustrating the effect of nurse education on the relationship between patient-to-nurse ratios on patient experiences with care through clinical care left undone



5.4.5 Sensitivity analysis

Several sensitivity analyses were performed to evaluate these findings. First, we tested for configural similarity of our construct relations across levels (Chen & Bliese, 2005). We aggregated the main covariates, mediators and moderator to the nursing unit level (n=887), rather than the hospital level, and again linked them to patient experience data using common unit identifiers. Significance tests and the direction of effects were consistent with the hospital-level analysis. This suggests that our theory and findings are homologous across levels. Second, results were also consistent when we used the original values for staffing and education in the moderated mediation analyses instead of group-mean centered values. Third, all analyses that included nurse education levels were repeated omitting the two countries with no variation in nurse education across hospitals (Germany and Spain). Omitting the two countries did not alter our findings.

5.5 Discussion

Nursing- and patient-oriented health services researchers should align the theoretical basis for their research with appropriate statistical techniques (Griffiths, 2009; Levy, Landerman, & Davis, 2011b; Mark, Hughes, et al., 2004; Mark, 2006). This paper responds to those calls by empirically testing a more

comprehensive conceptualization of the mechanisms underlying the association between patient experiences with care and the organizational context of nursing care. Our findings showed that higher patient-to-nurse ratios, worse work environments and higher proportions of nurses working overtime lead to more clinical care activities left undone as well as more planning and communication activities left undone. Previous reports using the RN4CAST study data have shown that nurses all over Europe report having performed non-nursing care and tasks below their skill level during their last shift (Bruyneel et al., 2013). Our findings suggest that, partly as a consequence of this, planning and communication activities are being omitted rather than clinical care activities. Our findings also showed that the omission of planning and communication activities, while decidedly more common than the omission of clinical care, did not however affect patient experiences. Work by Kalisch et al. (2012) has shown that patients are able to partially or fully report on missed nursing care activities that are comparable to the items included in our dimension of planning and communication activities. 'Preparing patients and families for discharge' was included in their study as 'Discharge planning' (fully reportable) and 'Educating patients and family' as 'Patient education' (partially reportable). It could be argued that for some items included in the planning and communication dimension, it is unclear whether patients recognize them, particularly 'developing or updating nursing care plans/care pathways' and 'planning care'. On the other hand, omitting these tasks might lead to disorganized patient care, which patients would be able to recognize. Thus our findings might suggest that patient recognition of omitted planning and communication activities and their consequences does not necessarily result in lower ratings or recommendations of the hospital, whereas this is the case with respect to clinical nursing care left undone. Future studies of the expectations of patients regarding professional hospital care would help to confirm this.

Findings on the moderating effect of nurse education suggest that in hospitals with relatively low proportions of nurses with at least a bachelor's degree, the positive effect of better staffing on patient experiences with hospital care, which occurs partly through reduced clinical care left undone, may be undermined. Bachelor prepared nurses have a wider range of skills and competencies and might especially increase the productivity of the nursing labor through better planning of nursing care and priority setting, which in turn results in lower nursing care activities undone. Important to note is that

nurse education seems not an isomorphic construct across individual and organizational levels. A multi-country analysis of RN4CAST study data at the individual nurse level showed that nurses with a bachelor's degree had higher reports on an overall composite measure of missed nursing care (Ausserhofer et al., 2013). A U.S. study did not find such association (Kalisch, Tschannen, Lee, & Friese, 2011). Repeating the 12-country RN4CAST analysis, we also did not find such effects for the organizational construct of nurse education, neither when we linked in to an organizational level construct (hospital or nursing unit) of care left undone nor when we related it to care left undone at the individual nurse level. Moreover, our findings on the interaction between nurse staffing and education also support our belief that evaluating such interaction effects leads to more tangible conclusions than studying the effect of the proportion of bachelor educated nurses on outcomes in multiple regressions also including nurse staffing, but without interaction effect.

Our findings suggest that patient experiences are an indicator of quality deficiencies in structures and processes of nursing care. A reasonable evolution of our findings would be to systematically register care activities left undone, to complement the already widely implemented reports on nursing care performed. This would provide a more complete understanding of care intensity and productivity, and illuminate actionable interventions to improve nurse work environments. Second, quality improvement strategies to efficiently allocating nurses' time and resources should be more widely implemented. Magnet Recognition is an evidence-based intervention shown to improve nurse work environments and produce superior patient outcomes, improved nurse retention, and to have a good return on investment to hospitals (Jayawardhana, Welton, & Lindrooth, 2014; McHugh et al., 2013). The Productive Ward - Releasing Time to Care programme, which is based on lean methodology, is one strategy that is drawing increased attention in this regard. This program aims to provide staff with more time to provide direct patient care and offers 15 modules to optimize the organization of patient care (Wilson, 2009). Another critically acclaimed frontline project is Transforming Care at the Bedside (TCAB), which also aims to increase time spend in direct patient care. TCAB is designed for medical and surgical nursing units in hospitals where it aims to empower nurses in redesigning work processes (The Robert Wood Johnson Foundation & The Institute for Healthcare Improvement, 2007).

Last, this study provides additional support for national and European policy makers to support international research agendas that focus on the quality of human capital in health care. Appropriate steps need to be taken to increase the proportion of hospital nurses with a bachelor's degree, in line with propositions made in the U.S. (Institute of Medicine (U.S.) Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing at the Institute of Medicine, 2011).

Our findings and policy implications assume that the associations we have uncovered are causal, though firm evidence of this is lacking since the data employed are cross-sectional. Omitted variable bias may have occurred by not including elements of nurse wellbeing, which has also been shown to link with the main explanatory variables used in this study and with patient assessments of care. We also did not study whether the same effects persist across morning, day and night shifts, or across shifts with different lengths. While random effects were included for the hospital and country level, we cannot conclude that all findings could exactly be replicated in each country individually.

5.6 Conclusions

This is the first study to explicate within an integrative theoretical and statistical framework that patient assessments of care comprise a critical overview of both the structure and process of nursing care. It is shown that less clinical care activities left undone and higher proportions of bachelor-prepared nurses exert important effects in patients denoting nursing excellence. These findings reinforce the need for nursing management to implement process improvement strategies that aim for nurses spending more time on direct patient care. Hospital hiring policies should reflect the growing body of research associating bachelor-educated nurses with safe, high quality patient care.

References

Aiken, L. H., Cimiotti, J. P., Sloane, D. M., Smith, H. L., Flynn, L., & Neff, D. F. (2011). Effects of Nurse Staffing and Nurse Education on Patient Deaths in Hospitals With Different Nurse Work Environments. Medical Care, 49(12), 1047–1053.

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.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., … Sermeus, W. (2014). Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. Lancet, 383(9931), 1824–30.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., & Sermeus, W. (2013). Nurses' reports of working conditions and hospital quality of care in 12 countries in Europe. International Journal of Nursing Studies, 50(2), 143–53.

Ausserhofer, D., Zander, B., Busse, R., Schubert, M., De Geest, S., Rafferty, A. M., … Schwendimann, R. (2013). Prevalence, patterns and predictors of nursing care left undone in European hospitals: results from the multicountry cross-sectional RN4CAST study. BMJ Quality & Safety.

Ball, J., Murrells, T., Rafferty, A. M., Morrow, E., & Griffiths, P. (2014). "Care left undone" during nursing shifts: associations with workload and perceived quality of care. BMJ Quality & Safety, 23(2), 116–25.

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), 1173–82.

Bruyneel, L., Li, B., Aiken, L. H., Lesaffre, E., Van den Heede, K., & Sermeus, W. (2013). A multi-country perspective on nurses' tasks below their skill level: reports from domestically trained nurses and foreign trained nurses from developing countries. International Journal of Nursing Studies, 50(2), 202–9.

Bruyneel, L., Li, B., Squires, A., Spotbeen, S., Meuleman, B., Lesaffre, E., & Sermeus, W. (2014). Bayesian Multilevel MIMIC Modeling for Studying Measurement Invariance in Cross-group Comparisons. Medical Care.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504.

Chen, G., & Bliese, P. D. (2005). Conceptual framework and statistical procedures for delineating and testing multilevel theories of homology. Academy of Management Annual Meeting Proceedings, 8, A1–A6.

Darby, C., Hays, R. D., & Kletke, P. (2005). Development and evaluation of the CAHPS hospital survey. Health Services Research, 40(6 Pt 2), 1973–6.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. Psychological Methods, 12(1), 1–22.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychological Methods, 12, 121–138.

Griffiths, P. (2009). RN+RN=better care? What do we know about the association between the number of nurses and patient outcomes? International Journal of Nursing Studies, 46(10), 1289–90.

Griffiths, P., Dall'Ora, C., Simon, M., Ball, J., Lindqvist, R., Rafferty, A.-M., … Aiken, L. H. (2014). Nurses' shift length and overtime working in 12 European countries: the association with perceived quality of care and patient safety. Medical Care, 52(11), 975–81.

Griffiths, P., Maben, J., & Murrells, T. (2011). Organisational quality, nurse staffing and the quality of chronic disease management in primary care: observational study using routinely collected data. International Journal of Nursing Studies, 48(10), 1199–210.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18(3-4), 117–44.

Institute of Medicine (U.S.) Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing at the Institute of Medicine. (2011). The Future of Nursing: Leading Change, Advancing Health. Washington, D.C.: The National Academies Press.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. Journal of Applied Psychology, 69(2), 307–321.

Jayawardhana, J., Welton, J. M., & Lindrooth, R. C. (2014). Is there a business case for magnet hospitals? Estimates of the cost and revenue implications of becoming a magnet. Medical Care, 52(5), 400–6.

Jha, A. K., Orav, E. J., Zheng, J., & Epstein, A. M. (2008). Patients' perception of hospital care in the United States. The New England Journal of Medicine, 359(18), 1921–31.

Kalisch, B. J. (2006). Missed nursing care: a qualitative study. Journal of Nursing Care Quality, 21(4), 306–13; quiz 314–5.

Kalisch, B. J., Doumit, M., Lee, K. H., & Zein, J. El. (2013). Missed nursing care, level of staffing, and job satisfaction: Lebanon versus the United States. The Journal of Nursing Administration, 43(5), 274–9.

Kalisch, B. J., Landstrom, G. L., & Hinshaw, A. S. (2009). Missed nursing care: a concept analysis. Journal of Advanced Nursing, 65(7), 1509–17.

Kalisch, B. J., McLaughlin, M., & Dabney, B. W. (2012). Patient perceptions of missed nursing care. Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 38(4), 161–7.

Kalisch, B. J., Tschannen, D., Lee, H., & Friese, C. R. (2011). Hospital variation in missed nursing care. American Journal of Medical Quality : The Official Journal of the American College of Medical Quality, 26(4), 291–9.

Kalisch, B. J., Tschannen, D., & Lee, K. H. (2011). Do staffing levels predict missed nursing care? International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 23(3), 302–8.

Kalisch, B. J., Xie, B., & Dabney, B. W. (2013). Patient-Reported Missed Nursing Care Correlated With Adverse Events. American Journal of Medical Quality : The Official Journal of the American College of Medical Quality.

Kane, R. L., Shamliyan, T. A., Mueller, C., Duval, S., & Wilt, T. J. (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. Medical Care, 45(12), 1195–204.

Kutney-Lee, A., McHugh, M. D., Sloane, D. M., Cimiotti, J. P., Flynn, L., Neff, D. F., & Aiken, L. H. (2009). Nursing: a key to patient satisfaction. Health Affairs (Millwood), 28(4), w669–77.

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in Nursing & Health, 25, 176–188.

Levy, J. A., Landerman, L. R., & Davis, L. L. (2011). Advances in mediation analysis can facilitate nursing research. Nursing Research, 60(5), 333–9.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. Annual Review of Psychology, 58, 593–614.

Mark, B. A. (2006). Methodological issues in nurse staffing research. Western Journal of Nursing Research, 28(6), 694–709.

Mark, B. A., Hughes, L. C., & Jones, C. B. (2004). The role of theory in improving patient safety and quality health care. Nursing Outlook, 52(1), 11–6.

McHugh, M. D., Kelly, L. A., Smith, H. L., Wu, E. S., Vanak, J. M., & Aiken, L. H. (2013). Lower mortality in magnet hospitals. Medical Care, 51(5), 382–8.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika.

Muthén, L. K., & Muthén, B. O. (2012). Mplus User's Guide. Seventh Edition. Losa Angeles, CA: Muthén & Muthén.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. Multivariate Behavioral Research, 42(1), 185–227.

Rogers, A. E., Hwang, W.-T., Scott, L. D., Aiken, L. H., & Dinges, D. F. (2004). The working hours of hospital staff nurses and patient safety. Health Affairs (Project Hope), 23(4), 202–12.

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., … Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nursing, 10, 6.

Sochalski, J. (2004). Is more better?: the relationship between nurse staffing and the quality of nursing care in hospitals. Medical Care, 42(2 Suppl), II67–73.

Squires, A., Aiken, L. H., van den Heede, K., Sermeus, W., Bruyneel, L., Lindqvist, R., … Matthews, A. (2013). A systematic survey instrument translation process for multi-country, comparative health workforce studies. International Journal of Nursing Studies, 50(2), 264–73.

Squires, A., Bruyneel, L., Aiken, L. H., Van den Heede, K., Brzostek, T., Busse, R., … Sermeus, W. (2012). Cross-cultural evaluation of the relevance of the HCAHPS survey in five European countries. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 24(5), 470–5.

Stimpfel, A. W., Sloane, D. M., & Aiken, L. H. (2012). The longer the shifts for hospital nurses, the higher the levels of burnout and patient dissatisfaction. Health Affairs (Project Hope), 31(11), 2501–9.

The Robert Wood Johnson Foundation, & The Institute for Healthcare Improvement. (2007). A New Era in Nursing: Transforming Care at the Bedside.

Vahey, D. C., Aiken, L. H., Sloane, D. M., Clarke, S. P., & Vargas, D. (2004). Nurse burnout and patient satisfaction. Medical Care, 42(2 Suppl), II57–66.

Wilson, G. (2009). Implementation of releasing time to care - The productive ward. Journal of Nursing Management, 17, 647–654.

You, L., Aiken, L. H., Sloane, D. M., Liu, K., He, G., Hu, Y., … Sermeus, W. (2013). Hospital nursing, care quality, and patient satisfaction: cross-sectional surveys of nurses and patients in hospitals in China and Europe. International Journal of Nursing Studies, 50(2), 154–61.

Appendix

A. Three-level mediation model

The equation for the three-level mediation model equals:

$$\text{probit}\left(\text{p}\left(\text{y}_{ijk}=1\middle|\text{x},\text{z},\mathbf{m},\mathbf{c}\right)\right)=\beta_0+\boldsymbol{\beta}_c^T\boldsymbol{c}_{jk}+u_{jk}+u_k+\beta_x x_{jk}+\beta_z z_{jk}+\boldsymbol{\beta}_m^T\boldsymbol{m}_{jk}$$

$$\boldsymbol{\beta}_m^T=(\beta_1,\beta_2\,),\qquad \boldsymbol{m}_{jk}=(m_{1jk},m_{2jk})^T$$

$$u_{jk}\sim N\left(0,\sigma_{hospital}^2\right),\qquad u_k\sim N\left(0,\sigma_{country}^2\right),$$

$$m_{1jk}=\gamma_{10}+\gamma_{1x}x_{jk}+\gamma_{1z}z_{jk}+\varepsilon_{1jk}$$

$$m_{2jk}=\gamma_{20}+\gamma_{2x}x_{jk}+\gamma_{2z}z_{jk}+\varepsilon_{2jk}$$

$$\varepsilon_{1jk}\sim N(0,\sigma_{m1}^2),\qquad \varepsilon_{2jk}\sim N(0,\sigma_{m2}^2)\,,$$

where y_{ijk} represents the binary score measured on patient i from hospital j in country k. The link function is probit() which corresponds to the inverse of the cumulative distribution function of the standard normal distribution. x_{jk} is the nurse staffing variable at the hospital level with coefficient β_x. z_{jk} is the measure of nursing work environment at the hospital level with coefficient β_z. $\boldsymbol{m}_{jk}$ represents the two potential hospital-level missed nursing care mediators with coefficient vector $\boldsymbol{\beta}_m$. $\boldsymbol{c}_{jk}$ represents the vector of all other hospital-level covariates included with $\boldsymbol{\beta}_c$ the coefficient vector. u_{jk} and u_k represent the random effects at the hospital level and country level respectively. In the models of the mediators, x_{jk} and z_{jk} are included with the coefficients γ_x and γ_z respectively.

Replacing m_{1jk} and m_{2jk}, we get:

$$\text{probit}\left(\text{p}\left(\text{y}_{ijk}=1\middle|\text{x},\text{z},\mathbf{m},\mathbf{c}\right)\right)$$

$$=\beta_0+\boldsymbol{\beta}_c^T\boldsymbol{c}_{jk}+u_{jk}+u_k+\beta_x x_{jk}+\beta_z z_{jk}+\boldsymbol{\beta}_m^T\boldsymbol{m}_{jk}$$

$$=\beta_0+\boldsymbol{\beta}_c^T\boldsymbol{c}_{jk}+u_{jk}+u_k+\beta_x x_{jk}+\beta_z z_{jk}+\beta_1 m_{1jk}+\beta_2 m_{2jk}$$

$$= (\beta_0 + \beta_1\gamma_{10} + \beta_2\gamma_{20}) + \boldsymbol{\beta}_c^T\boldsymbol{c}_{ijk} + u_{jk} + u_k + (\beta_x + \beta_1\gamma_{1x} + \beta_2\gamma_{2x})x_{jk} + (\beta_z + \beta_1\gamma_{1z} + \beta_2\gamma_{2z})z_{jk} + (\beta_1\varepsilon_{1jk} + \beta_2\varepsilon_{2jk})$$

β_x is the direct effect of x and $\beta_1\gamma_{1x}$ and $\beta_2\gamma_{2x}$ are its indirect effects via the hospital-level mediators m_1 and m_2 respectively. Similarly, β_z is the direct effect of z and $\beta_1\gamma_{1z}$ and $\beta_2\gamma_{2z}$ are its indirect effects via m_1 and m_2 respectively.

B. Three-level moderated mediation model

The equation for the three-level moderated mediation analysis equals:

$$\text{probit}\left(\text{p}\left(\text{y}_{ijk} = 1\middle|\text{x}, \text{m}, \text{w}, \mathbf{c}\right)\right) = \beta_0 + \boldsymbol{\beta}_c^T\boldsymbol{c}_{jk} + u_{jk} + u_k + \beta_x x_{jk} + \boldsymbol{\beta}_m^T\boldsymbol{m}_{jk} + \beta_w w_{jk} + \beta_{xw}x_{jk}w_{jk}$$

$$m_{jk} = \gamma_0 + \gamma_x x_{jk} + \gamma_w w_{jk} + \gamma_{xw}x_{jk}w_{jk} + \varepsilon_{jk,}$$

where y_{ijk} represents the binary score measured on patient i from hospital j in country k. x_{jk} is the nurse staffing variable at the hospital level with coefficient β_x. $\boldsymbol{m}_{jk}$ represents the two potential hospital-level missed nursing care mediators with coefficient β_m. w_{jk} is the hospital-level moderator with the coefficient β_w. β_{xw} is the interaction effect of x_{jk} and w_{jk}. $\boldsymbol{c}_{jk}$ represents the vector of all other hospital-level covariates included with $\boldsymbol{\beta}_c$ the coefficient vector. u_{jk} and u_k represent the random effects at the hospital level and country level respectively. In the model of the mediator m_{jk}, covariates x_{jk} , w_{jk} and their interaction are included with the respective coefficients γ_x , γ_w and γ_{xw} .

Replacing m_{jk}, we get:

$$\text{probit}\left(\text{p}\left(\text{y}_{ijk} = 1\middle|\text{x}, \text{m}, \text{w}, \mathbf{c}\right)\right) = \beta_0 + \boldsymbol{\beta}_c^T\boldsymbol{c}_{jk} + u_{jk} + u_k + \beta_x x_{jk} + \beta_m m_{jk} + \beta_w w_{jk} + \beta_{xw}x_{jk}w_{jk}$$

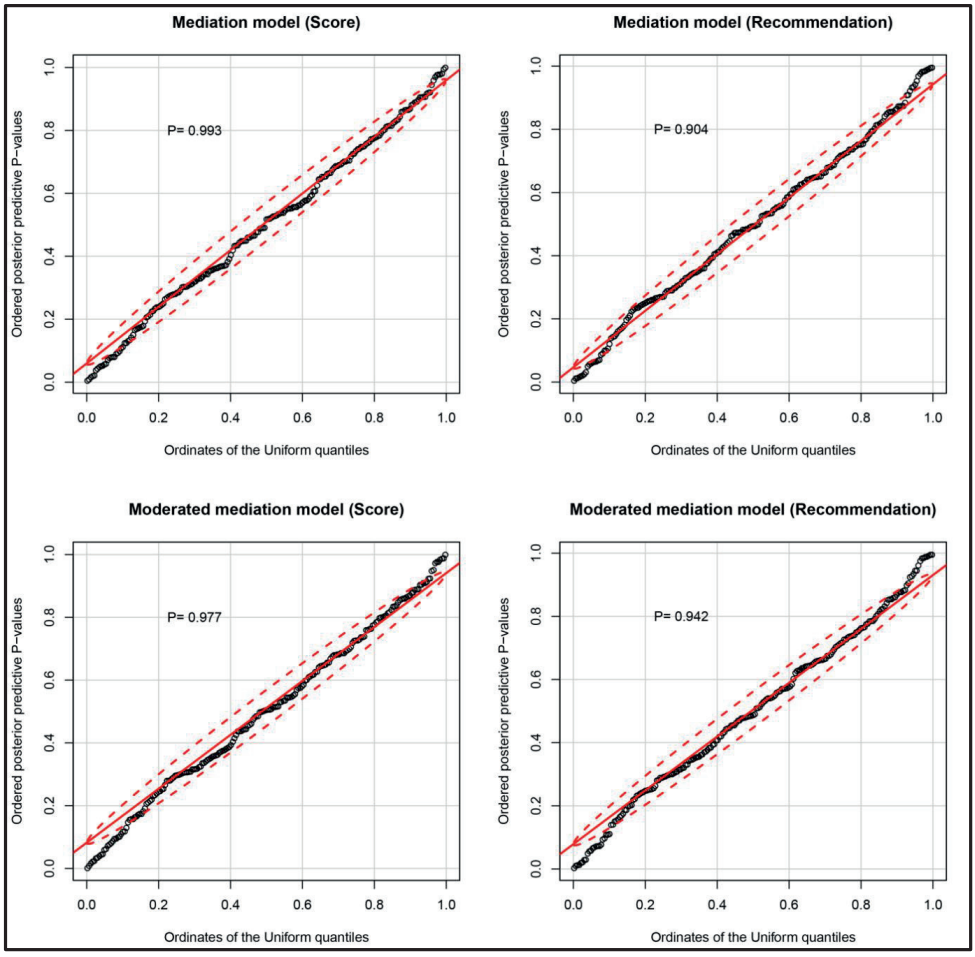
$$= (\beta_0 + \beta_m\gamma_0) + \boldsymbol{\beta}_c^T\boldsymbol{c}_{jk} + u_{jk} + u_k + \left(\beta_x + \beta_{xw}w_{jk} + \beta_m\gamma_x + \beta_m\gamma_{xw}w_{jk}\right)x_{jk} + (\beta_w + \beta_m\gamma_w)w_{jk} + \beta_m\varepsilon_{jk}$$

$\beta_x + \beta_{xw}w_{jk}$ is the direct effect of x with $\beta_{xw}w_{jk}$ the moderator effect, while $\beta_m\gamma_x + \beta_m\gamma_{xw}w_{jk}$ is its indirect effect with $\beta_m\gamma_{xw}w_{jk}$ the moderator effect in the first path.

C. Model fit evaluation

The default non-informative priors in Mplus (Muthén & Muthén, 2012) were used, that is, a normal prior of mean zero and variance 5 for all β parameters ($\beta_0, \boldsymbol{\beta}_c^T, \beta_x, \beta_z, \boldsymbol{\beta}_m^T, \beta_w . \beta_{xw}$), a normal prior of mean zero and variance 10^{10} for all γ parameters ($\gamma_0, \gamma_x, \gamma_z, \gamma_w, \gamma_{xw}$), and an improper inverse gamma prior which corresponds to a uniform distribution of $(0, +\infty)$ for all variance parameters ($\sigma^2_{hospital}, \sigma^2_{country}, \sigma^2_{m1}, \sigma^2_{m2}$). Given the large sample size, the priors are not very crucial in our case. The convergence of all parameters is assessed by the potential scale reduction factor (PSRF) (Gelman & Rubin, 1992) which should be smaller than .01 for all parameters in the model. Model fit was evaluated by using the mixed posterior predictive check (PPC) suggested by Marshall and Spiegelhalter (2003). The mixed PPC first generates the predictive values for the random effects. The final PPC (uses mean as discrepancy function) is subsequently based on the posterior estimates of the fixed effects and the predictive values of the random effects. This is said to be able to overcome the overoptimistic problem of the classic PPC that is based on the posterior estimates of all effects for hierarchical models. We obtain posterior predictive p-values (PPp) for each of the 32 PES-NWI items within each nursing unit. From this, we generate quantile-quantile (QQ) plots of the PPps for each item across all nursing units against the uniform distribution of (0,1). A uniform distribution suggests an acceptable model fit. As displayed in Figure 5.2 , the QQ plots did not violate the uniform distribution.

Figure 5.2. Bayesian multilevel model fit evaluation



Appendix references

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472.

Marshall, E. C., & Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine*, *22*, 1649–1660.

Muthén, L. K., & Muthén, B. O. (2012). Mplus User's Guide. Seventh Edition. Losa Angeles, CA: Muthén & Muthén.

6 BAYESIAN MULTILEVEL MIMIC MODELING FOR STUDYING MEASUREMENT INVARIANCE IN CROSS-GROUP COMPARISONS

This chapter is published as:

Luk Bruyneel, Baoyue Li, Allison Squires, Sara Spotbeen, Bart Meuleman, Emmanuel Lesaffre and Walter Sermeus. Bayesian multilevel MIMIC modeling for studying measurement invariance in cross-group comparisons Medical Care 2014 [Epub ahead of print].

Abstract

Background: Recent methodological advancements should catalyze the evaluation of measurement invariance across groups, which is required for conducting meaningful cross-group comparisons.

Objective: The aim of this study was to apply a state-of-the-art statistical method for comparing latent mean scores and evaluating measurement invariance across managers' and frontline workers' ratings of the organization of hospital care.

Methods: On the 87 nursing units in a single institution, French speaking and Dutch-speaking nursing unit managers' and staff nurses' ratings of their work environment were measured using the multidimensional 32-item practice environment scale of the nursing work index (PES-NWI). Measurement invariance and latent mean scores were evaluated in the form of a Bayesian 2-level multiple indicators multiple causes model with covariates at the individual nurse and nursing unit level. Role (manager, staff nurse) and language (French, Dutch) are of primary interest.

Results: Language group membership accounted for 7 of 11 PES-NWI items showing measurement noninvariance. Cross-group comparisons also showed that covariates at both within-level and between-level had significant effects on PES-NWI latent mean scores. Most notably, nursing unit managers, when compared with staff nurses, hold more positive views of several PES-NWI dimensions.

Conclusions: Using a widely used instrument for measuring nurses' work environment, this study shows that precautions for the potential threat of measurement noninvariance are necessary in all stages of a study that relies on survey data to compare groups, particularly in multilingual settings. A Bayesian multilevel multiple indicators multiple causes approach can accommodate for detecting all possible instances of noninvariance for multiple covariates of interest at the within-level and between-level jointly.

6.1 Introduction

Self-report instruments are frequently used to detect and explain differences between (sub)groups of respondents. A primary methodological issue with this kind of comparative research is that the measurement instrument may not function invariantly across the comparative groups. Measurement invariance pertains to the extent to which respondents across groups perceive and interpret the content of survey instrument items in the same way. Findings of differences between individuals and groups cannot be unambiguously interpreted if evidence of measurement invariance is lacking (Byrne & Watkins, 2003; Horn & McArdle, 1992; Vandenberg & Lance, 2000). Measurement invariance is predominantly studied in a structural equation modeling framework. By testing the equality of parameters of the measurement models for the respective comparison groups, researchers can study the comparability of measurements. A first approach to study these levels of measurement invariance in a structural equation modeling framework is through a multiple group confirmatory factor analytic (MGCFA) model. MGCFA entails the simultaneous analysis of ≥2 measurement models (Joreskog & Goldberger, 1975). Within this framework, several levels of measurement invariance are distinguished. Configural invariance pertains to identical factor structures across groups. Metric invariance refers to the equality of factor loadings across groups. Scalar invariance additionally requires equality of factor loadings and intercepts (thresholds for ordered categorical indicator scores) across groups (Meredith & Teresi, 2006; Meredith, 1993; Steenkamp & Baumgartner, 1998). A second approach is to study measurement invariance in the form of a multiple indicators multiple causes (MIMIC) model (Joreskog & Goldberger, 1975; B. O. Muthén, 1989). This technique is also referred to as confirmatory factor analysis (CFA) with covariates, as ≥1 group variables are treated as covariates. This model involves a single measurement model and input matrix. In a mimic model, cross-group comparisons are performed by including group membership as a covariate in the model. A significant effect of group membership on the latent variable indicates that the factor means are different for different levels of the covariates. Measurement invariance is then examined by estimating direct effects from a covariate on a manifest variable. A covariate with a direct effect on a manifest variable indicates that that manifest variable is not invariant across the levels of the covariate. Advantages of the MGCFA model over the MIMIC model are that it is more flexible in that it allows more parameters

to represent noninvariance. The MIMIC model is, however, more parsimonious and better suited for analyses where the researcher considers many covariates jointly. Other advantages of MIMIC modeling are that it can accommodate continuous predictors and usually has smaller sample size requirements than MGCFA (Brown, 2006).

Health services researchers have previously advocated for the incorporation of measurement invariance detection techniques in a study's analytic strategy (Borsboom, 2006; Teresi, 2006). With few exceptions (Cherepanov et al., 2011; Jones, 2006; Reeve et al., 2007; Rodriguez & Crane, 2011), such evaluations have, however, remained notably absent from the literature. Nonetheless, recent advancements in computational statistics for measurement invariance evaluation have greatly enhanced the feasibility of studies focusing on this important part of psychometric evaluation of measurement tools in fields like health services research. These advances have introduced important concepts that extend statistical techniques for measurement invariance evaluation to a multilevel (Asparouhov & Muthén, 2012) and Bayesian (B. O. Muthén & Asparouhov, 2012) framework. This is of significant relevance to health services research, where study designs often involve complex survey data. Health services researchers should therefore feel encouraged to incorporate statistical techniques that detect measurement invariance in their work.

The focus of this study is the case of managers' and frontline workers' views about features of the organization of hospital care. Studying measurement invariance is illustrated using a state-of-the-art statistical approach in the form of a Bayesian 2-level MIMIC model.

6.2 Methods

6.2.1 Study Sample

We analyze a convenience sample of one large non-academic Belgian hospital entity consisting of 118 patient care units. The study utilized extensively validated instruments from the Registered Nurse Forecasting (RN4CAST) study, a large European nurse workforce study (Sermeus et al., 2011). Dutch and French versions of the questionnaire, with translations previously validated (Squires, Aiken, et al., 2012),

were distributed to all 118 nursing unit managers and their staff nurses. A nursing unit manager is defined as a registered nurse who is largely responsible for administratively and clinically managing the nursing staff on a single nursing unit. Only responses from units where both the manager and staff responded to the survey are included in this analysis. Eighty-seven of 118 nurse managers completed the questionnaire (75.0% response rate). On these 87 nursing units, 821 of 1,159 staff nurses completed the questionnaire (70.8% response rate). The final dataset consists of 908 observations.

6.2.2 Measures

Nursing unit managers' and staff nurses' perceptions of the nursing work environment were measured using the practice environment scale of the nursing work index (PESNWI). This instrument consists of 32 statements measured on a 4-point Likert scale anchored between 1='completely disagree' to 4='completely agree'. In the factor analytic stage of the development of the PES-NWI, 5 dimensions of the nursing work environment were proposed. These include collegial nurse-physician relations; nurse managerial abilities, leadership, and support of nurses; nurse participation in hospital affairs; nursing foundations for quality of care; and staffing and resource adequacy (Lake, 2002).

6.2.3 Statistical Methods

First, we obtained the intraclass correlation (ICC) coefficient for the 32 PES-NWI items. This reflects the proportion of a single variable's variance that can be accounted for by each level. From this, we calculated the design effect as $\varphi = 1 + (\bar{n} - 1) * \mathrm{ICC}$, where n equals the average group size. Twenty-four of 32 items had a design effect >2, which indicates that the clustering in the data needs to be taken into account during estimation (Muthén & Satorra, 1995).

Second, we adopted a 2-level exploratory factor analytic model (EFA) to understand the factor structure underlying the PES-NWI items. The multilevel approach accounts for nurses being clustered in nursing units. We obtained a clear interpretable and parsimonious EFA solution. The within-level part of the model describes the factor structure for how the nurses' PES-NWI item perceptions covary within nursing units. The 6 within-level factors are: career development and opportunities (3 items), collegial nurse-physician relations (7 items), nurse staffing (2 items), frontline nurse management (2 items),

support for nurses (4 items), and nursing foundations for quality of care (4 items). Ten of 32 items are not included in any within-level dimension. The between-level part of the model describes the factor structure for how the nursing unit PES-NWI item means covary. The between-level factors are: career development, opportunities, and support for nurses (11 items), collegial nurse-physician relations (7 items), and frontline nurse management, nurse quality foundations, and staffing and resources adequacy (11 items). Thus, at the nurse level, several underlying dimensions are more clearly manifested than at the nursing unit level, where more general factors appear. The only exception is the factor of collegial nurse-physician relations, which is identical at both levels. Evaluation of good EFA model fit between the hypothesized model and the observed data was based on the work of Hu and Bentler (1999).

Third, we evaluated latent mean scores and measurement invariance of the PES-NWI across ratings provided by staff nurses and their managers. Jak et al. (2010) advise researchers investigating measurement bias to include as many possible violator variables as available. Next to role (1=manager, 0=frontline worker), covariates included in the analysis are nurses' language (0=French, 1=Dutch), degree (0=diploma degree, 1=bachelor degree), type of employment (0=part-time, 1=full-time), migratory status (0=domestically trained, 1=foreign trained), and sex (0=female, 1=male). For this reason, a mimic model is preferred over MGCFA because it can include many covariates jointly. However, as a preliminary step, using MGCFA, we conducted tests of configural and metric invariance for the suggested EFA factor solution for the values (groups) of the 6 covariates separately in order for the MIMIC model, which only studies a higher level of invariance (threshold invariance), to be meaningful. Findings provided evidence of configural and metric invariance. Model fit evaluation was based on Chen's (2007) suggestions on changes allowed in goodness-of-fit indexes when testing for invariance at different levels. Next, we specify a 2-level confirmatory factor analytic (CFA) model based on findings from the EFA model. Last, the MIMIC model is established by regressing the latent and observed variables from the CFA model on covariates. An effect of any of the 6 covariates on any of the 6 PES-NWI latent variables would indicate that the latent variable mean is different for different levels of the covariate. A direct effect of any of the 6 covariates on any of the 32 PES-NWI items, over and above the indirect effect via the factors, would indicate that a particular item does not behave the same for different levels of the covariate. Again, we

apply a multilevel approach that accounts for nurses being clustered in nursing units. The equation for the multilevel MIMIC model is:

$$\begin{gathered}\boldsymbol{y}_{ij}^{*}=\boldsymbol{B}\boldsymbol{x}_{ij}+\Lambda_{B}\boldsymbol{\eta}_{j}+\boldsymbol{u}_{j}+\Lambda_{W}\boldsymbol{\eta}_{ij}+\boldsymbol{\varepsilon}_{ij},\\ \boldsymbol{\eta}_{ij}=\boldsymbol{\Gamma}\boldsymbol{x}_{ij}+\boldsymbol{\delta}_{ij},\\ \boldsymbol{\eta}_{j}\sim N(0,\Sigma_{B}),\ \boldsymbol{u}_{j}\sim N\big(0,diag(\sigma_{u1}^{2},\sigma_{u2}^{2},\dots,\sigma_{uP}^{2})\big),\\ \boldsymbol{\delta}_{ij}\sim N(0,\Sigma_{w}),\ \boldsymbol{\varepsilon}_{ij}\sim N(0,\boldsymbol{I}_{P}),\end{gathered} \tag{6.1}$$

where $\boldsymbol{y}_{ij}^{*}$ represents a vector of P items coming from the ith nurse in the jth nursing unit. It is the underlying continuous measure of the observed ordinal response $\boldsymbol{y}_{ij}$ with the following relationship (let y_{ij}^{p} and y_{ij}^{*p} represent the pth element in $\boldsymbol{y}_{ij}$ and $\boldsymbol{y}_{ij}^{*}$ respectively):

$$y_{ij}^{p}=\begin{cases}1, & if\ y_{ij}^{*p}<\alpha_{1}^{p}\\ 2, if\ \alpha_{1}^{p}<y_{ij}^{*p}<\alpha_{2}^{p}\\ 3, if\ \alpha_{2}^{p}<y_{ij}^{*p}<\alpha_{3}^{p}\\ 4, & if\ \alpha_{3}^{p}<y_{ij}^{*p}\end{cases}$$

where αs are the thresholds. B is a $P\times Q$ matrix of the direct effects associated with the Q-dimensional vector $\boldsymbol{x}_{ij}$, while Γ is a $m_{W}\times Q$ matrix of the indirect effects associated with the same Q-dimensional vector $\boldsymbol{x}_{ij}$ where m_{W} is the number of common factors at the nurse level. $\boldsymbol{\eta}_{j}$ is the m_{B}-dimensional nursing unit level common factor following a multivariate normal distribution with mean zero and a covariate matrix Σ_{B}, and Λ_{B} is its $P\times m_{B}$ loading matrix with m_{B} the number of common factors at the nursing unit level. $\boldsymbol{\eta}_{ij}$ is the m_{W}-dimensional nurse level common factors with a multivariate normal distribution with mean zero and a covariance matrix Σ_{w}, and Λ_{W} is its loading matrix with a $P\times m_{W}$ dimension. $\boldsymbol{u}_{j}$ is the P-dimensional random intercept with independent normal distribution for each element with mean zero and a variance σ_{up}^{2}, while ε_{ij} is the P-dimensional residual, each element of which follows a standard normal distribution.

We adopted a Bayesian structural equation modeling approach for the proposed multilevel MIMIC model (Equation 6.1) (Muthén & Asparouhov, 2012). Measurement invariance for the PES-NWI has not been previously investigated. Freeing one parameter at a time may lead to a long search, whereas

allowing all direct effects from the 6 covariates to the 32-factor indicators to be freely estimated in a frequentist analysis, the model would not be identified. A Bayesian approach provides an alternative in that it identifies the model by introducing informative small-variance priors in the measurement part (cross-loadings in the CFA model) as well as in the structural part (direct effects from the covariates to the factor indicators).

Following an example of Bayesian MIMIC analysis by Muthén and Asparouhov (2012), normal priors with mean zero and variance .01 for all cross-loadings and variance .04 for the direct effects are considered a good starting point for the proposed MIMIC model. Smaller variances may produce worse model fit values and larger variances may lead to nonidentification. Similar to Muthén and Asparouhov (2012), we therefore studied sensitivity in the results by varying the cross-loadings from .01 to .10 for our data. Except for the direct effects and cross-loadings, the default noninformative priors in Mplus were used (Muthén & Muthén, 2012). Model fit was evaluated by using the mixed posterior predictive check (PPC) suggested by Marshall and Spiegelhalter (2003). The mixed PPC first generates the predictive values for the random effects. The final PPC (uses mean as discrepancy function) is subsequently based on the posterior estimates of the fixed effects and the predictive values of the random effects. This is said to be able to overcome the overoptimistic problem of the classic PPC that is based on the posterior estimates of all effects for hierarchical models (Marshall & Spiegelhalter, 2003) We obtain posterior predictive p-values (PPp) for each of the 32 PES-NWI items within each nursing unit. From this, we generate quantile-quantile (QQ) plots of the PPps for each item across all nursing units against the uniform distribution of (0,1). A uniform distribution suggests an acceptable model fit. All statistical analyses were conducted using Mplus version 7 (Muthén & Muthén, 2012) except for model fit evaluation, which was performed by the Markov Chain Monte Carlo software JAGS (Plummer, 2003) on the Markov Chain Monte Carlo data extracted from the Mplus output.

6.3 Results

6.3.1 Sample Characteristics

The response rate at the unit level varied from 23.1% to 100%. Seventy-eight of 87 nursing units had a response rate >50%. The number of staff nurses per nursing unit varied from 3 to 27 (mean = 9.4, median = 9). Table 6.1 displays the personal characteristics for nursing unit managers and staff nurses. The majority of nurses was female, spoke French, had obtained a bachelor degree, were domestically educated in Belgium, and worked full-time. A considerably higher proportion of nursing unit managers, when compared with staff nurses, worked full-time.

Table 6.1 Sample Characteristics

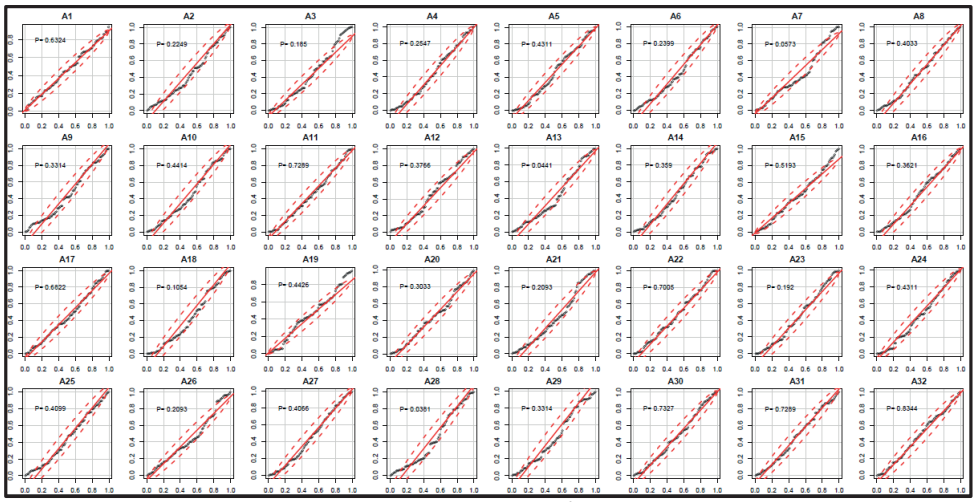
Parameter		Nurse managers (n=87)	Staff nurses (n=821)
Language	French	75.86% (n=66)	86.36% (n=709)
	Dutch	24.14% (n=21)	13.64% (n=112)
	Missing	-	-
Degree	Diploma degree	5.75% (n=5)	18.76% (n=154)
	Bachelor degree	90.80% (n=79)	74.06% (n=608)
	Missing	3.45% (n=3)	7.19% (n=59)
Type of employment	Part-time	8.05% (n=7)	38.37% (n=315)
	Full-time	91.95% (n=80)	58.95% (n=484)
	Missing	-	2.68% (n=22)
Migratory status	Domestically trained	98.85% (=86)	85.38% (n=701)
	Foreign trained	1.15% (n=1)	10.23% (n=84)
	Missing	-	4.38% (n=36)
Gender	Female	89.66% (n=78)	83.43% (n=685)
	Male	9.20% (n=8)	13.89% (n=114)
	Missing	1.15% (n=1)	2.68% (n=22)

6.3.2 The MIMIC Model

First, we describe model fit evaluation for the proposed Bayesian multilevel MIMIC model as this determined model selection. For our initial model described above, 9 of the 32 PES-NWI item QQ plots violated the uniform distribution with *p*-values <.05. We therefore included aggregated nursing unit level covariates to the MIMIC model. We estimated both the effects of these covariates to the 3 between-level PES-NWI latent variables as well as the effects to the 32 items in the between-level part of the model. Covariates included are the nursing unit level proportion of: nurses with a bachelor degree, nurses speaking Dutch, nurses working full-time, foreign trained nurses, and male nurses. We also controlled for

the size of the nursing unit (number of nurses in the nursing unit). For this new model, only 2 PES-NWI item QQ plots violated the uniform distribution with *p*-values slightly below .05, as displayed in Figure 6.1. This suggests that this model performed better than the initial model with within-level effects to the PES-NWI latent variables and items only. The sensitivity analysis for this model with both within-level and between-level effects suggested that the hypothesized pattern of cross-loadings did not substantially differ for changes in the prior variance (Table 6.2).

Figure 6.1 Bayesian 2-level MIMIC model with covariates at the within-level and between-level. QQ plots of the PP*ps* for each PES-nWI item across all nursing units against distribution of (0,1)



MIMIC indicates multiple indicators multiple causes. PES-NWI, Practice Environment Scale of the Nursing Work Index; PP*ps*, posterios predictive *p*-values; QQ, quantile-quantile plots.

Table 6.2 Bayesian two-level MIMIC model with covariates at the within-level and between-level: effects of using different variances for the informative priors of the cross-loadings and direct effects

			Within level			Between level		
Prior variance	95% cross-loading limit	Items with a significant posterior predictive *p*-value	Significant cross-loadings	Significant effect on PES-NWI items	Significant effects on PES-NWI latent variables	Significant cross-loadings	Significant effect on PES-NWI items	Significant effects on PES-NWI latent variables
Cross-loading (direct effect fixed at .04)								
.01	.20	2	27	12	5	1	0	4
.02	.28	2	25	9	5	1	0	3
.03	.34	3	26	10	5	1	0	3
.04	.39	3	25	9	4	1	0	3
.05	.44	2	25	9	4	1	0	3
.06	48	1	26	10	4	1	0	3
.07	.52	2	25	9	4	1	0	3
.08	.55	0	25	9	4	1	0	3
.09	.59	1	25	9	4	1	0	3
.10	.62	1	24	8	4	1	0	3
Direct effect (cross-loading fixed at .01)								
.01	.20	3	29	3	7	1	2	4
.02	.28	3	28	6	5	1	0	4
.03	.34	2	27	8	5	1	0	4
.04	.39	2	27	12	5	1	0	4
.05	.44	2	27	12	5	1	0	4
.06	.48	0	26	12	5	1	0	4
.07	.52	3	27	12	5	1	0	4
.08	.55	2	27	11	5	1	0	4
.09	.59	1	27	13	5	1	0	4
.10	.62	2	27	12	4	1	0	4

The number of identified effects from the covariates on the PES-NWI items and PES-NWI factors remained stable when the priors for the direct effects had at least a variance of 04 and cross-loadings were fixed at .01. For this reason, the final model was a Bayesian multilevel MIMIC model with covariates at the within-level and between-level and prior variances of .01 and .04 for the cross-loadings and direct effects, respectively. The coefficients for this model were standardized using the variances of the continuous latent variables. Findings should be interpreted as the change in *y* in *y* SD units when *x* changes from 0 to 1 (Muthén & Muthén, 2012). Significant effects, meaning that the Bayesian credibility interval did not cover 0, are marked with an asterisk. The findings for the CFA part of the MIMIC model are displayed in Table 6.3. Inclusion of the covariates did not cause distortion in the solution. The upper part of Table 6.4 shows the effects of the covariates to the latent mean scores. At the within-level, nursing unit managers perceived 4 of 6 latent nurse work environment variables significantly more positive than staff nurses. Foreign trained nurses had significantly higher ratings of career development and opportunities than domestically trained nurses. Similarly, at the between-level, in nursing units with a

higher proportion of foreign trained nurses, there were higher ratings on career development, opportunities, and support for nurses. The inverse is true for nursing units with a higher proportion of nurses working fulltime. Further in these units, nurses tend to give lower scores on the factor of frontline nurse management, nurse quality foundations, and staffing and resources adequacy. Nurses in units with a higher proportion of male nurses in general had better perceptions of collegial nurse-physician relations. The lower part of Table 6.4 showed evidence of measurement noninvariance for 11 PES-NWI items. There were 12 instances of direct effects from the covariates to the PESNWI items at the within-level, of which 7 related to language group membership. The interpretation is as follows. For example, the negative direct effect of the item 'Adequate support services allow me to spend time with my patients' on the covariate language means that for a given factor value, Dutch-speaking nurses have a lower probability of agreeing to this item than French-speaking nurses. No evidence of measurement noninvariance was found at the between-level. These estimates are not included in Table 6.4.

Table 6.3 Bayesian two-level MIMIC model with covariates at the within-level and between-level: Confirmatory factor analysis

Parameters		Within level loadings						Between level loadings		
		Career development and opportunities	Collegial nurse-physician relations	Nurse staffing	Frontline nurse management	Support for nurses	Nursing foundations for quality of care	Career development, opportunities, and support for nurses	Collegial nurse-physician relations	Frontline nurse management, nurse quality foundations, and staffing and resource adequacy
1	Adequate support services allow me to spend time with my patients.	.037	.013	.177*	.021	.083	.038	-.059	-.212	**.841***
2	Physicians and nurses have good working relationships.	-.030	**.758***	-.097	.023	.049	-.057	.022	**1.004***	-.011
3	A supervisory staff that is supportive of nurses.	.022	-.033	-.059	.051	**.891***	-.150*	**1.097***	.085	.018
4	Active staff development or continuing education programs for nurses.	**.591***	-.039	-.090	.109	-.009	.106	**.983***	.024	.028
5	Career development/clinical ladder opportunity.	**1.001***	-.027	-.028	-.038	-.020	-.041	**.810***	-.031	-.024
6	Opportunity for registered nurses to participate in policy decisions.	.142*	.062	.183*	.076	.323*	-.021	**1.158***	.212	-.025
7	Physicians value nurses' observations and judgments.	.024	**.712***	.069	.002	.003	-.065	.020	**.961***	-.002
8	Enough time and opportunity to discuss patient care problems with other nurses.	.030	.215*	.324*	.056	.049	.054	-.053	.116	**.916***
9	Enough registered nurses on staff to provide quality patient care.	-.002	-.068	**.925***	.036	-.114	-.012	.002	-.007	**1.035***
10	A nurse manager who is a good manager and leader.	.000	-.026	.011	**.856***	.041	.014	-.016	.004	**.764***
11	A chief nursing officer who is highly visible and accessible to staff.	-.033	.021	.008	.098	**.751***	-.159*	**1.170***	-.113	-.030
12	Enough staff to get the work done.	-.058	-.015	**.878***	-.075	.012	-.039	.027	.024	**1.011***
13	Physicians recognize nurses' contributions to patient care.	-.026	**.870***	.085	.070	-.047	-.124*	.006	**1.001***	.004
14	Praise and recognition for a job well done.	.092*	.207*	.191*	.064	.188*	.099	.097	.266*	.059
15	High standards of nursing care are expected by the management.	.043	.044	-.018	.104	.098	.235*	**.685***	-.207	-.038
16	A chief nursing officer is equal in power and authority to other top level hospital executives.	.042	.006	.102	.036	.218*	.054	**.787***	-.084	-.024
17	A lot of team work between nurses and physicians.	.028	**.747***	.000	.004	-.031	-.037	-.031	**1.006***	-.018
18	Opportunities for advancement.	**.780***	.036	.046	.050	.006	-.006	**1.044***	-.049	.024
19	A clear philosophy of nursing that pervades the patient care environment.	.065	.057	.099	.115*	.137*	.314*	.072	-.109	**.562***
20	Working with nurses who are clinically competent.	.037	.172*	.119	.048	-.045	.279*	-.099	.165	**.591***
21	Physicians respect nurses as professionals	-.029	**.838***	-.076	-.006	-.045	.064	-.030	**1.001***	.012
22	A nurse manager who backs up the nursing staff in decision making, even if the conflict is with a physician.	.008	.021	-.045	**.758***	-.001	.070	-.030	-.001	**.763**
23	Management that listens and responds to employee concerns.	-.061	-.041	-.050	-.005	**.931***	-.031	**.198***	.002	.007
24	An active quality assurance program.	.083	.028	.054	-.108	.310*	.311*	**.608***	-.119	.052
25	Registered nurses are involved in the internal governance of the hospital (e.g., practice and policy committees).	-.015	-.017	.035	-.153*	**.628***	.142	.063	-.166	.064
26	Collaboration between nurses and physicians.	.008	**.881***	-.066	-.054	-.050	.060	.006	**.997***	-.005

Parameters		Within level loadings						Between level loadings		
		Career development and opportunities	Collegial nurse-physician relations	Nurse staffing	Frontline nurse management	Support for nurses	Nursing foundations for quality of care	Career development, opportunities, and support for nurses	Collegial nurse-physician relations	Frontline nurse management, nurse quality foundations, and staffing and resource adequacy
27	A preceptor program for newly hired nurses.	-.038	-.088	.014	.021	-.127	**.911***	.045	-.046	.020
28	Nursing care is based on a nursing rather than a medical model.	-.073	.043	-.062	.091	-.074	**.572***	-.005	-.092	**.604***
29	Registered nurses have the opportunity to serve on hospital and nursing committees.	.063	-.005	.036	-.187*	.318*	.336*	**.613***	.012	-.011
30	Physicians hold nurses in high esteem.	-.018	**.603***	-.009	-.087	.162*	.061	-.013	**.989***	.022
31	Written, up-to-date care plans for all patients.	-.011	-.013	-.054	.066	-.118	**.628***	-.014	-.087	**.300**
32	Patient care assignments that foster continuity of care (i.e., the same nurse cares for the patient from one day to the next).	.044	.027	-.090	-.011	-.064	**.627***	.051	.016	**.225**

Factor correlations (within level)

Parameter	F1	F2	F3	F4	F5	F6
F1	1.000					
F2	.387*	1.000				
F3	.406*	.418*	1.000			
F4	.336	.416*	.317	1.000		
F5	.681*	.512*	.552*	.383	1.000	
F6	.610*	.555*	.408*	.373	.691*	1.000

Factor correlations (between level)

Parameter	F1	F2	F3
F1	1.000		
F2	-.072	1.000	
F3	.139	.334*	1.000

The bold entries are loadings that are largest for the item. An additional requirement for the entry to be bolded is that no severe cross-loadings can be present. MIMIC indicates multiple indicators multiple causes.

Table 6.4 Bayesian two-level MIMIC model with covariates at the within-level and between-level: Effects of covariates

Parameter		Role	Language	Degree	Working percentage	Migratory status	Gender	Unit Size
Factors (within-level)		Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Career development and opportunities		.138*	.026	-.061	.034	.101*	-.059	—
Collegial nurse-physician relations		.130*	-.011	.004	.053	-.014	-.047	—
Nurse staffing		.083	.031	-.084	.039	.039	.035	—
Frontline nurse management		.070	.043	-.035	-.011	-.016	.094	—
Support for nurses		.225*	.037	-.030	.029	.076	-.010	—
Nursing foundations for quality of care		.132*	-.086	-.064	.084	.041	-.012	—
Factors (between-level)								
Career development, opportunities, and support for nurses		—	-.028	.062	-.185*	.232*	-.187	.472
Collegial nurse-physician relations		—	-.071	.124	-.186	-.076	.253*	-.001
Frontline nurse management, nurse quality foundations and staffing and resources adequacy		—	-.117	.077	-.225*	.013	.110	.202
Items (within-level)								
1	Adequate support services allow me to spend time with my patients.	-.011	-.072*	.013	.029	-.032	.033	—
2	Physicians and nurses have good working relationships.	-.029	.003	-.002	-.031	-.044	-.019	—
3	A supervisory staff that is supportive of nurses.	.037	.140*	.007	.002	-.054*	.018	—
4	Active staff development or continuing education programs for nurses.	.013	-.077*	.042	.006	-.007	-.036	—
5	Career development/clinical ladder opportunity.	-.001	-.002	.000	-.019	.005	.004	—
6	Opportunity for registered nurses to participate in policy decisions.	.019	-.060	-.040	.011	.010	.001	—
7	Physicians value nurses' observations and judgments.	-.007	.026	.032	-.006	-.030	.040	—
8	Enough time and opportunity to discuss patient care problems with other nurses.	-.043	.049	.004	-.032	-.032	.008	—
9	Enough registered nurses on staff to provide quality patient care.	-.014	.020	.004	.006	.002	-.018	—
10	A nurse manager who is a good manager and leader.	-.016	.002	-.009	-.017	-.006	-.002	—
11	A chief nursing officer who is highly visible and accessible to staff.	.006	-.079*	.006	.021	.009	.001	—
12	Enough staff to get the work done.	.022	.012	.008	.003	-.001	.012	—
13	Physicians recognize nurses' contributions to patient care.	.024	.038	-.012	-.001	-.004	.024	—
14	Praise and recognition for a job well done.	-.007	.014	-.015	-.013	.024	.009	—
15	High standards of nursing care are expected by the management.	.062	.080*	.029	-.054	-.034	.025	—
16	A chief nursing officer is equal in power and authority to other top level hospital executives.	.033	.041	.005	-.041	.014	-.073*	—
17	A lot of team work between nurses and physicians.	-.022	-.004	.006	.033	-.012	.013	—
18	Opportunities for advancement.	.003	.029	-.016	.044	-.009	.003	—
19	A clear philosophy of nursing that pervades the patient care environment.	-.002	.029	-.019	-.017	-.047	-.019	—
20	Wore with nurses who are clinically competent.	-.019	-.058	.004	.001	.002	-.038	—
21	Physicians respect nurses as professionals	.003	-.060*	-.014	.001	.010	-.025	—
22	A nurse manager who backs up the nursing staff in decision making, even if the conflict is with a physician.	.003	-.013	.018	.018	.040	.008	—
23	Management that listens and responds to employee concerns.	-.009	-.073*	.015	.021	.021	-.014	—

Parameter		Role	Language	Degree	Working percentage	Migratory status	Gender	Unit Size
24	An active quality assurance program.	-.026	.014	.008	-.030	.060*	.014	—
25	Registered nurses are involved in the internal governance of the hospital (e.g., practice and policy committees).	-.011	-.032	-.019	-.036	.021	.012	—
26	Collaboration between nurses and physicians.	.009	.002	.018	.014	.027	-.030	—
27	A preceptor program for newly hired nurses.	.044	.028	-.030	.017	.011	.039	—
28	Nursing care is based on a nursing rather than a medical model.	.034	.019	.014	.021	-.065*	-.059	—
29	Registered nurses have the opportunity to serve on hospital and nursing committees.	-.079*	.000	.002	.018	.020	-.057	—
30	Physicians hold nurses in high esteem.	.006	-.013	-.025	-.023	.033	.033	—
31	Written, up-to-date care plans for all patients.	-.023	-.043	.007	-.051	-.013	.036	—
32	Patient care assignments that foster continuity of care (i.e., the same nurse cares for the patient from one day to the next).	-.035	.000	.037	.059	-.027	.002	—

Dashes indicate that this effect was not estimated. MIMIC indicates multiple indicators multiple causes

6.4 Discussion

6.4.1 Findings on Measurement Invariance

We demonstrated a process of measurement invariance evaluation in a cross-sectional study by applying a modern statistical technique to a widely used instrument for measuring nurses' work environments. A multilevel MIMIC analysis in a Bayesian framework revealed evidence of measurement noninvariance. Our main interest lies in evaluating the impact of measurement noninvariance in interpreting the 4 PES-NWI latent means for which nurse managers have significantly higher scores than staff nurses. As derived by Sass (2011) from the work of Cheung and Rensvold (1999) and Millsap and Kwok (2004), there are 3 appropriate options for how to deal with noninvariant measures: (a) only use the invariant items; (b) apply a partial measurement invariance model; (c) assume that for the items exhibiting measurement noninvariance, the differences are too small to influence the results and proceed using all the items. Sass states that "option 'c' is feasible when the degree of measurement noninvariance is minimal and the majority of items are invariant". For the 4 factors of interest, the ratings for all PES-NWI items were invariant across frontline workers and managers. We therefore conclude that the 4 latent means for which nurse managers have higher scores than staff nurses represent true differences, meaning that nurse managers have a higher probability of holding more positive views of career development and opportunities, collegial nurse-physician relations, support for nurses, and nursing foundations for quality of care.

6.4.2 Clinical Relevance

Many valuable insights about the organizational dynamics behind nursing role implementation in the acute care setting have resulted from research on nurses' work environment using the PES-NWI. Good hospital work environments are associated with higher rates of nurse wellbeing (Aiken et al., 2012), superior patient experiences with hospital care (Aiken et al., 2012; Kutney-Lee et al., 2009), and lower patient in-hospital mortality (Aiken et al., 2011; McHugh et al., 2013). This study extends nursing work environment research by using the instrument among nurse managers for the first time and simultaneously evaluating measurement invariance of the instrument across ratings provided by nurse

managers and staff nurses. In this study we found evidence of differences in staff nurses' and nursing unit managers' perceptions of the nursing work environment. That finding supports similar research in other fields that cites a disconnection between management and frontline workers. Other hospital-based organizational research analyzing both frontline workers' and managers' views also indicates the importance of consulting both frontline staff as well as managers in organizational decision making. For example, Price et al. (2007) showed that clinical nurses and their managers offered divergent views of deficiencies in quality improvement implementation. Studies including several types of hospital staff found that managers overall, when compared with frontline workers, have a more positive perception of the patient safety climate (Singer, Falwell, Gaba, & Baker, 2008) and perceive greater improvements in patient safety culture but lower improvements on the timeliness of care delivery (Parand et al., 2011). Our findings support endeavors to better understand the roots of differing views between nursing unit managers' and staff nurses' perceptions of their work environment and other organizational features of hospital care. Highly divergent views between frontline workers and management could destabilize work places, contribute to negative work environments, increase turnover rates, and could ultimately hinder any possible solutions and strategic direction to issues raised by frontline nurses to management. In addition, we found statistically significant differences in PES-NWI ratings across males and females, domestically and foreign trained nurses, and fulltime and part-time working nurses. This indicates that action plans on improving work environments should be tailored for specific groups.

6.4.3 Limitations

There are a number of methodological challenges that could affect the interpretation of our findings. First, with regard to the interpretation of the EFA model, we used Hu and Bentler's (1999) recommendations for cut-off values; however, there is still disagreement in setting acceptable levels of fit in goodness-of-fit indexes (Marsh, Hau, & Wen, 2004). Second, although our factor analytic models provided clearly interpretable factors, 2 of our factors consisted of 2 indicators only. To encompass the scope of the construct, Tabachnick and Fidell (Tabachnick & Fidell, 1983) recommend at least 3 indicators per factor as the interpretation of the factor defined by <3 variables might be hazardous. It has, however, also been stated that 2 items per factor can be sufficient and this is widely applied (Kenny & McCoach,

2003). More important is that the number of included items encompasses the scope of the construct (Bandalos & Finney, 2010). Although specifying only 2 variables per factor does not necessarily affect model identification, it might be preferable to have a larger number of variables per factor to better encompass the scope of the work environment construct. For the factor 'nurse staffing', only the 2 items that clearly referred to having enough nurses had high factor loadings. In Lake's development of the PES-NWI factor loadings for these 2 items were also much higher than factor loadings for the other 2 items. The same is true for the 2 items with highest salience on the factor of 'frontline nurse management'. Although the conceptual interpretation for these factors can be viewed as perfect subfactors of Lake's original factors, both might benefit from adding additional items.

6.4.4 Recommendations for Research

Two important recommendations emerge from this study for both analytic methods in cross-sectional organizational analyses and cross-cultural research. First, the presence of language-specific measurement noninvariance is of particular interest. Health services researchers examining measurement invariance across language groups for other survey instruments previously reported considerable evidence of measurement noninvariance (Jones, 2006; Ramirez, Teresi, Holmes, Gurland, & Lantigua, 2006). Ramirez et al (2006) specifically pointed to the need for "qualitative methods to address the conceptual adequacy and equivalence of items within and across populations". Such extensive cross-cultural content validation is exactly what the survey instruments underwent in the RN4CAST study (from which the questionnaires used in this study originated). In that study, the translation process standardized the interpretation of the items across all 12 participating countries (Squires et al., 2013). A translation manager was appointed to ascertain high standards of instrument translation that reduce item bias. In addition to forward and back translation techniques (Brislin, 1970), dissimilarity of constructs was assessed in the investigated countries through the application of content validity indexing procedures. This means that expert review panels evaluated cross-cultural relevance, including translation, of the survey instruments, allowing the research team to correct translations as appropriate. Our findings on measurement noninvariance across language groups indicate that, despite this process, additional efforts to reduce measurement noninvariance should be considered. An additional tool for examining

measurement invariance worth considering in studies that could be affected by cultural dynamics is the technique of anchoring vignettes (King, Murray, Salomon, & Tandon, 2004). These are descriptions of hypothetical people or situations included in data collection. Anchoring vignettes provide a common scale of measurement, which could account for response category differences. As such, they improve the problems of interpersonal and cross-cultural incomparability in survey research. Anchoring vignettes are highly recommended for health services research (Johnson, 2006) and positive experiences with this technique have been reported by various authors (Salomon, Tandon, & Murray, 2004; Van Soest, Delaney, Harmon, Kapteyn, & Smith, 2011). Second, mainly researchers interested in country comparisons have studied the effect of higher level covariates in explaining measurement noninvariance. For example, cross-cultural researchers recently used data from the European Social Survey in 26 countries and demonstrated how measurement noninvariance can be explained by including country-level covariates in multilevel structural equation modeling (Davidov, Dulmer, Schluter, Schmidt, & Meuleman, 2012). In our study, including both within-level and between-level covariates improved model fit. Although between-level covariates did not show evidence of measurement noninvariance, there were instances of significant effects on latent mean scores. For these reasons, we recommend that researchers explore the effects of both within-level and between-level covariates in evaluating measurement invariance and latent mean scores.

In summary, if measurement invariance is evaluated, meaningful cross-group comparisons can be made with greater confidence. A number of innovative methods for studying measurement invariance have recently been developed for use with popular statistical software. A Bayesian approach allows for comprehensively evaluating measurement invariance. Applying this technique to the most widely used instrument for measuring nurses' work environment, this study mainly found evidence of measurement noninvariance across language groups. These findings underscore the importance of a rigorous translation procedure and cultural adaptation of survey instruments. We hope our application of a method to assess measurement invariance will encourage health services researchers to take on the challenge of applying appropriate statistical techniques to explicitly test the suitability of their analysis to the study's design and implementation challenges.

References

Aiken, L. H., Cimiotti, J. P., Sloane, D. M., Smith, H. L., Flynn, L., & Neff, D. F. (2011). Effects of Nurse Staffing and Nurse Education on Patient Deaths in Hospitals With Different Nurse Work Environments. Medical Care, 49(12), 1047–1053.

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.

Asparouhov, T., & Muthén, B. O. (2012). Multiple group multilevel analysis.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In The reviewer's guide to quantitative methods in the social sciences (pp. 93–114).

Borsboom, D. (2006). When does measurement invariance matter? Medical Care, 44(11 Suppl 3), S176–81.

Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. Methodology in the social sciences (p. 475 p.).

Byrne, B. M., & Watkins, D. (2003). The Issue Of Measurement Invariance Revisited. Journal of Cross-Cultural Psychology, 34(2), 155–175.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504.

Cherepanov, D., Palta, M., Fryback, D. G., Robert, S. A., Hays, R. D., & Kaplan, R. M. (2011). Gender differences in multiple underlying dimensions of health-related quality of life are associated with sociodemographic and socioeconomic status. Medical Care, 49(11), 1021–30.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. Journal of Management, 25(1), 1–27.

Davidov, E., Dulmer, H., Schluter, E., Schmidt, P., & Meuleman, B. (2012). Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance. Journal of Cross-Cultural Psychology, 43(4), 558–575.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18(3-4), 117–44.

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

Jak, S., Oort, F. J., & Dolan, C. V. (2010). Measurement bias and multidimensionality; an illustration of bias detection in multidimensional measurement models. Advances in Statistical Analysis, 94(2), 129–137.

Johnson, T. P. (2006). Methods and frameworks for crosscultural measurement. Medical Care, 44(11 Suppl 3), S17–20.

Jones, R. N. (2006). Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling. Medical Care, 44(11 Suppl 3), S124–S133.

Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. Journal of the American Statistical Association, 70, 631–639.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal, 10(3), 333–351.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. American Political Science Review, 98(1), 191–207.

Kutney-Lee, A., McHugh, M. D., Sloane, D. M., Cimiotti, J. P., Flynn, L., Neff, D. F., & Aiken, L. H. (2009). Nursing: a key to patient satisfaction. Health Affairs (Millwood), 28(4), w669–77.

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in Nursing & Health, 25, 176–188.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. Structural Equation Modeling: A Multidisciplinary Journal, 11(3), 320–341.

Marshall, E. C., & Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease mapping models. Statistics in Medicine, 22, 1649–1660.

McHugh, M. D., Kelly, L. A., Smith, H. L., Wu, E. S., Vanak, J. M., & Aiken, L. H. (2013). Lower mortality in magnet hospitals. Medical Care, 51(5), 382–8.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. Medical Care, 44(11 Suppl 3), S69–77.

Millsap, R. E., & Kwok, O.-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. Psychological Methods, 9(1), 93–115.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54(4), 557–585.

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. Psychological Methods, 17(3), 313–35.

Muthén, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. Sociological Methodology, 25, 267–316.

Muthén, L. K., & Muthén, B. O. (2012). Mplus User's Guide. Seventh Edition. Losa Angeles, CA: Muthén & Muthén.

Parand, A., Burnett, S., Benn, J., Pinto, A., Iskander, S., & Vincent, C. (2011). The disparity of frontline clinical staff and managers' perceptions of a quality and patient safety initiative. Journal of Evaluation in Clinical Practice, 17(6), 1184–90.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3rd International Workshop on … (p. ISSN 1609–395X).

Price, M., Fitzgerald, L., & Kinsman, L. (2007). Quality improvement: the divergent views of managers and clinicians. Journal of Nursing Management, 15(1), 43–50.

Ramirez, M., Teresi, J. A., Holmes, D., Gurland, B., & Lantigua, R. (2006). Differential item functioning (DIF) and the Mini-Mental State Examination (MMSE). Overview, sample, and issues of translation. Medical Care, 44(11 Suppl 3), S95–106.

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., … Cella, D. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care, 45, S22–S31.

Rodriguez, H. P., & Crane, P. K. (2011). Examining multiple sources of differential item functioning on the Clinician & Group CAHPS® survey. Health Services Research, 46(6pt1), 1778–802.

Salomon, J. A., Tandon, A., & Murray, C. J. L. (2004). Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. British Medical Journal, 328(7434), 258.

Sass, D. A. (2011). Testing Measurement Invariance and Comparing Latent Factor Means Within a Confirmatory Factor Analysis Framework. Journal of Psychoeducational Assessment, 29(4), 347–363.

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., … Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. BMC Nursing, 10, 6.

Singer, S. J., Falwell, A., Gaba, D. M., & Baker, L. C. (2008). Patient safety climate in U.S. hospitals: variation by management level. Medical Care, 46(11), 1149–1156.

Squires, A., Aiken, L. H., van den Heede, K., Sermeus, W., Bruyneel, L., Lindqvist, R., … Matthews, A. (2013). A systematic survey instrument translation process for multi-country, comparative health workforce studies. International Journal of Nursing Studies, 50(2), 264–73.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. Journal of Consumer Research, 25(1), 78–107.

Tabachnick, B. G., & Fidell, L. S. (1983). Using Multivariate Statistics. PsycCRITIQUES (Vol. 28, p. 746).

Teresi, J. A. (2006). Overview of quantitative measurement methods. Equivalence, invariance, and differential item functioning in health applications. Medical Care, 44(11 Suppl 3), S39–49.

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the Use of Anchoring Vignettes for the Correction of Response Scale Differences in Subjctive Questions. Journal of the Royal Statistical Society. Series A, (Statistics in Society), 174(3), 575–595.

Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, Practices, and Recommendations for Organizational Research. Organizational Research Methods, 3(1), 4–70.

7 GENERAL DISCUSSION

If there is one guaranteed conversation in hospitals, it is the shortage of staffing and growing concerns about the quality of patient care. The Registered Nurse Forecasting study (RN4CAST) quantified the institutional variation in nurse staffing, nurse education and the quality of nurse work environments. Rigorous pan-European analyses have shown that such variation in the organization and practice of nursing is associated with variation in nurse wellbeing, nurse perceived quality, patient satisfaction and patient mortality. This PhD study demonstrated that bringing advanced methodological rigor to research linking nursing and outcomes yields new original knowledge and understanding in this field. Summarized below are the main findings, avenues for further research, and policy implications.

7.1 What this PhD study adds to what was known

The first three thematic chapters of this PhD study extend our knowledge and articulate more clearly the mechanisms of recently introduced concepts linking the organization and practice of nursing to nurse wellbeing and patient outcomes. Several advancements have been made.

Our first thematic chapter provided new insights into the concept of non-nursing tasks. First, very high proportions of nurses in Europe reported having performed non-nursing tasks and other tasks below their skill level during their last shift. This was later also evidenced in the South African branch of RN4CAST (Bekker, Coetzee, Klopper, & Ellis, 2014). Second, our study is unique in its focus on migratory status of nurses as one possible explanatory variable of reports on performing tasks below skill level. This analysis was instigated by two factors. First, anecdotal evidence suggest that foreign-trained nurses more often practice below their skill level compared to domestically trained nurses. Second, around the time of the RN4CAST study, WHO adopted the ground-breaking Global Code of Practice on the International Recruitment of Health Personnel. This instrument includes a specific section on the equal employment of domestically trained and foreign trained health workers. Our study provides evidence that foreign-trained nurses from developing countries are particularly vulnerable to inefficiencies in the organization of nursing care. Using a three-level analytic approach with nurses clustered in hospitals within countries, we showed that this issue is largely consistent across institutions within health systems and across health systems.

The second chapter of this PhD study is unique in testing the homology of effects across multiple systems levels. Prior to this PhD study, the four-level RN4CAST data structure had not yet been fully exploited. Studies had rather passively accounted for the multilevel structure. That is, RN4CAST researchers have used multilevel regression techniques to examine how higher level explanatory variables, aggregated at one level of the system, impact lower level outcomes. Two recent studies in this field have however explicitly tested the similarity of relationships across levels (i.e. homologous effects (Chen & Bliese, 2005)). Gabriel, Erickson, Moran, Diefendorff, & Bromley (2013) tested the effects of RN4CAST-like nursing work environment components on emotional exhaustion, job satisfaction and turnover intentions at the individual and nursing unit level. They showed that these components have different effects on these outcomes and that not all effects are homologous across these levels of analysis. A study of Italian nurses showed that perceived organizational support affected voluntary turnover at the individual and nursing unit level (El Akremi, Colaianni, Portoghese, Galletta, & Battistelli, 2014). In our study, we confirmed the hypothesis that country, hospital and nursing unit nurse work environment dimensions impact nurse burnout dimensions of emotional exhaustion, depersonalization and reduced personal accomplishment differently. As expressed in our overview of the patient safety literature (Chapter 1), several patient safety experts (Vincent et al., 1998; Waterson, 2009) and health services researchers with a specific focus on nursing (Meyer & O'Brien-Pallas, 2010) have called for studies testing whether effects are homologous across multiple systems levels. This type of research had also been specifically suggested by the developer of the PES-NWI, which is the multidimensional instrument that was used in RN4CAST to measure the nursing work environment (Lake, 2002).

In the third thematic chapter, we extended on the theoretical model of Ausserhofer et al. (2013). In their multicountry analysis of RN4CAST nurse survey data, Ausserhofer et al. theorized that inputs of nursing are linked with nursing care left undone, which subsequently impacts patient and nurse outcomes. This equals the specification of nursing care left undone as a mediator in this model, i.e. it explains the relationship between nursing inputs and outcomes. The authors tested the first part of this model, which showed that fewer nurses reported leaving nursing care left undone if there are lower patient-to-nurse ratios, better work environments for nurses, and when lower proportions of nurses

reported carrying out non-nursing tasks. We tested and confirmed the assumption that nursing care left undone mediates the relationship between nurse input (nurse staffing, nurse work environment and non-nursing tasks) and patient experiences with care. This is in line with previous research that showed that missed care, a construct that is very similar to nursing care left undone, mediates the relationship between nurse staffing and patient falls (Kalisch, Tschannen, & Lee, 2012). In addition, we showed that the effect of worse nurse staffing on more care left undone diminishes with an increasing proportion of bachelor-prepared nurses. This study is unique in its theory-driven detailed examination of mechanisms underlying the relationship between the organization and practice of nursing and patient outcomes.

A second part of this PhD study related to what should be the inevitable challenge of evaluating measurement invariance. This pertains to the comparability of survey findings across groups, over time or for different methods of data collection. Researchers in this field have not been sufficiently sensitive to the question of measurement invariance in evaluating cross-group comparisons in survey research. Using newly collected data in one large hospital entity, we compared nurse managers' and staff nurses' perceptions of their work environment, and evaluated whether the difference we observed were true differences, or were caused by non-invariant measures. From a substantive point of view, this study is unique in showing that managers seem to rate the work environment better than their staff nurses. From a methodological point of view, this study is unique in evaluating measurement invariance for within-level and between-level covariates jointly. We showed that researchers should especially be cautious in comparing findings across language groups. It was also unique in proposing a method for model evaluation using posterior predictive p-values (PPP) in our application of Bayesian two-level models. It has been recognized as an exemplary paper applying a Bayesian analysis using Mplus (Muthén & Muthén, 2014). The most recent version (v7.3) of the Mplus software includes posterior predictive p-values (PPP) for Bayesian two-level, three-level and cross-classified models.

7.2 Avenues for further research

7.2.1 Decision-points at which nursing inputs affect outcomes

Many researchers in this domain conclude their study with a closing remark on what patient outcomes would look like under different scenarios of input. From the RN4CAST study for example, Aiken and colleagues (2014) concluded that patients are almost 30% less likely to die in hospitals in which 60% of the nurses had bachelor's degrees and nurses cared for an average of six patients compared to hospitals in which only 30% of the nurses had bachelor's degrees and nurses cared for an average of eight patients. Using a logistic regression model Schubert, Clarke, Glass, Schaffert-Witvliet and De Geest (2009) studied at which levels rationing of nursing care in Swiss hospitals becomes problematic. They showed that, already when nurses reported that rationing of care happened rarely, this negatively affected nosocomial infections, pressure ulcers and patient satisfaction. Ball et al. (2014) showed that significantly more care is being left undone in hospitals where nurses take care of more than 7.4 patients on average. As a last example, Tervo-Heikkenen et al. (2008) demonstrated that eight patients per nurse is the cut-off point for patient satisfaction.

Policies that seek an evidence-based approach to improve nurse and patient outcomes through nursing delivery system strategies would likely benefit from a larger number of studies that identify specific minimum as well as maximum thresholds at which inputs affect outcomes. Staggs and Dunton (2014) phrase this issue as 'how much staffing is enough or how much might be too much?'. A similar reasoning had previously instigated Zhang, Unruh, Liu & Wan (2006) to search for decision points between staffing and quality in nursing homes using non-linear modeling techniques. Using an S-shaped production function they searched for the optimum level of production. It was hypothesized that as nurse staffing increases, quality first increases at a faster rate (increasing marginal returns), then a slower rate (diminishing marginal returns), and finally at some point stops increasing at all (diminishing absolute returns). The optimum staffing point is hypothesized to be in the area of diminishing marginal returns. They showed that different minimum staffing levels can be set for achieving different levels of quality. At a certain level of quality, staffing ratios need to increase less substantively to reach ever higher levels of

quality. Non-linear relations have been shown for various outcomes in multiple settings and using different analytic techniques in both a frequentist and Bayesian framework. For example, Mark et al. (2004) had also found signs of diminishing marginal returns, showing that better nurse staffing decreases the mortality ratio for staffing levels up to 4.62 patients per nurse. This corresponded to the 88th percentile value in their study sample. Blegen and Vaughn's (1998) findings even demonstrated that if the proportion of nurses in the staff mix (including nurses and unlicensed assistive personnel) was higher than 85%, the rate of adverse patient outcomes increased. Most recently, Pitkäaho, Partanen, Miettinen, & Vehviläinen-Julkunen (2014) used a Bayesian dependency model that allowed them to model relationships between several categorical variables and a non-linear relationship between nurse staffing and patients' length-of-stay in acute care units.

As evidenced by a large number of studies, many factors beside nurse staffing determine the level of quality. Based on this premise that staffing adequacy is more than a single patient-to-nurse ratio, Schmalenberg and Kramer (2009) searched for a comprehensive list of factors that influence nurse-perceived staffing adequacy. Their findings suggested that even with sufficient numbers of nurses, nurses may still perceive that staffing is inadequate. It was concluded that success of explaining nurse-perceived staffing adequacy is determined by including staff nurses' work environment perceptions. Mark, Salyer & Harless (2002) had previously shown that structures such as smaller unit size, lower patient acuity and consistent support services are positively associated with nurses' perceptions of adequate staffing. Nurse-perceived staffing adequacy has also been shown to predict patient outcomes. Tourangeau et al. (2007) showed that a 10% increase in nurse-reported staffing adequacy was associated with decreased patient mortality rates. Swedish RN4CAST researchers showed that perceived staffing adequacy is also related with nurses' perception of patient safety (Smeds Alenius, Tishelman, Runesdotter, & Lindqvist, 2014).

For these reasons, future efforts using the RN4CAST study data aim to take forward in greater detail our understanding of which factors determine nurse-perceived adequate staffing levels, and at which level of the system. Its role of both dependent and independent variable opens opportunities for theorizing a mediating relationship of nurse-perceived staffing adequacy. Alternatively, it could be investigated using a multilevel covariance regression (MCR) model (Li et al., 2014). This is an extension we

developed on the basis of our findings in Chapter 4. The MCR model is a Bayesian multivariate multilevel factor model that allows for level-specific covariates in the mean structure and covariance matrix of the three burnout dimensions. Findings revealed that as nurses' work experience increases, the correlation between emotional exhaustion and personal accomplishment also increases (Li et al., 2014). In the case of perceived staffing adequacy, it could be of interest to study how the covariance matrix between nurse-perceived staffing adequacy and patient outcomes depends on explanatory factors. An additional perspective may result from the fact that two questions about staffing adequacy were asked (on a four-point Likert scale): 'Enough staff to get the work done' and 'Enough registered nurses on staff to provide quality patient care'. It will be of interest to study if the same explanatory factors are associated with these questions, and whether the magnitude of this association is similar. Ultimately, the goal would be to isolate the factors that determine quality from nurses' point of view.

7.2.2 Evaluating cost offsets of nursing system delivery strategies

The question of which level of quality needs to be targeted and what are the required inputs inevitably involves cost considerations when informing decision-makers. Only a limited number of studies have investigated the payoff in costs savings from modifying nursing delivery system strategies. The groundbreaking studies by Aiken et al. (2002; 2003) and Needleman et al. (2002) that linked nurse staffing and education to patient outcomes were followed by analyses of costs and cost savings from increasing nurse staffing and nurse education levels. Rothberg, Abraham, Lindenauer, & Rose (2005) concluded on the basis of Aiken et al.'s (2002) study that modifying patient to nurse ratios from eight patients per nurse to four patients per nurse becomes progressively less cost effective. In terms of costs per life saved, a ratio of four patients per nurse would however still be reasonably priced compared to commonly accepted interventions such as thrombolytic therapy in acute myocardial infarction and routine cervical cancer screening with PAP tests. Two other cost-effectiveness studies evaluated several scenarios that not only focused on the economic value of improving patient-to-nurse ratios, but also on education levels. Needleman et al. (2006) found that increasing each U.S. hospital below the 75th percentile for the proportion of bachelor-prepared nurses to that level (cost of $811 million) pays for itself (net cost of -$1821 million) when taking into account cost offsets of avoided patient deaths and complications (e.g.

urinary tract infections, pneumonia) and reduced length of stay. Based on Aiken et al.'s (2003) findings, Newbold (2008) came to similar conclusions. Needleman (2008) also estimated from Aiken's study (2002) that economic and social cases for increased nursing can be made from reduced nurse turnover rates that would result from better patient-to-nurse ratios. Dall, Chen, Seifert, Maddox, & Hogan (2009) calculated that each additional patient care nurse will generate at least 72% of her salary in medical savings. Recently, Yakusheva, Lindrooth and Weiss (2014) conducted an economic analysis that supported the IOM's recommendation to increase the proportion of bachelor-educated nurses to 80% by 2020 (Institute of Medicine (U.S.) Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing at the Institute of Medicine, 2011).

Despite this evidence, Palese and Watson (2014) expressed their fear that RN4CAST findings on the association of nurse education and workload with patient mortality (Aiken et al., 2014) would be deemed too expensive to act upon. Unfortunately, European evidence on the economic value of nursing is scarce. Only Van den Heede et al. (2010) expanded on their findings of an association between nurse staffing levels and lower mortality rates in Belgian general cardiac postoperative nursing units (Van den Heede, Lesaffre, et al., 2009). Using a design similar to that of Needleman et al. (2006), it was concluded that increasing nurse staffing levels to the 75th percentile might be a cost-effective intervention when compared to other cardiovascular interventions.

More research in a European setting should evaluate if the quality and safety of patient care can be improved in an economical way by increasing nurse inputs. Also, economic evaluations have focused on workload and education, while the association between positive work environments and better nurse and patient outcomes has also been established across very different health systems. Future economic evaluations may therefore want to evaluate the cost of interventions to improve nursing work environments. In line with our research findings on differential effects of nursing work environment components at the hospital and nursing unit level, such studies should aim to allocate costs to their respective system levels. It could be of particular interest to evaluate for example if interventions implemented at the nursing unit level can subsequently be applied in other nursing units at diminishing

costs. This in itself might then produce multiple level systems theories that lend themselves to further development.

7.2.3 Opportunities in a growing field of research

Many interpretations of findings linking nursing input to outcomes, including the estimation of costs, assume causality of the association. In their systematic review on the association between nurse staffing and patient outcomes, Kane et al. (2007) concluded that the arguments for a causal effect are mixed and recommended longitudinal studies to be designed. For this reason, researchers at the University of Pennsylvania who vice-coordinated the RN4CAST study, initiated RN4CAST-U.S. This longitudinal study of panels of hospitals recently received funding from the U.S. National Institute of Nursing Research. Its aim is to study associations between changes over time (1999, 2006, and 2014) in nursing resources and outcomes for patients. A previous analysis of panel data of 137 Pennsylvania hospitals showed lower rates of high burnout levels, intention to leave the present job, and job dissatisfaction in 2006 compared to 1999. More detailed analyses showed that hospitals in which nurses over time became more satisfied with their work environment, also have favorable changes in these nurse outcomes. (Kutney-Lee, Wu, Sloane, & Aiken, 2013).

Important opportunities also remain from extending this research to non-hospital settings. Nursing homes and home healthcare are increasingly the focus of policy efforts to reduce expensive hospital use. Relatively few research evidence linking nursing system delivery strategies to outcomes is however available for these settings. RN4CAST-U.S. will therefore extend to nursing homes and home health agencies using measures similar to those used in hospital studies. Similar research in Europe is needed.

The RN4CAST study and this PhD study have demonstrated the added value of extending our knowledge of the process of nursing care. However, our measures of nursing care left undone and tasks below skill level could be further refined. The MISSCARE survey (Kalisch, Landstrom, & Williams, 2009) holds more detailed factors of missed care and looked into additional explanations of missed care, such as communication, material resources and labor resources (Kalisch, Terzioglu, & Duygulu, 2012).

It might be useful to theorize if other process variables underlie the relationship between nursing and outcomes. Dubois et al. (2012) proposed a theoretical model of nursing care organization in which the work environment was divided into aspects of the practice environment and aspects related to the capacity for innovation. To measure the practice environment, Dubois et al. recommend the PES-NWI, which was used in RN4CAST. The five main features that Dubois and colleagues assessed in terms of capacity for innovation were based on the work of Kimball, Joynt, Cherner, & O'Neil (2007) and included elevated RN role, sharpened focus of care on the patient, mechanisms to ensure smooth patient transitions across care settings and links with communities, mechanisms for monitoring results to improve performance, and leveraging of technologies. Kimball et al. also identified that "beyond these five common elements, successful new care delivery models also share a key lesson for implementation success: early and regular involvement of care givers in the design and implementation of new models". One model that proposes such bottom-up approach is the Productive Ward - Releasing Time to Care program (Wilson, 2009). This program aims to provide staff with more time to provide direct patient care. It heavily relies on lean methodology (Womack, Jones, & Roos, 1990) and offers 15 modules to optimize the organization of patient care. A recent longitudinal study in a Belgian hospital showed that implementing this program may result in improved nurse-physician relations, nurse management at the unit level, hospital management and organizational support, and nurses' feelings of personal accomplishment (Van Bogaert et al., 2014). Another critically acclaimed frontline project is Transforming Care at the Bedside (TCAB), which also aims to increase time spend in direct patient care. TCAB is designed for medical and surgical nursing units in hospitals where it aims to empower nurses in redesigning work processes (The Robert Wood Johnson Foundation & The Institute for Healthcare Improvement, 2007).

Last, studies should examine which elements in the work environment are important in the relationship with process variables such as missed care. Of particular interest are nurse-physician relations, which have been found to impact nurse and patient outcomes. Better nurse-reported relations between nurses and physicians link to lower patient mortality (Estabrooks et al., 2005; Knaus, Draper, Wagner, & Zimmerman, 1986; Mitchell, Armstrong, Simpson, & Lentz, 1989). For outcomes of nurse

wellbeing, Bruyneel et al. (2009) showed that better ratings of nurse-physician relationships are associated with an increase in the odds of reporting high job satisfaction and better nurse-perceived quality of care. This Belgian study confirmed reports from England (Rafferty, Ball, & Aiken, 2001) and Iceland (Gunnarsdóttir et al., 2009). Another Belgian study established associations with increased intention to stay employed at the current hospital and increased feelings of personal accomplishment (Van Bogaert, Clarke, et al., 2009). Finally, Li et al. (2013) used a four-level model to explore the effect of nursing unit, hospital, and country-level variability on the relationship between nurses' work environment dimensions and nurse burnout. Findings showed that nurse-physician collegial relations affected burnout measures at the nursing unit level, but not at the hospital level.

7.2.4 Measurement invariance evaluation in an cross-country perceptive

RN4CAST has provided detailed findings on interrelations between explanatory variables of nursing organization on the one hand, and nurse and patient outcomes on the other hand. This has offered unique insights into the type of nurse workforce policies that are desirable. The RN4CAST experience had shown however that mere descriptive findings, between-country comparisons in particular, have also received considerable attention from specialist and non-specialist audiences. Particularly patient-to-nurse ratios, education levels and the quality of the work environment have often formed the basis of country comparisons or rankings. For cross-national comparisons of subjective questions such as the work environment items to be valid, all items would need to function in exactly the same way in all participating countries.

Future work should therefore evaluate cross-cultural measurement invariance of nurse work environment measures, and examine ways in which individual propensities and cultural context explain possible non-invariance. Health services researchers previously recognized that cross-cultural measurement strategies will benefit from greater attention to theories of culture that may have measurement ramifications. This amounts to 'unpacking culture by means of context variables' (Matsumoto & Van de Vijver, 2011). The cross-cultural framework developed by Hofstede (1980) is best-known for this type of analysis. Hofstede's dimensions reflect work-related values defined at the country

level (power distance, individualism, masculinity, uncertainty avoidance, long- versus short-term orientation). Country-level scores for these dimensions could thus be merged with the RN4CAST data to study whether there are consistent associations between Hofstede's dimensions of national culture and the latent work environment constructs captured in the nurse survey.

7.3 Policy implications

Following the Institute of Medicine's famous reports on patient safety issues (2000) and quality improvement (2001), patient safety research and quality improvement initiatives received massive attention. But although great accomplishments have been achieved in the safety field, the rate of progress is slow (Wachter, 2010). RN4CAST strengthened the believe that improvements can be made at an incremental rate if policy makers and human resources managers would influence the direction of travel along a trajectory that recognizes the vital role of nurses in providing safe patient care. RN4CAST used its evidence base to inform policy at all levels with the ultimate goal of improving patient care. A range of capacity building and knowledge dissemination activities have been offered, including peer-reviewed manuscripts in the leading medical, nursing and health care policy journals, high-level conferences to engage stakeholders, and wide media coverage. Presented below are the implications that result from RN4CAST and this PhD study, which are structured around investments in a better educated nurse workforce, improved nurse staffing, and improving nursing work environments as levers for improving quality of patient care. The starting assumption is that policy makers and human resources managers at the very least acknowledge that these characteristics of nursing are modifiable properties of a health care organization.

The priority task that we assign to policy-makers and human resources managers is to place a premium on well-educated nurses to secure an optimal contribution of nursing to healthcare delivery. There is now well-established evidence in the U.S. as well as in Europe that higher proportions of bachelor-prepared nurses are associated with lower patient mortality. As one possible explanation, findings from our study linking nurse staffing levels to missed care suggest that a better educated nurse workforce may increase the productivity, in the sense of allocating tasks more efficiently across the

available labor supply. What should facilitate the debate on nurse education is that an increasing number of independent research has shown that education levels can be improved in a cost effective way. It has also been repeatedly shown that this strategy is favored over increases in nurse staffing if both scenarios are on the table and are mutually exclusive. Although evidence from cost-effectiveness research is not available on a large scale in Europe, we believe that these U.S. findings should contribute appropriately to the European policy debate. We cling to the hope that EU Member States will take responsibility in this area, where the European Commission has chosen not to opt for the most courageous option by allowing vocational nurse training after only 10 years of general education. Member States should go through debates that touch upon the very design of nurse education. They should take targeted action to improve nurse education levels. Such initiative may also be catalytic in designing and developing new ways of reconfiguring tasks and roles of nurses and other caregivers across Europe. In the U.S., the IOM's Future of Nursing report (2011) aims for 80% bachelor prepared nurses. Health care organizations carry a large responsibility for this plan to succeed. They are recommended to encourage lower educated nurses to enter bachelor programs within five years of graduation. Incentives include tuition reimbursement among others. Researchers have found that this would present a strong business case at the institutional level (Yakusheva et al., 2014). Findings from this PhD suggest that, at least in terms of completing necessary nursing tasks, benefits would be at their maximum from bringing nurse education levels across institutions to the country average. This value neglects potential benefits that increased nurse education may have on other outcomes. Two other recommendations from the Future of Nursing report that many EU Member States would benefit from are: 1) that nurses should practice to the full extent of their education and training – which has been evidenced by this PhD study to be a huge issue in European hospitals; and 2) that nurses should engage in lifelong learning. We believe that these might be connected, as engagement in lifelong learning may improve the quality and commitment of the nurse workforce and stimulate career progression, which in turn could trigger the removal of scope-of-practice barriers. A last recommendation that may positively impact the competence level of nurses is the wider implementation of preceptor programs for newly hired nurses. Specific programs should be in place for

foreign-trained nurses, as many European health systems are becoming more and more reliant on the supply of migrant nurses.

A further priority is to maintain a sufficient number of nurses relative to the number of patients. Findings from RN4CAST and this PhD study clearly link nurse staffing to patient mortality, patient satisfaction, and nurse productivity. A well-known measure taken against staffing inadequacy are mandatory hospital nurse staffing ratios. California is still the only U.S. state that has enacted such (unfunded) fixed ratio mandates. There is no doubt that these ratios have increased nurse staffing levels over time (Burnes Bolton et al., 2007; Donaldson & Shapiro, 2010; McHugh et al., 2012). Recent findings have also shown that this improvement did not come at the cost of an increased proportion of vocational nurses (who generally cost less) (McHugh et al., 2012), a fear that had previously been expressed (Coffman, Seago, & Spetz, 2002). However, fixed ratios remain highly controversial, as several experts raised concerns or provided evidence (mostly using a before and after study design) about the bad fit between fixed ratios and the variability in the intensity of care delivery (de Véricourt & Jennings, 2008), high costs (Coffman et al., 2002; Welton, 2007), absence of evidence on increased nurse job satisfaction (Spetz, 2008) and improve nurse-sensitive patient outcomes (Burnes Bolton et al., 2007; Cook, Gaynor, Stephens, & Taylor, 2012; Donaldson & Shapiro, 2010; Sochalski, Konetzka, Zhu, & Volpp, 2008), hiring younger, less experienced nurses because they are cheaper (Griffiths, 2009), deteriorated decision-making control (Chapman et al., 2009), and a decrease of other ancillary staff (Chapman et al., 2009; Griffiths, 2009) with the risk of nurses performing more non-nursing care tasks (Coffman et al., 2002; Welton, 2007). Alternative methods that might be worth considering, but for which evidence is also lacking, include directly linking the costs and billing for inpatient nursing care with hospital reimbursement (Welton, 2007), mandating the process by which hospitals determine staffing by having a staffing committee that matches the skills and experience of the nursing staff to the need of patients (required by hospitals in 7 U.S. states), and reporting of staffing levels to the public and/or a regulatory body (required by hospitals in 5 U.S. states) (American Nurses Association, 2014). In the RN4CAST study, hospital managers were asked if ward staffing levels are based on the result of matching staffing to patient acuity using a formal system. Unpublished data show that less than half of the hospitals in

RN4CAST use such system, ranging from 17% of the hospitals in Finland, to 63% of the hospitals in Poland. A blended approach of requiring hospitals to have a nurse driven staffing committee and publicly report their staffing levels may thus hold the most promise.

The success of these strategies hinges on creating a positive work environment for nurses. This PhD study brought new levels of evidence to this issue. First, assuming causality of the associations, we showed at which levels interventions might be successful. Second, nurses rather than their managers should serve as informants of what is going on in their work environment. The accumulating evidence on the importance of positive nurse work environments makes it more easy to make a persuasive argument that there is an urgent need for innovative solutions to redesign these environments. We believe that blueprints for excellent nursing care environments are to a large degree exchangeable between hospitals. We therefore propose that hospitals that perform exceptionally well – such hospitals were identified in almost every country participating to RN4CAST – exchange 'what works' principles. This process can be viewed as a shared responsibility between researchers, human resources managers and policy makers.

Whereas similar studies in the U.S. instantly fueled debates about the direction of nursing's future, past experiences have shown that Europe is still in its infancy compared to evidence-based policy implementation in the U.S. The work of RN4CAST therefore continues in the U.S., among the original collaborators in Europe, and in countries that recently joined this unique consortium of researchers with the common objective to highlight the importance of nursing in improving patient outcomes.

References

Aiken, L. H., Clarke, S. P., Cheung, R. B., Sloane, D. M., & Silber, J. H. (2003). Educational levels of hospital nurses and surgical patient mortality. JAMA : The Journal of the American Medical Association, 290(12), 1617–23.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA : The Journal of the American Medical Association, 288(16), 1987–93.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., … Sermeus, W. (2014). Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. Lancet, 383(9931), 1824–30.

American Nurses Association. (2014). Nurse staffing plans and ratios. Retrieved from http://www.nursingworld.org/MainMenuCategories/Policy-Advocacy/State/Legislative-Agenda-Reports/State-StaffingPlansRatios

Ausserhofer, D., Zander, B., Busse, R., Schubert, M., De Geest, S., Rafferty, A. M., … Schwendimann, R. (2013). Prevalence, patterns and predictors of nursing care left undone in European hospitals: results from the multicountry cross-sectional RN4CAST study. BMJ Quality & Safety.

Ball, J., Murrells, T., Rafferty, A. M., Morrow, E., & Griffiths, P. (2014). "Care left undone" during nursing shifts: associations with workload and perceived quality of care. BMJ Quality & Safety, 23(2), 116–25.

Bekker, M., Coetzee, S. K., Klopper, H. C., & Ellis, S. M. (2014). Non-nursing tasks, nursing tasks left undone and job satisfaction among professional nurses in South African hospitals. Journal of Nursing Management.

Blegen, M. A., & Vaughn, T. (1998). A multisite study of nurse staffing and patient occurrences. Nursing Economic$, 16(4), 196–203.

Burnes Bolton, L., Aydin, C. E., Donaldson, N., Brown, D. S., Sandhu, M., Fridman, M., & Aronow, H. U. (2007). Mandated nurse staffing ratios in California: a comparison of staffing and nursing-sensitive outcomes pre- and postregulation. Policy, Politics & Nursing Practice, 8(4), 238–50.

Chapman, S. A., Spetz, J., Seago, J. A., Kaiser, J., Dower, C., & Herrera, C. (2009). How have mandated nurse staffing ratios affected hospitals? Perspectives from California hospital leaders. Journal of Healthcare Management / American College of Healthcare Executives, 54(5), 321–33; discussion 334–5.

Chen, G., & Bliese, P. D. (2005). Conceptual framework and statistical procedures for delineating and testing multilevel theories of homology. Academy of Management Annual Meeting Proceedings, 8, A1–A6.

Coffman, J. M., Seago, J. A., & Spetz, J. (2002). Minimum nurse-to-patient ratios in acute care hospitals in California. Health Affairs (Project Hope), 21(5), 53–64.

Cook, A., Gaynor, M., Stephens, M., & Taylor, L. (2012). The effect of a hospital nurse staffing mandate on patient health outcomes: evidence from California's minimum staffing regulation. Journal of Health Economics, 31(2), 340–8.

Dall, T. M., Chen, Y. J., Seifert, R. F., Maddox, P. J., & Hogan, P. F. (2009). The economic value of professional nursing. Medical Care, 47(1), 97–104.

De Véricourt, F., & Jennings, O. B. (2008). Nurse-To-Patient Ratios in Hospital Staffing: A Queueing Perspective.

Donaldson, N., & Shapiro, S. (2010). Impact of California mandated acute care hospital nurse staffing ratios: a literature synthesis. Policy, Politics & Nursing Practice, 11(3), 184–201.

Dubois, C.-A., D'Amour, D., Tchouaket, E., Rivard, M., Clarke, S., & Blais, R. (2012). A taxonomy of nursing care organization models in hospitals. BMC Health Services Research, 12(1), 286.

El Akremi, A., Colaianni, G., Portoghese, I., Galletta, M., & Battistelli, A. (2014). How organizational support impacts affective commitment and turnover among Italian nurses: a multilevel mediation model. The International Journal of Human Resource Management, 25(9), 1185–1207.

Gabriel, A. S., Erickson, R. J., Moran, C. M., Diefendorff, J. M., & Bromley, G. E. (2013). A multilevel analysis of the effects of the practice environment scale of the nursing work index on nurse outcomes. Research in Nursing & Health.

Griffiths, P. (2009). RN+RN=better care? What do we know about the association between the number of nurses and patient outcomes? International Journal of Nursing Studies, 46(10), 1289–90.

Institute of Medicine (U.S.) Committee on Quality of Health Care in America. (2000). To Err Is Human: Building a Safer Health System. (L. T. Kohn, J. M. Corrigan, & S. Molla, Eds.)Medicine (Vol. 126, p. 312). Washington, D.C.: National Academies Press.

Institute of Medicine (U.S.) Committee on Quality of Health Care in America. (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, D.C.: National Academies Press.

Institute of Medicine (U.S.) Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing at the Institute of Medicine. (2011). The Future of Nursing: Leading Change, Advancing Health. Washington, D.C.: The National Academies Press.

Kalisch, B. J., Landstrom, G., & Williams, R. A. (2009). Missed nursing care: errors of omission. Nursing Outlook, 57(1), 3–9.

Kalisch, B. J., Terzioglu, F., & Duygulu, S. (2012). The MISSCARE Survey-Turkish: psychometric properties and findings. Nursing Economic$, 30(1), 29–37.

Kalisch, B. J., Tschannen, D., & Lee, K. H. (2012). Missed nursing care, staffing, and patient falls. Journal of Nursing Care Quality, 27(1), 6–12.

Kane, R. L., Shamliyan, T. A., Mueller, C., Duval, S., & Wilt, T. J. (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. Medical Care, 45(12), 1195–204.

Kimball, B., Joynt, J., Cherner, D., & O'Neil, E. (2007). The quest for new innovative care delivery models. The Journal of Nursing Administration, 37(9), 392–8.

Kutney-Lee, A., Wu, E. S., Sloane, D. M., & Aiken, L. H. (2013). Changes in hospital nurse work environments and nurse job outcomes: an analysis of panel data. International Journal of Nursing Studies, 50(2), 195–201.

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in Nursing & Health, 25, 176–188.

Li, B., Bruyneel, L., & Lesaffre, E. (2014). A multivariate multilevel Gaussian model with a mixed effects structure in the mean and covariance part. Statistics in Medicine, 33(11), 1877–99.

Mark, B. A., Harless, D. W., McCue, M., & Xu, Y. (2004). A longitudinal examination of hospital registered nurse staffing and quality of care. Health Services Research, 39(2), 279–300.

Mark, B. A., Salyer, J., & Harless, D. W. (2002). What explains nurses' perceptions of staffing adequacy? The Journal of Nursing Administration, 32(5), 234–42.

McHugh, M. D., Brooks Carthon, M., Sloane, D. M., Wu, E., Kelly, L., & Aiken, L. H. (2012). Impact of nurse staffing mandates on safety-net hospitals: lessons from California. The Milbank Quarterly, 90(1), 160–86.

Meyer, R. M., & O'Brien-Pallas, L. L. (2010). Nursing Services Delivery Theory: an open system approach. Journal of Advanced Nursing, 66(12), 2828–38.

Muthén, B. O., & Muthén, L. K. (2014). Papers Using Special Mplus Features. Bayesian analysis. Retrieved from http://www.statmodel.com/papers.shtml

Needleman, J. (2008). Is what's good for the patient good for the hospital? Aligning incentives and the business case for nursing. Policy, Politics & Nursing Practice, 9(2), 80–7.

Needleman, J., Buerhaus, P. I., Stewart, M., Zelevinsky, K., & Mattke, S. (2006). Nurse staffing in hospitals: is there a business case for quality? Health Affairs (Project Hope), 25(1), 204–11.

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), 1715–22.

Newbold, D. (2008). The production economics of nursing: a discussion paper. International Journal of Nursing Studies, 45(1), 120–8.

Palese, A., & Watson, R. (2014). Nurse staffing and education in Europe: if not now, when? Lancet, 383(9931), 1789–90.

Pitkäaho, T., Partanen, P., Miettinen, M., & Vehviläinen-Julkunen, K. (2014). Non-linear relationships between nurse staffing and patients' length of stay in acute care units: Bayesian dependence modelling. Journal of Advanced Nursing.

Rothberg, M. B., Abraham, I., Lindenauer, P. K., & Rose, D. N. (2005). Improving nurse-to-patient staffing ratios as a cost-effective safety intervention. Medical Care, 43(8), 785–91.

Schubert, M., Clarke, S. P., Glass, T. R., Schaffert-Witvliet, B., & De Geest, S. (2009). Identifying thresholds for relationships between impacts of rationing of nursing care and nurse- and patient-reported outcomes in Swiss hospitals: a correlational study. International Journal of Nursing Studies, 46(7), 884–93.

Smeds Alenius, L., Tishelman, C., Runesdotter, S., & Lindqvist, R. (2014). Staffing and resource adequacy strongly related to RNs' assessment of patient safety: a national study of RNs working in acute-care hospitals in Sweden. BMJ Quality & Safety, 23(3), 242–9.

Sochalski, J., Konetzka, R. T., Zhu, J., & Volpp, K. (2008). Will mandated minimum nurse staffing ratios lead to better patient outcomes? Medical Care, 46(6), 606–13.

Spetz, J. (2008). Nurse satisfaction and the implementation of minimum nurse staffing regulations. Policy, Politics & Nursing Practice, 9(1), 15–21.

Staggs, V. S., & Dunton, N. (2014). Associations between rates of unassisted inpatient falls and levels of registered and non-registered nurse staffing. International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 26(1), 87–92.

Tervo-Heikkinen, T., Kvist, T., Partanen, P., Vehviläinen-Julkunen, K., & Aalto, P. (2008). Patient satisfaction as a positive nursing outcome. Journal of Nursing Care Quality, 23(1), 58–65.

The Robert Wood Johnson Foundation, & The Institute for Healthcare Improvement. (2007). A New Era in Nursing: Transforming Care at the Bedside.

Tourangeau, A. E., Doran, D. M., McGillis Hall, L., O'Brien Pallas, L., Pringle, D., Tu, J. V, & Cranley, L. A. (2007). Impact of hospital nursing care on 30-day mortality for acute medical patients. Journal of Advanced Nursing, 57(1), 32–44.

Van Bogaert, P., Van heusden, D., Somers, A., Tegenbos, M., Wouters, K., Van der Straeten, J., … Havens, D. S. (2014). The Productive Ward programTM: a longitudinal multilevel study of nurse perceived practice environment, burnout, and nurse-reported quality of care and job outcomes. The Journal of Nursing Administration, 44(9), 452–61.

Van den Heede, K., Lesaffre, E., Diya, L., Vleugels, A., Clarke, S. P., Aiken, L. H., & Sermeus, W. (2009). The relationship between inpatient cardiac surgery mortality and nurse numbers and educational level: analysis of administrative data. International Journal of Nursing Studies, 46(6), 796–803.

Van den Heede, K., Simoens, S., Diya, L., Lesaffre, E., Vleugels, A., & Sermeus, W. (2010). Increasing nurse staffing levels in Belgian cardiac surgery centres: a cost-effective patient safety intervention? Journal of Advanced Nursing, 66(6), 1291–6.

Vincent, C., Taylor-Adams, S., & Stanhope, N. (1998). Framework for analysing risk and safety in clinical medicine. BMJ (Clinical Research Ed.), 316(7138), 1154–7.

Wachter, R. M. (2010). Patient safety at ten: unmistakable progress, troubling gaps. Health Affairs (Project Hope), 29(1), 165–73.

Waterson, P. (2009). A critical review of the systems approach within patient safety research. Ergonomics, 52(10), 1185–95.

Welton, J. M. (2007). Mandatory Hospital Nurse to Patient Staffing Ratios: Time to Take a Different Approach. The Online Journal of Issues in Nursing, 12. Retrieved from http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume122007/No3Sept07/MandatoryNursetoPatientRatios.html#Mitchell

Wilson, G. (2009). Implementation of releasing time to care - The productive ward. Journal of Nursing Management, 17, 647–654.

Womack, J. P., Jones, D. T., & Roos, D. (1990). The Machine that Changed the World: The Story of Lean Production. World (pp. 1–11).

Yakusheva, O., Lindrooth, R., & Weiss, M. (2014). Economic evaluation of the 80% baccalaureate nurse workforce recommendation: a patient-level analysis. Medical Care, 52(10), 864–9.

Zhang, N. J., Unruh, L., Liu, R., & Wan, T. T. H. (2006). Minimum nurse staffing ratios for nursing homes. Nursing Economic$, 24(2), 78–85, 93, 55.

8 SUMMARY - SAMENVATTING

Summary

The past two decades have witnessed a growing body of U.S. research evidence supporting contentions that hospitals with better nurse staffing, higher proportions of bachelor-educated nurses and superior nurse practice environments have better outcomes of patient mortality, patient satisfaction with care, and nurse wellbeing. More recently, the Registered Nurse Forecasting (RN4CAST) research consortium confirmed the direction and consistency of these associations across European hospitals at a pan-European scale. This PhD study capitalizes on the RN4CAST study. It encourages the use of contemporary statistical techniques by refining our understanding of the mechanisms underlying established associations, by studying phenomena at multiple organizational levels, and by promoting an evaluation of similarity of concepts across groups. Two important issues are tackled.

A first issue is a lack of alignment between statistical techniques and the many comprehensive theories that have been proposed to explain why inputs of nursing are associated with nurse and patient outcomes. A first challenge in the statistical analysis is to account for the complex hierarchical nature of the data and theories. A hierarchical data structure is also called a multilevel data structure. In the RN4CAST study, a complex four-level data structure occurred because the study design corresponds with the clustering of nurses and patients (level 1) within nursing units (level 2) within hospitals (level 3) within countries (level 4). This complex data structure requires the use of multilevel modeling techniques. A second complexity arises from the theoretical specification of mediator and moderator variables. Mediators illuminate the mechanisms that explain why associations can be observed. Moderators are contextual factors that define the conditions under which such associations can be predicted. A third challenge relates to understanding underlying dimensions, also called latent variables, from a set of survey items. Hypotheses involving latent variables require a statistical technique called structural equation modeling. The analysis of mediators, moderators and latent variables must be performed in a multilevel framework to comply with suggested theoretical models.

A second issue relates to a lack of measurement invariance evaluation in the cross-group comparisons that are so important to this field of research. Measurement invariance pertains to whether

measurement instruments and the constructs being measured operate in the same way across populations of interest. Lack of measurement invariance can result in the incorrect interpretation of what the researcher believed were true cross-group differences. Measurement invariance is studied in a latent variable framework and, like mediation and moderation analysis, relies on the use of structural equation models.

In a first introductory chapter, we provide an overview of the patient safety literature and theoretical models and landmark studies in this field of research. It concludes with the research objectives of this PhD study. A second introductory chapter provides an overview of the RN4CAST study design and gives a concise introduction to the fundamental principles of the two main statistical techniques that are used in the PhD study: multilevel modeling and structural equation modeling.

The next part of this PhD study is organized into four thematic chapters that all extend on RN4CAST data or study concepts. The first core chapter shows that a high proportion of nurses working in European hospitals reported having performed tasks below their skill level during their last shift, which is a newly theorized concept in the evaluation of nurses' work structure. In line with anecdotal evidence, we confirm a pronounced overall effect of being a foreign trained nurse from a developing country and an increase in reports of tasks below skill level. In the second core chapter, we hypothesize and confirm that nursing work environment components exhibit different effects on nurse outcomes at different systems levels. These new linkages have rich potential for policy makers and hospital human resources managers, as it may provide them with empirically-based strategies in designing level-specific interventions to improve nurse wellbeing and patient safety. In a third core chapter, we show that necessary nursing care left undone, another only recently theorized concept, is associated with a poor nursing work environment, performing more tasks below skill level, and a complex interplay between nurse staffing and nurses' education levels. We also showed that nursing care left undone explains the association between these components of nursing organization of care and patient satisfaction. This had recently been theorized, but was not evaluated statistically. In the fourth and last thematic chapter, we evaluate measurement invariance of our work environment survey items in Belgium. This study evidenced that some items are not invariant across Dutch and French speaking nurses. This study also allowed us to conclude that

nursing unit managers, when compared to staff nurses, hold more positive views of several work environment dimensions.

In a concluding chapter, we argue that the mounting evidence linking inputs of nursing to quality and patient safety allows human resources managers and policy makers to identify the impact of different systems policies of nursing organization on outcomes. It should at the very least trigger a critical reflection on good practices and on the transferability, implementation and evaluation of investments in nurse education, work environments and staffing levels.

Samenvatting

De afgelopen twee decennia werden gekenmerkt door toenemende evidentie omtrent het belang voor ziekenhuizen om te investeren in een goede personeelsinzet van verpleegkundigen, een hoge proportie bachelor-opgeleide verpleegkundigen, en een uitstekende werkomgeving voor verpleegkundigen. Deze kenmerken van de organisatie van verpleegkundige zorg zijn immers gerelateerd aan een lager risico op overlijden van patiënten en een hogere graad van tevredenheid onder patiënten en verpleegkundigen. Tot voor enkele jaren geleden was deze evidentie hoofdzakelijk gebaseerd op onderzoeksbevindingen in een Amerikaanse studiepopulatie. Het Registered Nurse Forecasting (RN4CAST) consortium bevestigde deze bevindingen in een Europese setting. Deze doctoraatsstudie bouwt verder op RN4CAST. Dit werk stimuleert het gebruik van hedendaagse statistische technieken door het verfijnen van onze kennis over mechanismen die ten grondslag liggen aan gevestigde associaties, door het bestuderen van verschijnselen op verschillende organisatieniveaus, en door het bevorderen van een evaluatie van de gelijkenis van concepten tussen groepen. Twee belangrijke uitdagingen worden aangepakt.

Een eerste uitdaging betreft de afstemming tussen statistische technieken en de vele uitgebreide theorieën die werden voorgesteld om te verklaren waarom kenmerken van de organisatie van de verpleegkundige zorg geassocieerd zijn met zorgresultaten en met het welzijn van verpleegkundigen. Een eerste bijzonder uitdagend aspect in de statistische analyse betreft de complexe hiërarchische aard van de gegevens. Een hiërarchische datastructuur wordt ook een multilevel datastructuur genoemd. De RN4CAST studie wordt gekenmerkt door een complexe vier-level structuur. Deze treedt op omdat

het studiedesign aansluit bij de clustering van verpleegkundigen en patiënten (niveau 1) in verpleegeenheden (niveau 2) in ziekenhuizen (niveau 3) in landen (niveau 4). Deze complexe structuur vereist het gebruik van multilevel modelleringstechnieken. Een tweede complexiteit komt voort uit de theoretische specificatie van mediatoren en moderatoren. Mediatoren illustreren de mechanismen die verklaren waarom associaties worden waargenomen. Moderatoren zijn contextuele variabelen die de associatie tussen twee variabelen beïnvloeden. Een derde complexiteit heeft betrekking op het begrijpen van onderliggende dimensies, ook wel latente variabelen genoemd, uit een set van geobserveerde variabelen. Hypotheses waarbij latente variabelen worden gespecificeerd vereisen het gebruik van zogenaamde structurele modellen, ook wel aangeduid met de term structurele vergelijkingsmodellen. De analyse van mediatoren, moderatoren en latente variabelen moet worden uitgevoerd binnen een multilevel analyse om overeen te stemmen met theoretische modellen.

Een tweede uitdaging betreft het evalueren van meetinvariantie wanneer vergelijkingen tussen groepen worden gemaakt. Het evalueren van meetinvariantie betreft de vraag of meetinstrumenten er in slagen een concept op dezelfde manier te meten in groepen die men wil vergelijken. Een gebrek aan meetinvariantie kan leiden tot de onjuiste interpretatie van wat door de onderzoeker werd geïnterpreteerd als echte verschillen tussen groepen. Meetinvariantie wordt bestudeerd aan de hand van latente variabelen en maakt, net als mediatie en moderatie, gebruik van structurele modellen.

In een eerste inleidend hoofdstuk geven we een overzicht van de literatuur omtrent patiëntveiligheid. Vervolgens komen de meest voorname theoretische modellen en studies in dit onderzoeksdomein aan bod. Dit hoofdstuk eindigt met de onderzoeksdoelstellingen van deze doctoraatsstudie. Een tweede hoofdstuk geeft een overzicht van het RN4CAST studiedesign en een beknopte inleiding in de fundamentele beginselen van de twee belangrijkste statistische technieken die worden gebruikt in deze doctoraatsstudie: multilevel modellen en structurele modellen.

Vervolgens volgen vier thematische hoofdstukken die gebruik maken van de RN4CAST studiegegevens. Het eerste thematisch hoofdstuk toont aan dat een groot deel van de verpleegkundigen die werkzaam zijn in Europese ziekenhuizen aangeeft dat ze taken onder hun competentieniveau

uitvoeren. Dit is een recentelijk getheoretiseerd begrip dat meer en meer aandacht krijgt in de evaluatie van de werkstructuur van verpleegkundigen. In lijn met anekdotische rapporten bevestigde deze studie dat verpleegkundigen die werden opgeleid in een ontwikkelingsland meer van deze taken rapporteren. In een tweede thematisch hoofdstuk bevestigen we de hypothese dat componenten uit de werkomgeving van verpleegkundigen verschillende effecten op het welzijn van verpleegkundigen uitoefenen op verschillende systeemniveaus. Deze inzichten laten toe dat beleidsmakers en human resources managers empirisch-gebaseerde en systeemniveau-specifieke strategieën kunnen ontwikkelen en implementeren wanneer men het welzijn onder verpleegkundigen wil verhogen. Een derde thematisch hoofdstuk handelt over noodzakelijke verpleegkundige zorg die niet wordt uitgevoerd. We tonen aan dat dit geassocieerd is met een slechte verpleegkundige werkomgeving, het uitvoeren van meer taken onder het vaardigheidsniveau, en een complex samenspel tussen verpleegkundige personeelsinzet en het opleidingsniveau van verpleegkundigen. Er wordt tevens aangetoond dat nagelaten verpleegkundige zorg het verband verklaart tussen deze kenmerken van verpleegkundige organisatie en de tevredenheid van de patiënt. Dit werd reeds getheoretiseerd, maar niet statistisch geëvalueerd. In het vierde en laatste thematisch hoofdstuk evalueren we de meetinvariantie inzake de perceptie van Belgische verpleegkundigen omtrent hun werkomgeving. Uit deze studie blijkt dat sommige items voor het meten van de werkomgeving niet invariant zijn voor wat betreft de vergelijking tussen Nederlandstalige en Franstalige verpleegkundigen. In deze studie wordt ook aangetoond dat verpleegkundig managers de kwaliteit van de werkomgeving hoger inschatten dan hun verpleegkundigen. Deze verschillen zijn echte verschillen die niet onderhevig zijn aan problemen met betrekking tot meetinvariantie.

In een afsluitend hoofdstuk argumenteren we dat er voldoende evidentie is dat kenmerken van de organisatie van de verpleegkundige zorg gerelateerd zijn aan de kwaliteit en de veiligheid van de patiëntenzorg. Dit moet op zijn minst leiden tot een kritische reflectie over goede praktijken en over de overdraagbaarheid, implementatie en evaluatie van investeringen in de opleiding van verpleegkundigen, hun werkomgeving, en personeelsinzet.

9 CURRICULUM VITAE

Biography

Luk Bruyneel is a PhD student at the Department of Public Health and Primary Care at KU Leuven, Belgium. He received a Bachelor in Nursing from KU Leuven University College (2006) and a Double Degree Master in Nursing from KU Leuven and the University of Alicante, Spain (2008), both with great distinction. After graduation, he joined the Registered Nurse Forecasting (RN4CAST) consortium, directed by Professor Walter Sermeus at KU Leuven. From 2009 to 2011, this consortium of 16 countries studied the association between the organization of nursing care, nurse wellbeing and patient outcomes. In 2010 Luk became project manager and was largely responsible for the statistical analysis of the unique international RN4CAST data set and for disseminating the study findings.

Luk became passionate about this field of research, with a particular interest in knowledge gaps that result from statistical challenges. In 2012 he started his PhD under the supervision of Professor Walter Sermeus and Professor Emmanuel Lesaffre. It is devoted to the application of contemporary statistical techniques for analyzing the complex associations and cross-group comparisons that emerge in nursing- and patient-oriented health services research. During his PhD, Luk earned a Master in Economic Policy (2013) and a Master in Statistics with great distinction (2015) from KU Leuven.

Selected honors include the KU Leuven Professor Evers Award (2008) for innovative research in the field of nursing for his master thesis in which he critically evaluated Belgian nurses' reports on the quality of their work environment. Luk also received two scholarships for participation to the Young Forum Gastein (2011, 2014). This initiative is funded by the European Health Forum Gastein and the European Commission and brings together promising EU health researchers and policy makers. He was also recipient of a conference award from the Association of Schools of Public Health in the European Region for his work on migratory nurses' reports on performing tasks below their skill level (2012).

From January to April 2015, Luk is a visiting scholar at King's College London. For this research stay he was awarded a scholarship by the KU Leuven Research Council within the YouReCa Junior Mobility Programme.

Publications in international peer-reviewed journals

Bruyneel, L., Li, B., Ausserhofer, D., Dumitrescu, I., Lesaffre, E., Smith, H. L., Sloane, D. M., Aiken, L. H., Sermeus, W. (2015). Organization of hospital nursing, provision of nursing care, and patient assessments of care in Europe. *Medical Care Research and Review, submmitted for publication.*

Griffiths, P., Dall'Ora, C., Simon, M., Ball, J., Lindqvist, R., Rafferty, A.-M., Schoonhoven, L., Tishelman, C., & Aiken, L. H. for the RN4CAST consortium (2014). Nurses' shift length and overtime working in 12 European countries: the association with perceived quality of care and patient safety. *Medical Care*, *52*(11), 975–81.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Griffiths, P., & Sermeus, W. (2014). Staffing and education of nurses and hospital mortality in Europe-Authors' reply. *Lancet*, *384*(9946), 851–2.

Bruyneel, L., Li, B., Squires, A., Spotbeen, S., Meuleman, B., Lesaffre, E., & Sermeus, W. (2014). Bayesian Multilevel MIMIC Modeling for Studying Measurement Invariance in Cross-group Comparisons. *Medical Care [Epub ahead of print]*.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E., McHugh, M. D., Moreno-Casbas, M. T., Rafferty, A. M., Schwendimann, R., Scott, P. A., Tishelman, C., van Achterberg, T., & Sermeus, W. (2014). Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *Lancet*, *383*(9931), 1824–30.

Li, B., Bruyneel, L., & Lesaffre, E. (2014). A multivariate multilevel Gaussian model with a mixed effects structure in the mean and covariance part. *Statistics in Medicine*, *33*(11), 1877–99.

Ausserhofer, D., Zander, B., Busse, R., Schubert, M., De Geest, S., Rafferty, A. M., Ball, J., Scott, A., Kinnunen, J., Heinen, M., Strømseng Sjetne, I., Moreno-Casbas, T., Kózka, M., Lindqvist, R., Diomidous, M., Bruyneel, L., Sermeus, W., Aiken, L. H., & Schwendimann, R. (2013). Prevalence,

patterns and predictors of nursing care left undone in European hospitals: results from the multicountry cross-sectional RN4CAST study. *BMJ Quality & Safety*.

Seys, D., Deneckere, S., Sermeus, W., Van Gerven, E., Panella, M., Bruyneel, L., Mutsvari, T., Bejarano, R. C., Kul, S., & Vanhaecht, K. (2013). The Care Process Self-Evaluation Tool: a valid and reliable instrument for measuring care process organization of health care teams. *BMC Health Services Research*, *13*, 325.

Najjar, S., Hamdan, M., Baillien, E., Vleugels, A., Euwema, M., Sermeus, W., Bruyneel, L., & Vanhaecht, K. (2013). The Arabic version of the hospital survey on patient safety culture: a psychometric evaluation in a Palestinian sample. *BMC Health Services Research*, *13*(1), 193.

Aiken, L. H., Sloane, D. M., Bruyneel, L., Van den Heede, K., & Sermeus, W. (2013). Nurses' reports of working conditions and hospital quality of care in 12 countries in Europe. *International Journal of Nursing Studies*, *50*(2), 143–53.

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), 281–91.

Bruyneel, L., Li, B., Aiken, L. H., Lesaffre, E., Van den Heede, K., & Sermeus, W. (2013). A multi-country perspective on nurses' tasks below their skill level: reports from domestically trained nurses and foreign trained nurses from developing countries. *International Journal of Nursing Studies*, *50*(2), 202–9.

Squires, A., Bruyneel, L., Aiken, L. H., Van den Heede, K., Brzostek, T., Busse, R., Ensio, A., Schubert, M., Zikos, D., & Sermeus, W. (2012). Cross-cultural evaluation of the relevance of the HCAHPS survey in five European countries. *International Journal for Quality in Health Care*, *24*(5), 470–5.

Squires, A., Aiken, L. H., van den Heede, K., Sermeus, W., Bruyneel, L., Lindqvist, R., Schoonhoven, L., Stromseng, I., Busse, R., Brzostek, T., Ensio, A., Moreno-Casbas, M., Rafferty, A. M., Schubert, M., Zikos, D., & Matthews, A. (2013). A systematic survey instrument translation process for multi-country, comparative health workforce studies. *International Journal of Nursing Studies*, *50*(2), 264–73.

Aiken, L. H., Sermeus, W., Van den Heede, K., Sloane, D. M., Busse, R., McKee, M., Bruyneel, L., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Tishelman, C., Scott, A., Brzostek, T., Kinnunen, J., Schwendimann, R., Heinen, M., Zikos, D., Sjetne, I. S., Smith, H. L., & 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.

Van den Heede, K., Florquin, M., Bruyneel, L., Aiken, L. H., Diya, L., Lesaffre, E., & Sermeus, W. (2013). Effective strategies for nurse retention in acute hospitals: a mixed method study. *International Journal of Nursing Studies*, *50*(2), 185–94.

Sermeus, W., Aiken, L. H., Van den Heede, K., Rafferty, A. M., Griffiths, P., Moreno-Casbas, M. T., Busse, R., Lindqvist, R., Scott, A. P., Bruyneel, L., Brzostek, T., Kinnunen, J., Schubert, M., Schoonhoven, L., & Zikos, D. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. *BMC Nursing*, *10*, 6.

Bruyneel, L., Van den Heede, K., Diya, L., Aiken, L. H., & Sermeus, W. (2009). Predictive validity of the International Hospital Outcomes Study questionnaire: an RN4CAST pilot study. *Journal of Nursing Scholarship*, *41*(2), 202–210.