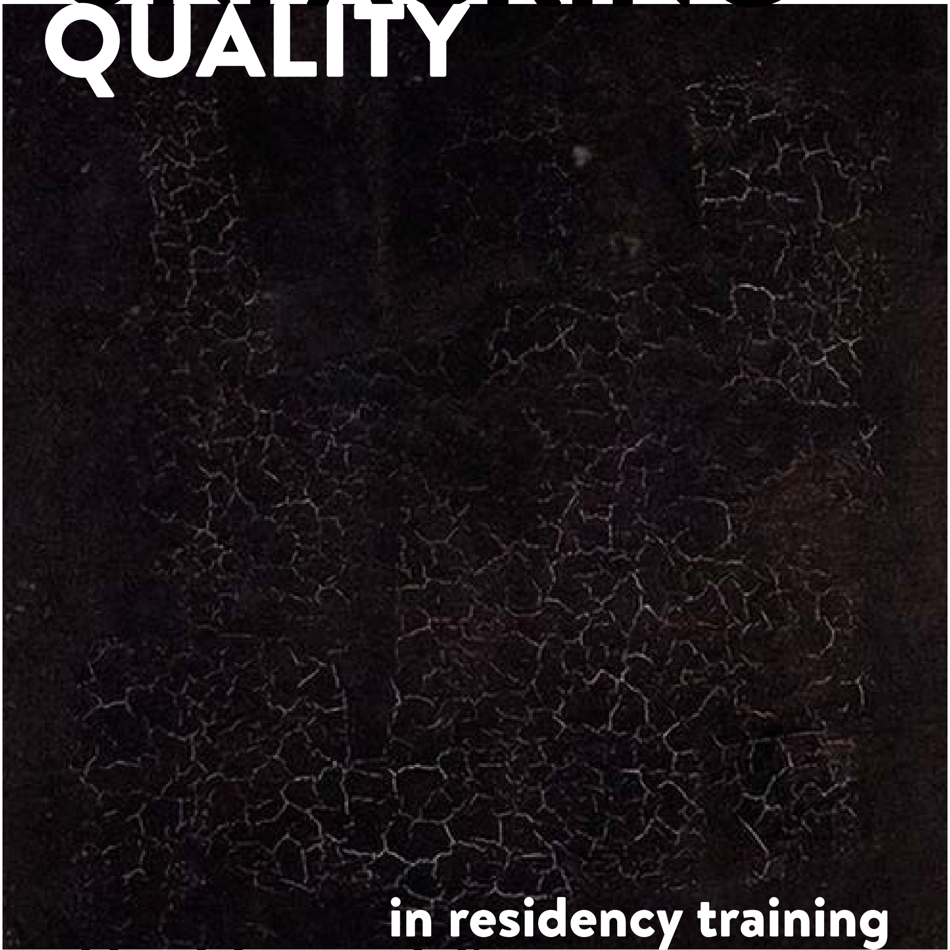
Alina Smirnova

UNPACKING QUALITY



in residency training and health care delivery

ISBN: 978-94-6233-912-5

The research reported here was carried out at

Maastricht University

and Professional Performance Research Group, Academic Medical Centre, Amsterdam.

School of Health Professions Education

The research reported here is part of the research project 'Quality of Clinical Teachers and Residency Training Programs' co-financed by the Dutch Ministry of Health, the Faculty of Health and Life Sciences of Maastricht University, and the Academic Medical Centre, Amsterdam.

Layout: Sander de Haan – www.youneedtodesign.nl
Cover design: Matt Kirkwood
Printed by: Gildeprint – www.gildeprint.nl

Printed with financial support from Maastricht University, Academic Medical Centre, the Dutch Association for Medical Education (NVMO) and MEDOX.nl

UNPACKING QUALITY IN RESIDENCY TRAINING AND HEALTH CARE DELIVERY

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday, 4th of April 2018, at 14:00 hours.

by

Alina Aleksandrovna Smirnova

Supervisors:	Prof.dr. C.P.M. van der Vleuten Prof.dr. M.J.M.H. Lombarts, University of Amsterdam
Co-supervisors:	Prof.dr. O.A. Arah, University of California Los Angeles Dr. R.E. Stalmeijer
Assessment Committee:	Prof.dr. W.N.K.A. van Mook (chairman) Dr. J.A.A.M. van Diemen – Steenvoorde, Health and Youth Care Inspectorate, Ministry of Health, Welfare and Sport Prof.dr. I.C. Heyligers Prof.dr. F.W.J.M. Smeenk Prof.dr. C. Wagner, Dutch Institute for Health Services Research

To my parents and Matt
for their unconditional love

Table of contents

CHAPTER 1

Introduction

Introduction

In 2016 a visitation report of the otolaryngology department in the Academic Medical Centre in Amsterdam fell into the hands of an investigative television program, made waves in the news and ultimately cost the head of the department her job. The visitation committee brought to light the unsafe learning climate at the department which was considered to be so toxic that it endangered patient safety. The committee stated that a safe learning climate, in which residents are free to discuss their uncertainties in regard to patient care as well as their mistakes, is a prerequisite for safe patient care. The otolaryngology department's culture was characterized by fear mongering and bullying. This resulted in reputation damage for both the department and the hospital. This news came shortly after another news-making scandal at University Medical Centre Utrecht where a culture of fear and disagreements between staff contributed to two patients' deaths being unreported while putting countless other patients at risk.

Both of these cases took place in large university teaching hospitals, where future doctors (medical students and residents) also receive their training. The cases show that the training environment and the clinical environment are intrinsically intertwined and what happens in one ultimately affects the other. The wide media attention that both of these cases received shows the public's invested interest in quality and safety of patient care, but also demonstrates the damaging effect a department's culture and climate can have on functioning of a whole department and, more importantly, concerns about patient safety.

This thesis investigates the relationship between quality of post-graduate medical education (PGME) and quality of care in the Netherlands. PGME is a big business, costing taxpayers $15 billion per year in the United States alone.[1] In the Netherlands, costs of educating doctors in specialist training amount to about €1 billion per year, which are higher per capita than the United States.[2] Fueled by increasing healthcare costs and patients' expectations, there is an increasing need for public accountability in how residents are educated and incorporated into clinical contexts. This need has led to recent reform calls for greater transparency in PGME[1] and greater use of patient outcomes in medical education research.[3,4] The first step towards designing PGME programs that are better aligned with the goals of high quality healthcare is understanding the relationship between quality of PGME and quality of care in clinical teaching departments. This thesis aims to shed light on the relationships between PGME quality and quality of care.

The purpose of this section is to introduce the setting of medical specialist training as well as familiarize the reader with the major concepts and latest developments in the quality of PGME and care relevant to the research in this thesis. This chapter also provides a critical overview of research in medical education on the relationship between PGME and quality of care and argues that the current state of research is limited by its focus on the resident. The chapter will then introduce the main research questions concluding with an overview of the studies included in the thesis.

Medical specialist training

After finishing medical school, newly qualified doctors undergo a period of supervised training aimed at gaining specialization in a particular specialty usually referred to as post-graduate medical education (PGME) in the United Kingdom and Canada or graduate medical education (GME) training in the United States. In this thesis, we will use the term PGME or residency training referring to this period. This training period usually lasts a number of years, with or without an initial "internship" phase, in which the trainee practices medicine without a particular specialization route. Sometimes, the "internship" is included as a part of the medical specialist training program. Although the terminology, content and length of training differ between countries, in all cases the trainees gain progressive independence as they move further through the training taking more responsibility for the care of their patients. The end of the training period marks the point when the trainee is officially allowed to practice without supervision. Table 1 is aimed at familiarizing the reader with the peculiarities of medical specialist training in various countries. Since the research studies in this thesis were conducted in the Netherlands, this training is compared with other countries below.

Table 1. Overview of medical specialist training and terminology

	The Netherlands	United Kingdom	USA	Canada
Pre-specialist training phase	6 years (entry after high school)	5 years (entry after high school)	4 years (entry after bachelor)	4 years (entry after bachelor, except Quebec[1])
Initial training phase (internship)	Resident-not-in-training (1-3 years, common)	Foundation doctor (2 years, required)	Intern or transitional year doctor (1 year, optional)	Resident (PGY1) (1 year, required)
Specialist training phase	Resident	Specialty registrar	Resident	Resident (PGY 2 and up)
Typical length of training	3-6 years	3-6 years	3-7 years	2-6 years
Position after certification	Attending physician, specialist	Consultant	Attending physician, faculty	Attending physician, faculty

[1] In Quebec, graduates with a diploma of college studies (Diplôme d'Études Collégiales; "DEC") in a pre-university program are also eligible to apply if they completed a 90-credit program from a Quebec university.

Organization of medical specialist training in the Netherlands

In the Netherlands, upon attainment of a medical degree, Dutch medical graduates automatically qualify for registration as a medical doctor in a national registry system (BIG-register), which allows them to work as licensed doctors and enter medical specialty training. Since entry into medical specialist training programs is competitive, newly qualified doctors usually undergo a 1-3 years of internship equivalent as resident-not-in-training and/or complete a PhD program in order to increase their chances of gaining a training spot in a specialty of their choice. There are 28 recognized hospital-based medical specialties, 6 non-hospital based specialties and 12 profile specialties. In 2016, there were 10231 registered residents in the Netherlands.[5] Hospital-based specialist training consists of clinical rotations in one of eight clusters of hospitals, each organized around a single university teaching hospital which coordinates the training programs according to the national curriculum dictated by the specialty bodies. In addition to clinical rotations, residents follow formal educational sessions, offered either by their specialty body or by the teaching hospital. In the Dutch context, residents are trained based on the competency framework, in which the seven CanMEDS roles have been supplemented with four additional themes: medical leadership, patient safety, elderly care, cost-effectiveness.[6] Residency training in the Netherlands has recently been "modernized", whereby residents are able to individualize the national education plan by gaining recognition for earlier competencies, shortening the total length of training through entrustment of professional activities or through choosing in-depth electives in their area of interest. Progress exams ensure residents fulfill the European requirements of specialist training. Each teaching hospital coordinates residency training programs through the hospital-wide education committee (in Dutch *Centrale Opleidingscommissie*, COC), which consists of the local program directors (or specialty tutors) in addition to residents' representatives, assistant program directors and representatives of the hospital's board of directors. At the end of the residency training, the resident can register as a specialist with the Royal Dutch Medical Association (in Dutch *Koninklijke Nederlandsche Maatschappij tot bevordering der Geneeskunst*, KNMG).

Quality of medical specialist training

"Quality" is a subjective term defined by the different stakeholders, who have a vested interest in improving quality.[7,8] In general terms quality can be defined as 1) conformance to process requirements[9], or 2) the extent to which the product/service meets the expectations and wishes of the end-user/consumer/customer.[10] So quality is a relative term the meaning of which can vary based on the stakeholders that are being considered.[7]

The stakeholders in PGME are the residents, residency programs, hospital-wide education committees, teaching hospitals and clinical teaching departments, the Ministry of Health, and ultimately patients and society. Teaching hospitals, hospital-wide education committees or individual residency programs, can define quality as audit success, higher ranking among other programs or higher graduate certification rates. Residents are also important stakeholders because they expect residency training to prepare them to be high-performing doctors. For residents, a better training program may mean performing better on exams, on in-training evaluations and being prepared to take on responsibilities in practice later on. High quality residency programs can be characterized by high quality learning environments that ensure high quality learning for residents and safe and effective care for patients. From the residents' perspectives, a high quality learning environment includes a safe and supportive learning climate and high quality teaching, which facilitates residents' meaningful learning experiences with patient care, while honing important clinical skills and develop professional identity, self-confidence and progressive autonomy in preparation for unsupervised practice.[11] In authoring the Global Standards for Quality Improvement in Postgraduate Medical Education, the World Federation for Medical Education has gone as far as saying that the goal of medical education is "improving health of all peoples"[12], essentially uniting the medical education goals with those of high quality patient care. Ultimately, patients and society should benefit from PGME that delivers better-prepared doctors.

However, since notions of quality evolve over time, we will first outline how ideas of quality in PGME have changed over the last century before discussing current thinking about quality in PGME and how it can be monitored and improved.

Historical perspectives on PGME quality

Our conceptions of quality of PGME have evolved over time, beginning with Flexner's 1910 report on *Medical Education in the United States and Canada*.[13] The Flexner report advocated for increasing standards in medical training and the implementation of scientific knowledge in patient care compatible with latest scientific discoveries. To be a good doctor meant to be a good scientist who possessed a mindset of scientific inquiry without separation in research or practice, where knowledge and practice were based on observation, experiment and induction.[14] As a result, science entered the mainstream of medical education as preparation for one's work as a future physician. Since the 1970's fueled by fragmentation of the medical curriculum and de-emphasis on basic skills, the role of the doctor has been transformed beyond a medical expert or a scientist towards a doctor who is competent to handle the population's needs.[15] The role of PGME also shifted towards instilling and evaluating a number of competencies that doctors should demonstrate at the end of their training, which would ultimately allow them to better serve their population. This type of PGME training was called outcome-based, or competency-based, medical education (CBME). Since then, the focus in medical education has been on defining, refining and evaluating these competencies.[15]

Current conceptions of quality in PGME

While the Flexner Report shaped the landscape of medical education for the rest of the 20th century, the publication of the report by the committee on the Quality of Health Care in America in 2001 called *Crossing the Quality Chasm: A New Health System for the 21st Century* changed the course of healthcare at the turn of the century.[16] By putting the patient's needs first, the report fundamentally shifted the role of the physician in patient care. Physicians are now seen less as lone practitioners of science, but members of patient-care teams responsible for delivering care that is safe, effective, patient-centered, timely, efficient and equitable. As a result, physicians could no longer function as isolated islands of medical expertise, but were expected to work collaboratively in teams to deliver care and regard the patient as an equal partner in the decision-making process.[17] This complexity is highlighted in the *Lancet* report *Health professionals for a new century: transforming education to strengthen health systems in an interdependent world,* which reinforces the notion that high quality healthcare education should be patient-centered and team-based.[18]

In order to align new standards of quality in patient care with medical education and to prepare physicians for rapidly changing practice, in 2005 the Royal College of Physicians and Surgeons of Canada released its seminal report on competencies, evidenced on societal need, that all physicians should possess in order to meet the needs of society and the patients they serve, which quickly became one of the standards for medical education worldwide.[19] The CanMEDS framework consists of seven competencies: medical expert, communicator, collaborator, manager, health advocate, scholar and professional. Similarly, the Accreditation Council for Graduate Medical Education (ACGME) started its Outcome Project in 1998 with the goal of developing physicians' competence in six domains: patient care, medical knowledge practice-based learning and improvement, inter-personal and communication skills, professionalism, and systems-based practice.[20] All competency-based frameworks have a major focus on developing abilities in pre-defined areas of competence, de-emphasis of time spent in training, and a major focus on the trainee as an active participant in their own learning (learner-centeredness).[15,21]

Regardless of CBME's popularity, it has experienced mounting criticism in the past years. Competencies have been seen as reductionist - something that simplifies the complexity into a list of competencies to be 'checked off' while losing sight of reality and complexity of medical practice.[22] There is fear that supervisors in residency training are too busy filling out evaluation forms, which residents consider as a hurdle or an administrative task, which could even result in harm being done if competency is identified too early.[23,24] Others have criticized CBME for lack of evidence on how it ultimately affects patient outcomes.[15,25]

Monitoring and improving PGME quality

Ensuring quality of medical specialist training in the Netherlands is the responsibility of the KNMG, materialized in the College of Medical Specialists (in Dutch *College Geneeskundige Specialismen,* CGS) and the Registration Committee for Medical Specialists (in Dutch *Registratiecommissie Geneeskundig Specialisten*, RGS). The RGS provides accreditation to Dutch residency training programs for up to 5 years based on an external program evaluation in the form of a visitation (summative evaluation). Aside from periodic summative assessments in the form of visitations, training programs are required to plan and execute a continuous quality improvement cycle as a part of their efforts to maintain and improve quality in residency training (formative evaluation). The hospital-wide education committees, together with local program directors and clinical supervisors, are responsible for monitoring and improving the internal quality of training. The Scherpbier 2.0 report describes four major areas for continuous improvement in PGME, including organization and development, learning climate, professionalization of residency training, and competency development.[26] For this purpose, several instruments have been developed to gain better insight into the functioning of their training, including the System for Evaluation of Teaching Qualities (SETQ)[27], the Maastricht Clinical Teaching Questionnaire (MCTQ)[28], Evaluation and Feedback for Effective Clinical Teaching (EFFECT)[29], the Dutch Residency Educational Climate Test (D-RECT)[30], TeamQ[31] for evaluation of teamwork of clinical educators, and end-of-rotation evaluations. Training programs may use these instruments in order to get feedback into strengths and areas for own quality improvement efforts, depending on the needs of the program.[26] In this thesis we will focus on two aspects of PGME quality: the residency learning climate and clinical teaching, both of which we will describe in more detail below.

The residency learning climate

The concept of the learning climate originates in organizational psychology. Organizational climate generally constitutes the collective perception of individuals about an organization's policies, practices and procedures. Denison distinguished organizational climate from culture in that climate researchers usually use quantitative methods to study perceptions of observable practices and procedures and their effects on the organizational performance, while culture researchers study the deep underlying assumptions and values of an organization using in-depth observational qualitative methods.[32,33] While both are important in organizations, climate research is advantageous because it focuses on the effects on performance that is generalizable across organizations. Previous research has linked organizational climate in healthcare to both employees' and patients' outcomes, although the link with patient outcomes is less consistent.[34] In medical education literature, the concept of the "learning" climate (also called educational climate or climate for learning) has taken off in both undergraduate[35-39] and post-graduate medical training[30,40,41], for which a number of instruments have been developed.[42] The learning climate reflects how supportive the clinical learning environment is for residents' learning. Supportive learning climates in residency training have been associated with better residents' outcomes including better performance on examinations[43], lower likelihood of burnout[44], lower likelihood to leave practice[45], as well as better quality of clinical teaching.[46] However, the evidence of the relationship between learning climates in residency training and patient care is missing.

Clinical teaching quality

Clinical teachers also play an important role in ensuring PGME quality in residency training. They are responsible for establishing a safe and supportive learning climate for residents to discuss their concerns about patient care as well as reflect on their strengths and weaknesses.[47] Furthermore, faculty supervise residents' clinical activities to ensure patient safety, as well as assess and coach residents to aid their professional development.[48] Moreover, faculty with good clinical teaching skills are more likely to be considered role models by residents.[49] While a lot of research has focused on evaluation of clinical teaching skills[50-52], much less light has been shed on the effects of clinical teaching in PGME on patient care. In a systematic review, Farnan et al. showed that enhanced supervision was beneficial for both educational and patient outcomes.[53] In another systematic review Van der Leeuw et al. reported that in addition to adequate supervision, evaluation of and attention to individual competence of residents was also associated with more favourable clinical outcomes.[54] While these reviews have focused on specific supervisory behaviours, which are also attributable to high quality clinical teaching, in this thesis we will focus on the role that overall teaching quality and role modeling plays in clinical care and patient outcomes.

Measuring quality in health care

Similar to quality in PGME, our conceptions of quality in health care experienced major shifts over time. For instance, the definition of "health" itself has been recently updated from its original 1946 World Health Organization (WHO) definition as a state of 'complete well-being' to 'the ability to adapt and implement own control, in light of the physical, emotional and social challenges of life'.[55] Aging populations, high prevalence of chronic and non-communicable diseases are all placing higher demands on healthcare systems and their organization, with patients gaining a central role.[16,56] Issues such as patient-centeredness and shared-decision making have become as important as treatment of disease or prevention of complications.[16] With the increasing focus on the patient, the process of delivery and the responsibility for the outcomes of care have shifted from individual providers towards multi-disciplinary teams of providers and systems that deliver healthcare that is safe, effective, efficient, personalized, timely and equitable.[16] However, achieving high value in healthcare requires healthcare to be monitored and evaluated.[57,58]

The concept of evaluation of quality in healthcare has been around since the 1960s, when Avedis Donabedian first defined the three categories of metrics in healthcare: structure, process, and outcomes.[59] Structure usually refers to the presence of certification or accreditation of the institution, as well as the organization, provision of necessary facilities, and the presence of protocols. Process measures usually refer to adherence to clinical guidelines or following protocols. Outcome measures have to do with patient outcomes, such as mortality, morbidity, complications, readmissions, or length of stay. Both structure and process measures are required to achieve improvement in patient outcomes. The resulting quality assurance (QA) movement focused on meeting the requirements for structure and processes in healthcare organizations.

Since Donabedian, our approach to quality management in health care has undergone two major shifts. The first shift took place in the 1990's from quality assurance (QA) to continuous quality improvement (CQI) in health care.[60,61] While QA involves meeting pre-set standards and invokes punitive measures if these standards are not met, CQI, on the other hand, focuses on future improvement in addition to meeting the requirements. CQI involves cyclical measurement and improvement activities according to the Plan-Do-Check-Act (PDCA) cycle aimed at achieving targeted improvements in outcomes and processes of care. Critics of this era, in which we are still finding ourselves, assert that the resulting obsession of measurement in healthcare has led to an explosion of evaluation tools, associated over-measurement, reductionism and over-bureaucratization of healthcare delivery.[57]

With growing emphasis on the role of the patient, patient-centeredness and shared decision-making in healthcare, patient-related measures have also gained increasing importance in evaluation of health care quality. Patient-reported experience measures and patient-reported outcome measures (PREM's and PROM's) are becoming as important as the traditional measures of care quality. Taken together, these measurements in health care are used today to guide improvement in quality and safety of health care, assist patients in making choices regarding providers and healthcare organizations, for pay-for-performance schemes, and to define efficiency in healthcare.[62]

Research into quality of care in medical education

Quality of care research in medical education has focused primarily on studying the quality of care delivered by residents focusing on the aspects of training that are necessary to ensure that residents deliver safe, high quality patient care (Figure 1). To this extent, several frameworks have been developed.[63-66] These frameworks have primarily focused on studying the quality of care delivered by residents and identifying the conditions necessary to ensure that residents deliver safe, high quality patient care. We summarize these frameworks below.

In assessing the ultimate impact of medical training on care delivered by graduates, Tamblyn[63] considers the impact of optimal, average and suboptimal medical practice on patient populations, as well as the degree to which individual physicians contribute to the standard of care received and/or resulting health outcomes, taking prevalence and treatability of the condition as well as risk of adverse effects into account. Asch et al.[65] also considered the effect of the institution of training on the quality of care delivered by physicians after graduation as a logical way to evaluate training programs. Although the authors recognize that physician characteristics can affect the quality of care delivered by a physician, their focus is to discern the effect of the site of training. They present a research agenda to study this relationship and discuss a number of important issues that need to be considered when researching the relationship. Kalet et al.[66] proposed another framework for educationally sensitive patient outcomes to be used in medical education research and, in particular, to investigate effects of education on the quality of care delivered by residents. The authors recognized that patient outcomes need to be both important and sensitive to provider education. They identified two interdependent educationally sensitive patient outcomes: patient activation and clinical microsystem activation, both of which could be evaluated validly and reliably using existing tools and methods, across the educational continuum. Dauphinee[64] argued for the use of patient-based and patient-reported outcomes at all stages of training to the extent that these are productive and technically feasible, as well as identifying high-risk behaviours early on in training. The outcomes should be predictive of the residents' performance expected at the next level of functioning.

Although the different approaches propose different patient outcomes relevant to PGME, they also have some similarities (see Figure 1 on the next page). One similarity is that they are based on the assumption that inputs (PGME characteristics) will have an effect on the development of competencies of an individual trainee, which in turn will affect health care provided by the trainee and result in improved patient outcomes. This linear thinking focuses on the (competent) doctor as an end-product of PGME, rather than considering the patient as the starting point of all PGME activities. Although the current conceptual models recognize the multi-factorial nature of patient outcomes (i.e. outcomes do not just depend on an individual health care provider) and complexity of healthcare practice, they sideline complexity as too difficult to study in favour of defining outcomes that can be attributed to a single provider. As mentioned earlier, the empirical research based on the current conceptual framework can be divided into two areas: 1) quality of care provided by individual residents, and 2) elements of institutions, programmes and systems that can support optimal resident development in relation of improving patient outcomes. We summarize the findings of this research in the following paragraphs.

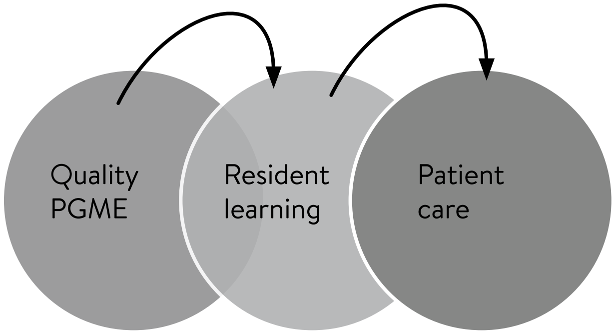


Figure 1. Current model of the relationship of PGME quality and patient care, mainly conceptualized through PGME effects on resident learning

Research on quality of care provided by residents

The empirical research on quality of care provided by individual residents shows that although residency is not without risk, residents generally deliver quality of care that is of similar quality to that of specialists, provided a number of important conditions are met.[54] In particular, residency transitions, including the start of residency training or the transition to unsupervised practice are the most risky to the patient.[54] Resident participation in care is associated with increased lab test use[67] and increased operation time and patient morbidity without increased mortality.[68] However, important aspects of residency training could minimize these risks including suitability of the training to the competence level of the trainee, allowance of extra operating time for residents, and focus on development and evaluation of competencies.[54] In particular, good quality of clinical supervision has also been found to be important for better patient outcomes in care provided by residents.[53,54,69] Also, specific initiatives aimed at improving quality of care, including patient handoff programs and simulation training, have shown to improve targeted patient outcomes.[70,71] However, research on the value of workplace-based assessments, such as the mini-Clinical Evaluation Exercise (mini-CEX), direct observation of procedural skills (DOPS) and case based discussion, so far has shown no evidence for subsequent improvement in individual performance.[72] Certification of physicians and higher scores on licensing exams were shown to be predictive of better patient outcomes and better preventive care.[73-75] While years of experience tend to improve patient outcomes of residents and young graduates, research shows that quality of care tends to decrease the further away one gets from graduation.[76,77]

Research on training programs, systems and institutions and quality of care

The research on programs, systems and institutions showed that the clinical environment in which residents train has a lasting effect on the quality of care residents deliver after graduation. Research in two states in the United States demonstrated that obstetricians, who graduated from the top quintile medical schools, had systematically fewer complication rates as compared to the graduates of lowest quintile schools, independent of licensing scores.[78] Similarly, spending patterns of family and internal medicine physicians mimic those of the location of their training.[79] Internists' ability to practice conservatively has also been related to practice patterns of their training location.[80] This effect, however, is not stable for all outcomes and decreases over time, when researched in general surgery programs.[81] Program certification and accreditation could also be used as proxies for future care quality. Medical programs that have been ranked in U.S. News and World Report's "America's Best Hospitals" had lower rate of adverse actions and demonstrated a positive, albeit small, relationship with certification rates.[82] While there is no evidence that accreditation status of the school affect the quality of physician practice[83], graduates of accredited schools perform better on certification exams[84], which has been related to a number of quality indicators in practice.[73-75] Research has also revealed that effects of residency training persist over time. Those obstetricians with poor patient health outcomes after graduation had consistently poorer patient outcomes than their colleagues even 15 years after graduation.[85]

The problem of duality

As discussed in the seminal report *To Err is Human: Building a Safer Health System* published by the National Academy of Medicine (formerly known as the Institute of Medicine (IOM)), linear thinking, which has been followed by research described above, assumes that each part of the system only interacts with the part(s) immediately preceding/following it, and does not interact with any other elements of the system (Figure 1). Although in such linear systems the risk of errors is low, problems can arise when a part of a system serves multiple functions making a system more complex and, as a result, more prone to failure.[86] The duality that is engrained in PGME may represent an important source of such complexity.

All PGME training takes place in a workplace-based setting, usually in the hospital or a clinic, which forms a natural learning environment for residents (i.e. the clinical learning environment). Duality in PGME stems from several sources, including the clinical learning environment itself and the actors in the clinical learning environment (i.e. the patient, the learner and the supervisor(s)). Unlike clinical rotations in undergraduate medical education, residents are expected to take a sizeable responsibility for patient care. Aside from formal educational sessions that residents are expected to attend, clinical activities act as a pivotal point and important source of learning for residents.[87] On the other hand, clinical activities that do not directly add to resident learning, but are an essential part of patient care activities, are seen as "service".[88] As a result of the educational value of clinical activities, patients also have a dual role of being the *object* and the *subject* of learning. Clinical teaching and supervision is usually performed by attending physicians, who also carry the ultimate responsibility for the quality of care delivered to the patient. However, the clinical learning environment is not designed for learning. Its primary function of delivering high quality care in an efficient and cost-effective manner can create tensions between "production", time pressures, accommodating multiple learners' needs and varying levels of experience.[89-92] Ideally, the clinical environment where residents learn and work, should be perfectly situated to benefit both the patient and the learner, however, in reality practice dictates learning.

Main argument and research questions

In sum, current thinking and research on the relationship between PGME and quality of care suffers from two shortcomings. First, it assumes that PGME quality affects patient care through its effects on the resident. This view is likely too simplistic given the complexity of the learning environment in which residency training takes place. This linear thinking overlooks possible ways in which PGME quality can be related to patient care, and as a result can overlook its potential unintended consequences for patient care. Second, current thinking about the relationship between PGME and quality of care does not consider how the clinical learning environment itself may contribute to patient care and outcomes. As a result, the study of the relationship between PGME quality and quality of care should recognize the role of the clinical context as well as the complexity of the interplay between the parts.

In order to understand how the goals of PGME and high quality care can be better aligned, the initial step in unravelling this complexity is to investigate the nature of the relationship between PGME quality and quality of patient care in clinical learning environments. Therefore, the overarching research question of this thesis is:

1) How are indicators of PGME quality related to indicators of care quality and patient outcomes in clinical teaching departments?

However, in order to answer this question, it is necessary to ensure that the quality of the instruments used to measure quality are valid and reliable in the context of their use. Therefore, the supporting research question is:

2) What is the quality of the instruments used to measure quality of PGME and quality of care?

A pragmatic perspective

We believe that the breadth of the overarching research question and the complexity of the research concepts necessitates a pragmatic approach. We will use pragmatism as the overarching paradigm in approaching the research questions. Pragmatism, which was first traced out by Pierce in 1905, in its essence aims at clarifying meaning of concepts by identifying their "conceivable practical consequences of the affirmation or denial of the concept".[93] In other words, pragmatism posits that an idea is only meaningful to the extent that it has an effect on the set of things that are of interest.[94] Like scientific realism, pragmatism recognizes that there is an independent world outside, however, it also takes into account that we are historically, socially, and politically situated[95], on which our conceptions of what is meaningful or of interest are based. Unlike other research traditions that rely on antecedent descriptions, theories and explanations, pragmatism tends to focus on clarifying meaning and looking for consequences.[95] Pragmatism serves the overarching aim of this thesis by allowing us to look at and identify measures that are inherently relevant to both patients and residents in the clinical learning enviornment. In taking the pragmatic approach to studying the relationships between PGME quality and quality of care, this thesis applies PGME quality measures that are highly relevant to residents, namely the residency learning climate and clinical teaching quality, to clinical care measures that are inherently meaningful and relevant to patients.

Overview of studies in the thesis

Chapter 2 investigates the validity and reliability of an instrument to measure the residency learning climate in clinical teaching departments. The first association study in chapter 3 aims to test the association of the residency learning climate with patient outcomes by investigating the odds of adverse maternal and perinatal outcomes in 16 nonacademic obstetrics and gynecology teaching departments in the Netherlands. Chapter 4 follows up the investigation of quality from the perspective of the patient by bringing the validity and reliability of a frequently used patient questionnaire under the magnifying glass. Going deeper into testing the association between learning climate and care quality, chapter 5 investigates the role of the domains of the learning climate in relation to patient experiences of their hospitalization across 15 specialties in 86 clinical teaching departments. Chapter 6 investigates the hypothesis that good clinical teachers are also good clinicians in a retrospective observational study of anesthesiology faculty's teaching and clinical performance in an academic teaching hospital. In chapter 7 we weigh up the evidence in light of the existing literature, embed findings in theory, provide recommendations for practice and set out a plan for future research. Figure 2 provides an overview of the studies in the thesis and Table 2 provides their accompanying research questions and corresponding research methods. Since this thesis is based on published journal articles, some overlap will be inevitable.

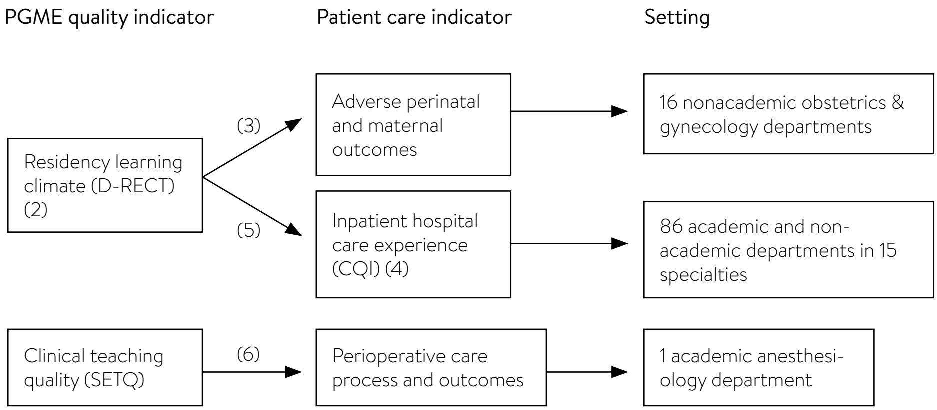


Figure 2. Overview of studies in the thesis including chapter numbers

Table 2. Studied research questions, data sources, corresponding research methods and analytical procedures, and outcomes

Chapter	Research question	Data sources	Analytic approach	Type of outcome
2	What are the psychometric properties of the Dutch Residency Educational Climate Test (D-RECT) at the resident and the department level and what number of residents' evaluations is needed per department to generate reliable assessments?	Resident surveys	Exploratory and confirmatory factor analyses, reliability and generalizability analyses	Learning climate
3	How is the residency learning climate associated with the odds of adverse perinatal and maternal outcomes in obstetrics and gynecology teaching departments?	Resident surveys, Netherlands Perinatal Registry data	Multilevel regression methods, sensitivity & bias analyses	Adverse perinatal and maternal outcomes
4	What are the psychometric properties of the Consumer Quality Index (CQI) Inpatient Hospital Care on patient and department levels and what number of patient evaluations is needed per department and per hospital to generate reliable assessments?	Patient surveys	Confirmatory factor analysis, reliability, generalizability analyses	Inpatient hospital care experience
5	How is the residency learning climate associated with patients' experience of their care during hospitalization and which facets of the learning climate affect patient care experience?	Resident surveys, patient surveys	Multilevel regression methods, False Discovery Rate, sensitivity analyses	Inpatient hospital care experience
6	What is the predictive value of anesthesiology faculty's clinical teaching performance and/or their role modeling status of clinical performance in perioperative care 6 months after their teaching evaluation and how does the presence of the resident influence this association?	Resident surveys, electronic health record data	Hierarchical panel analysis of longitudinal data using multilevel regression methods	Perioperative care process and outcome measures

References:

1. Iglehart JK. Institute of Medicine report on GME--a call for reform. *N Engl J Med*. 2015;372(4):376-381.
2. Meijerink MH, van Blerck-Woerdman A, Bosma H, et al. Numerus Fixus in Medicine: let go or keep(in Dutch). The Hague: Council for Public Health and Care; 2010.
3. Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med*. 2004;79(10):955-960.
4. Chen FM, Burstin H, Huntington J. The importance of clinical outcomes in medical education research. *Med Educ*. 2005;39(4):350-351.
5. The Registration Committee for Medical Specialists (RGS). Number of doctors in training per specialty/profile on the survey date of December 31st of the year (in Dutch). https://www.knmg.nl/opleiding-herregistratie-carriere/rgs/registers/aantal-registraties-specialistenaois.htm. Published March 8, 2017. Accessed November 7, 2017.
6. Borleffs J, Habets J, van Loon K, et al. From CanMEDS to CanBetter: How do you teach residents general competencies? (in Dutch). Utrecht: The Royal Dutch Medical Association (KNMG); 2015.
7. Harvey L, Green D. Defining quality. *Assess Eval High Educ*. 1993;18(1):9-34.
8. Tam M. Measuring quality and performance in higher education. *Quality in Higher Education*. 2001;7(1):47-54.
9. Crosby P. Quality is free: the art of making quality certain. New York: New American Library; 1979.
10. Parasuraman A, Zeithaml VA, Berry LL. A conceptual-model of service quality and its implications for future-research. *J Marketing*. 1985;49(4):41-50.
11. The Royal Dutch Medical Association (KNMG). Directive of the Central College of Medical Specialists (in Dutch). https://www.knmg.nl/opleiding-herregistratie-carriere/cgs/regelgeving/oude-regelgeving/oude-regelgeving-ccms.htm. Published November 11, 2015. Accessed November 7, 2017.
12. Postgraduate Medical Education: WFME global standards for quality improvement. Copenhagen: World Federation for Medical Education; 2003.
13. Flexner A. Medical education in the United States and Canada. From the Carnegie Foundation for the Advancement of Teaching, Bulletin Number Four, 1910. *Bull World Health Organ*. 2002;80(7):594-602.
14. Whitehead C. The Good doctor in medical education 1910-2010: a critical discourse analysis. Dissertation, Faculty of Pharmacy, University of Toronto. Toronto, Canada: University of Toronto; 2011. http://hdl.handle.net/1807/32161. Accessed April 21, 2017.
15. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Acad Med*. 2002;77(5):361-367.
16. Crossing the quality chasm: a new health system for the 21st Century. The National Academies Press; 2001.
17. Medical professionalism in the new millennium: a physician charter. *Ann Intern Med*. 2002;136(3):243-246.
18. Frenk J, Chen L, Bhutta ZA, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. 2010;376(9756):1923-1958.
19. Frank JR. The CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care. Ottawa: The Royal College of Physicians and Surgeons of Canada; 2005.
20. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach*. 2007;29(7):648-654.
21. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-645.
22. Lombarts MJ. Competence-based education misses the essence of the medical profession. *Perspect Med Educ*. 2015;4(6):326-328.
23. Dornan T, Mann K, Scherpbier A, Spencer J. Medical education: theory and practice. China: Elsevier; 2011.
24. Holmboe ES. Realizing the promise of competency-based medical education. *Acad Med*. 2015;90(4):411-413.
25. Morcke AM, Dornan T, Eika B. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv Health Sci Educ Theory Pract*. 2013;18(4):851-863.
26. Stimulus for internal quality improvement of medical specialist training programs (Scherpbier 2.0) (in Dutch). Utrecht: The Royal Dutch Medical Association (KNMG); 2015.
27. Lombarts MJ, Bucx MJ, Rupp I, Keijzers PJ, Kokke SI, Schlack W. An instrument for the assessment of the training qualities of clinician-educators (in Dutch). *Ned Tijdschr Geneeskd*. 2007;151(36):2004-2008.
28. Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, Scherpbier AJ. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med*. 2010;85(11):1732-1738.
29. Fluit C, Bolhuis S, Grol R, et al. Evaluation and feedback for effective clinical teaching in postgraduate medical education: validation of an assessment instrument incorporating the CanMEDS roles. *Med Teach*. 2012;34(11):893-901.
30. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011;33(10):820-827.
31. Slootweg IA, Lombarts KM, Boerebach BC, Heineman MJ, Scherpbier AJ, van der Vleuten CP. Development and validation of an instrument for measuring the quality of teamwork in teaching teams in postgraduate medical training (TeamQ). *PLoS One*. 2014;9(11):e112805.

32. Denison DR. What IS the difference between organizational culture and organizational climate? A native's point of view on a decade of paradigm wars. *Acad Manag Rev.* 1996;21(3):619-654.
33. Benzer JK, Young G, Stolzmann K, et al. The relationship between organizational climate and quality of chronic disease management. Health Services Research. 2011;46(3):691-711.
34. MacDavitt K, Chou SS, Stone PW. Organizational climate and health care outcomes. *Jt Comm J Qual Patient Saf.* 2007;33(11 Suppl):45-56.
35. Genn JM. AMEE Medical Education Guide No. 23 (Part 1): Curriculum, environment, climate, quality and change in medical education-a unifying perspective. *Med Teach.* 2001;23(4):337-344.
36. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. *Med Teach.* 2001;23(5):445-454.
37. Roff S, McAleer S. What is educational climate? *Med Teach.* 2001;23(4):333-334.
38. Boor K, Scheele F, van der Vleuten CP, Teunissen PW, den Breejen EM, Scherpbier AJ. How undergraduate clinical learning climates differ: a multi-method case study. *Med Educ.* 2008;42(10):1029-1036.
39. Strand P, Sjoborg K, Stalmeijer R, Wichmann-Hansen G, Jakobsson U, Edgren G. Development and psychometric evaluation of the Undergraduate Clinical Education Environment Measure (UCEEM). *Med Teach.* 2013;35(12):1014-1026.
40. Bennett D, Dornan T, Bergin C, Horgan M. Postgraduate training in Ireland: expectations and experience. *Ir J Med Sci.* 2014;183(4):611-620.
41. Piek J, Bossart M, Boor K, et al. The work place educational climate in gynecological oncology fellowships across Europe: the impact of accreditation. *Int J Gynecol Cancer.* 2015;25(1):180-190.
42. Schonrock-Adema J, Bouwkamp-Timmer T, van Hell EA, Cohen-Schotanus J. Key elements in assessing the educational environment: where is the theory? *Adv Health Sci Educ Theory Pract.* 2012;17(5):727-742.
43. Shimizu T, Tsugawa Y, Tanoue Y, et al. The hospital educational environment and performance of residents in the General Medicine In-Training Examination: a multicenter study in Japan. *Int J Gen Med.* 2013;6:637-640.
44. van Vendeloo SN, Brand PL, Verheyen CC. Burnout and quality of life among orthopaedic trainees in a modern educational programme: importance of the learning climate. *Bone Joint J.* 2014;96-B(8):1133-1138.
45. Cross V, Hicks C, Parle J, Field S. Perceptions of the learning environment in higher specialist training of doctors: implications for recruitment and retention. *Med Educ.* 2006;40(2):121-128.
46. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One.* 2014;9(1):e86512.
47. Ramani S, Leinster S. AMEE Guide no. 34: Teaching in the clinical environment. *Med Teach.* 2008;30(4):347-364.
48. Kilminster S, Cottrell D, Grant J, Jolly B. AMEE Guide No. 27: Effective educational and clinical supervision. *Med Teach.* 2007;29.
49. Lombarts KM, Heineman MJ, Arah OA. Good clinical teachers likely to be specialist role models: results from a multicenter cross-sectional survey. *PLoS One.* 2010;5(12):e15202.
50. Fluit CR, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med.* 2010;25(12):1337-1345.
51. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med.* 2004;19(9):971-977.
52. Snell L, Tallett S, Haist S, et al. A review of the evaluation of clinical teaching: new perspectives and challenges. *Med Educ.* 2000;34(10):862-870.
53. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad Med.* 2012;87(4):428-442.
54. van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med.* 2012;10:65.
55. Huber M, Knottnerus JA, Green L, et al. How should we define health? *Brit Med J.* 2011;343.
56. McPake B, Squires A, Mahat A, Araujo EC. The economics of health professional education and careers: insights from a literature review. Washington, DC: World Bank; 2015.
57. Berwick DM. Era 3 for medicine and health care. *JAMA.* 2016;315(13):1329-1330.
58. Porter ME. What is value in health care? *N Engl J Med.* 2010;363(26):2477-2481.
59. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q.* 1966;44(3):Suppl:166-206.
60. Appel F. From quality assurance to quality improvement: the Joint Commission and the new quality paradigm. *J Qual Assur.* 1991;13(5):26-29.
61. Batalden PB. Organizationwide quality improvement in health care. *Topics in health record management.* 1991;11(3):1-12.
62. Lazar EJ, Fleischut P, Regan BK. Quality measurement in healthcare. *Annu Rev Med.* 2013;64:485-496.
63. Tamblyn R. Outcomes in medical education: what is the standard and outcome of care delivered by our graduates? *Adv Health Sci Educ Theory Pract.* 1999;4(1):9-25.
64. Dauphinee WD. Educators must consider patient outcomes when assessing the impact of clinical training. *Med Educ.* 2012;46(1):13-20.
65. Asch DA, Epstein A, Nicholson S. Evaluating medical training programs by the quality of care delivered by their alumni. *JAMA.* 2007;298(9):1049-1051.

66. Kalet AL, Gillespie CC, Schwartz MD, et al. New measures to establish the evidence base for medical education: identifying educationally sensitive patient outcomes. *Acad Med*. 2010;85(5):844-851.
67. Iwashyna TJ, Fuld A, Asch DA, Bellini LM. The impact of residents, interns, and attendings on inpatient laboratory ordering patterns: a report from one university's hospitalist service. *Acad Med*. 2011;86(1):139-145.
68. Davis SS, Jr., Husain FA, Lin E, Nandipati KC, Perez S, Sweeney JF. Resident participation in index laparoscopic general surgical cases: impact of the learning environment on surgical outcomes. *J Am Coll Surg*. 2013;216(1):96-104.
69. Aiken CE, Aiken AR, Park H, Brockelsby JC, Prentice A. Factors associated with adverse clinical outcomes among obstetrics trainees. *Med Educ*. 2015;49(7):674-683.
70. McMahon GT, Katz JT, Thorndike ME, Levy BD, Loscalzo J. Evaluation of a redesign initiative in an internal-medicine residency. *N Engl J Med*. 2010;362(14):1304-1311.
71. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med*. 2014;371(19):1803-1812.
72. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;341:c5064.
73. Norcini JJ, Kimball HR, Lipner RS. Certification and specialization: do they matter in the outcome of acute myocardial infarction? *Acad Med*. 2000;75(12):1193-1198.
74. Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. *JAMA*. 2002;288(23):3019-3026.
75. Prystowsky JB, Bordage G, Feinglass JM. Patient outcomes for segmental colon resection according to surgeon's training, certification, and experience. *Surgery*. 2002;132(4):663-670; discussion 670-662.
76. Epstein AJ, Srinivas SK, Nicholson S, Herrin J, Asch DA. Association between physicians' experience after training and maternal obstetrical outcomes: cohort study. *BMJ*. 2013;346:f1596.
77. Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. Patients of doctors further from medical school graduation have poorer outcomes. *Med Educ*. 2017;51(5):480-489.
78. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA*. 2009;302(12):1277-1283.
79. Chen C, Petterson S, Phillips R, Bazemore A, Mullan F. Spending patterns in region of residency training and subsequent expenditures for care provided by practicing physicians for Medicare beneficiaries. *JAMA*. 2014;312(22):2385-2393.
80. Sirovich BE, Lipner RS, Johnston M, Holmboe ES. The association between residency training and internists' ability to practice conservatively. *JAMA Intern Med*. 2014;174(10):1640-1648.
81. Bansal N, Simmons KD, Epstein AJ, Morris JB, Kelz RR. Using patient outcomes to evaluate general surgery residency program performance. *JAMA Surg*. 2016;151(2):111-119.
82. Philibert I. Can hospital rankings measure clinical and educational quality? *Acad Med*. 2009;84(2):177-184.
83. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094-1102.
84. Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. *Arch Pediatr Adolesc Med*. 2000;154(3):256-260.
85. Asch DA, Nicholson S, Srinivas SK, Herrin J, Epstein AJ. How do you deliver a good obstetrician? Outcome-based evaluation of medical education. *Acad Med*. 2014;89(1):24-26.
86. Kohn LT, Corrigan JM, Donaldson MS. To err is human: building a safer health system. Washington (DC): National Academies Press; 2000.
87. Teunissen PW, Scheele F, Scherpbier AJ, et al. How residents learn: qualitative evidence for the pivotal role of clinical activities. *Med Educ*. 2007;41(8):763-770.
88. Kesselheim JC, Cassel CK. Service: an essential component of graduate medical education. *N Engl J Med*. 2013;368(6):500-501.
89. Jippes E, Van Luijk SJ, Pols J, Achterkamp MC, Brand PL, Van Engelen JM. Facilitators and barriers to a nationwide implementation of competency-based postgraduate medical curricula: a qualitative study. *Med Teach*. 2012;34(8):e589-602.
90. van Rossum TR, Scheele F, Scherpbier AJ, Sluiter HE, Heyligers IC. Dealing with the complex dynamics of teaching hospitals. *BMC Med Educ*. 2016;16:104.
91. Stroud L, Bryden P, Kurabi B, Ginsburg S. Putting performance in context: the perceived influence of environmental factors on work-based performance. *Perspect Med Educ*. 2015;4(5):233-243.
92. Famiglio LM, Thompson MA, Kupas DF. Considering the clinical context of medical education. *Acad Med*. 2013;88(9):1202-1205.
93. Pierce C. What pragmatism is. *The Monist*. 1905;15:161-181.
94. Campbell PL. Peirce, pragmatism, and the right way of thinking. Albuquerque, New Mexico: Sandia National Laboratories; 2011.
95. Cherryholmes CH. Notes on pragmatism and scientific realism. *Educ Res*. 1992;21(6):13-17.

CHAPTER 2

Revisiting the D-RECT tool: validation of an instrument measuring residents' learning climate perceptions

*A. Smirnova**
*M.E.W.M. Silkens**
R.E. Stalmeijer
O.A. Arah
A.J.J.A. Scherpbier
C.P.M. van der Vleuten
K.M.J.M.H. Lombarts
**These authors should be considered as first authors in this study*

Medical Teacher, 2016; 38: 476-481

Abstract

Introduction

Credible evaluation of the learning climate requires valid and reliable instruments in order to inform quality improvement activities. Since its initial validation the Dutch Residency Educational Climate Test (D-RECT) has been increasingly used to evaluate the learning climate, yet it has not been tested in its final form and on the actual level of use – the department.

Aim

Our aim was to re-investigate the internal validity and reliability of the D-RECT at the resident and department levels.

Methods

D-RECT evaluations collected during 2012–2013 were included. Internal validity was assessed using exploratory and confirmatory factor analyses. Reliability was assessed using generalizability theory.

Results

In total, 2306 evaluations and 291 departments were included. Exploratory factor analysis showed a 9-factor structure containing 35 items: teamwork, role of specialty tutor, coaching and assessment, formal education, resident peer collaboration, work is adapted to residents' competence, patient sign-out, educational atmosphere, and accessibility of supervisors. Confirmatory factor analysis indicated acceptable to good fit. Three resident evaluations were needed to assess the overall learning climate reliably and eight residents to assess the subscales.

Conclusion

This study reaffirms the reliability and internal validity of the D-RECT in measuring residency training learning climate. Ongoing evaluation of the instrument remains important.

Introduction

The learning climate in residency has been increasingly recognized as an important contributor to high-quality graduate medical education (GME).[1-3] The learning climate can be conceptualized as residents' perceptions of the formal and informal aspects of education[4], including perceptions of the overall atmosphere[5] as well as policies, practices, and procedures within the teaching hospital[6]. When residents perceive their learning climate as positive, they are more likely to make more use of their existing knowledge base[7] and effective learning styles.[8] Furthermore, a positive learning climate is instrumental in preventing resident burnout[9] and can among others promote career satisfaction and professional identity development in residents.[10]

Due to the significance of the learning climate for resident education, a key role has been assigned to evaluation and improvement of the learning climate.[1] In order to accurately evaluate the learning climate, literature has stressed the importance of enhancing the validity of instruments used in measuring the learning climate.[11,12] One instrument that has been increasingly used to measure the learning climate is the Dutch Resident Educational Climate Test (D-RECT). The preliminary D-RECT was validated in 2011[13], which lead to the final 50-item questionnaire. Pinnock et al. have investigated the internal consistency and applicability of the D-RECT in an Australian teaching hospital[14], however, no extensive validation of the final 50-item D-RECT has been performed so far. Furthermore, to our knowledge, the concepts of the D-RECT have not yet been tested in their aggregated form. Literature on climate research regards that an organization's climate constitutes individual responses aggregated to the unit of analysis.[15] Similarly, for the D-RECT, this means that individual responses should be aggregated to the level of the department in order to inform a department's learning climate. As a result, the identified structure of the D-RECT may not be the optimal structure for evaluating the learning climate of the department.

In light of the increasing importance allotted to the learning climate as well as the need for further validity evidence of learning climate instruments, a re-evaluation of the initial psychometric properties of the D-RECT is necessary. Since its introduction, D-RECT has been widely used in post-graduate training in the Netherlands as a part of internal quality improvement efforts, which provides an opportunity for a larger-scale analysis of existing data. Our study investigated the internal validity of the D-RECT as well as its internal consistency and generalizability on the resident and the department levels.

Methods

Setting

The study included departments that provide hospital-based residency training in the Netherlands. During a pre-determined period (usually one month once per year), the residents rotating on the service are invited to fill out the D-RECT. D-RECT measurements are usually repeated every year in order to gain insight into and monitor the strengths and weaknesses of a department's learning climate and fuel quality improvement initiatives.

Data collection

For this study, we included resident evaluations of the departments' learning climate using D-RECT between January 2012 and December 2013. If a department was evaluated more than once during the study period, only the most recent evaluation was included. D-RECT evaluations were completed via an online platform or a paper questionnaire depending on the training program. In case of online evaluations, participants were reminded up to three times by e-mail to participate in the online D-RECT evaluations. For paper-based evaluations, no reminders were provided. Participation in the D-RECT evaluations was anonymous and voluntary for all participants. The institutional ethical review board of the Academic Medical Centre of the University of Amsterdam provided a waiver of ethical approval for this study. Written permission was asked and granted from the departments using the web-based platform and from hospitals using paper questionnaires.

The D-RECT questionnaire

The D-RECT was developed based on qualitative research, expert opinion and a Delphi panel by Boor et al.[13] The initial questionnaire consisted of 75 items rated on a 5-point Likert scale (1 = totally disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = totally agree). An additional "not applicable" option was also included. Exploratory factor analysis with oblimin rotation based on 1276 resident evaluations revealed 50 items covering 11 constructs describing the learning climate, which were confirmed by a confirmatory factor analysis. The subscale reliability (Cronbach's α) ranged from 0.64 to 0.85. Generalizability analysis showed that 11 resident evaluations were needed to reliably evaluate all subscales and three for the overall score.

Data analysis

To describe relevant characteristics of the study sample and the D-RECT, descriptive statistics and frequencies were used. Evaluations with more than 50% of the items missing were excluded from the analysis. For evaluations with less than 50% missing, data were assumed to be missing at random and imputed using the expectation–maximization (EM) technique. To be able to differentiate between resident and department level evaluations, departments with only one resident evaluation were excluded from the analysis.

In developing the new structure, first, practical importance of the items was discussed within the research group. Items with a lack of practical importance were excluded from the analysis. One-third of the sample containing resident evaluations was randomly selected for exploratory factor analysis (EFA), which was deemed a sufficient sample.[16] Principal axis factoring was chosen rather than principal component analysis (PCA), because PCA tends to inflate factor loadings and limits the possibility for confirmation by CFA and replication in other samples.[16] Similar to Boor et al., we chose oblique rotation, since the data showed that factors were correlated. Based on the pattern matrix of the first factor solution, items with factor loadings <0.4 were excluded from further analysis.[17] A second EFA was performed on the remaining items. The Kaiser–Guttman criterion (eigenvalue >1.0) was used to determine the number of factors. The soundness of this model was compared with alternate models based on the scree plot and the amount of explained variance as suggested previously.[18] Items were assigned to the factor on which the factor loading was highest. The placement of each item was further discussed within the research group.

To assess the fit of the structure obtained by the EFA, a confirmatory factor analysis (CFA) was performed on the remaining two-thirds of the sample containing resident evaluations. In order to compare the fit of the new structure to the original, the CFA was then repeated on the same sample using the original 11-factor structure as identified by Boor et al. The structure with the best fit was tested on the department level using a CFA. Department level data were attained by calculating the means for each question per department.

The CFA models were estimated by using robust maximum likelihood. The fit of the model was assessed by using the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). Cut off values for these fit-indices were pre-determined (SRMR <0.08 for good fit and <0.12 for acceptable fit; RMSEA <0.06 for good fit and <0.10 for acceptable fit; CFI and TLI >0.95 for good fit and >0.90 for acceptable fit).[19]

The complete sample of resident evaluations as well as the aggregated sample was used to determine internal consistency, inter-scale correlations, and the item-total correlations. Internal consistency of the subscales was checked by calculating Cronbach's α for each subscale. Cronbach's α >0.70 was considered satisfactory.[20] Inter-scale correlations were deemed satisfactory when <0.70. The homogeneity of each scale was assessed by item-total correlations, which should be >0.40.[21]

Generalizability analysis was conducted using generalizability theory to determine the optimal number of resident evaluations needed for reliable estimation of the subscale and total scores. We regarded the departments to be the unit of analysis and the number of items as fixed. The resulting design was an unbalanced single-facet nested study with persons (*p*) nested within departments (*d*) (*p:d*).[22] We estimated variance components associated with variance across departments (*Sd*) and persons nested within departments (*Sp:d*), the reproducibility coefficient (*G*), and standard error of measurement (SEM) for varying number of trainees for the mean score and the subscale scores. Similar to Boor et al., we used SEM <0.26 ($1.96 \times 0.26 \times 2 \approx 1.0$) representing a 1 unit "noise level" on the scale 1–5 as the maximum value for 95% confidence interval interpretation.

The CFA was performed using the lavaan package in R statistical software version 3.1.0. Variance components were estimated using UrGENOVA. The remaining analyses were performed with SPSS version 20 (IBM Corp., Armonk, NY).

Results

Study participants

Between 2012 and 2013, a total of 2347 D-RECT evaluations were completed, of which nine were excluded based on more than 50% missing data. Thirty-two departments were excluded because they had only one D-RECT evaluation. As a result, 2306 resident evaluations for 291 departments in 48 teaching hospitals including five academic teaching hospitals were included in the study. Seventy-two percent of the residents completed online evaluations, yielding a response rate of 62%. The response rate for the paper-based sample (28%) could not be calculated because the number of invited trainees was not collected. However, based on the literature, we suspect a similar or even higher response rate for the paper-based questionnaires.[23] A detailed description of the study population is provided in Table 1.

Psychometric properties of D-RECT

Two items ("Observation forms are used to structure my feedback" and "Observation forms are used periodically to monitor my progress") were deemed redundant, because observation forms are now used to evaluate all residents and were therefore removed (Directive of the Central College of Medical Specialists 2009). Seven hundred sixty nine (769) evaluations were randomly assigned for the EFA and 1537 for the CFA. The first EFA resulted in 10 factors. Thirteen items were removed based on their factor loading, leaving 35 items for further analysis (Appendix, Table A1).

The final EFA resulted in a 9-factor structure, which explained 65.5% of the total variance. Overall, items clustered into the following subscales: (1) educational atmosphere, (2) teamwork, (3) role of specialty tutor, (4) coaching and assessment, (5) formal education, (6) resident peer collaboration, (7) work is adapted to residents' competence, (8) accessibility of supervisors, and (9) patient sign-out (Appendix, Table A2). Table 2 shows the results of the CFA performed on the resident and department levels as well as the results of the original 11-factor structure.

Table 1. Characteristics of the study population

Characteristics	N	(%)
Number of teaching hospitals	48	(100)
Academic	5	(10)
Nonacademic	43	(90)
Number of departments	291	(100)
Surgical[a]	97	(33)
Medical[b]	165	(57)
Auxilliary[c]	29	(10)
Number of resident evaluations	2306	(100)
Gender of respondents		
Male	899	(39)
Female	1366	(59)
Missing	41	(2)
Year of training		
1	287	(13)
2	309	(13)
3	259	(11)
4	209	(9)
5	156	(7)
6	94	(4)
Doctor not in training	292	(13)
Fellow	46	(2)
Missing	654	(28)

[a]Surgical: general surgery, obstetrics and gynecology, orthopedic surgery, urology, otolaryngology, ophthalmology, plastic surgery, oral and maxillofacial surgery, thoracic surgery

[b]Medical: internal medicine, pediatrics, neurology, cardiology, emergency medicine, psychiatry, rehabilitation medicine, gastroenterology, pulmonology, anesthesiology, radiology, dermatology, intensive care, geriatrics, pathology, sports medicine, nuclear medicine, medical microbiology, radiotherapy

[c]Auxiliary: clinical chemistry, clinical genetics, clinical physics, medical psychology, pharmacy

Table 2. Fit indices of the 9-factor structure at the resident and department level compared to the original 11-factor structure at resident level

	Resident level (n=1537)		**Department level (n=291)**
	35-item (9-factor) questionnaire	50-item (11-factor) questionnaire	35-item (9-factor) questionnaire
CFI	0.92	0.86	0.89
TLI	0.91	0.85	0.88
SRMR	0.04	0.05	0.06
RMSEA	0.04	0.05	0.06

Cronbach's α for subscales ranged from 0.71 to 0.86 at the resident level and from 0.80 to 0.91 for the department level (Appendix, Table A2). Corrected item-total correlations ranged from 0.41 to 0.75 at the resident level, and 0.53 to 0.84 at the department level (Appendix, Table A2). Inter-scale correlations ranged from 0.32 to 0.52 for the resident level, and from 0.37 to 0.66 at the department level. Summary statistics for the D-RECT subscales are provided in the appendix (Appendix, Table A3).

Table A2 reports the minimum number of trainees needed to reliably assess each subscale and the mean score. The minimum number of resident evaluations needed ranged from three for the mean score to eight for patient sign-out and resident peer collaboration subscales.

Discussion

Main findings

The aim of this study was to test the internal validity and the reliability of the D-RECT on both the resident and the department level. The results showed that the learning climate can be evaluated on both the level of the resident and the department using 35 questions grouped into nine subscales: educational atmosphere, teamwork, role of specialty tutor, coaching and assessment, formal education, resident peer collaboration, work is adapted to residents' competence, accessibility of supervisors, and patient sign-out. Furthermore, eight residents per department were needed to evaluate all subscales of the learning climate reliably.

Explanation of findings

Overall the new structure reflects the original D-RECT questionnaire. Although the subscale "feedback" was dropped, the topic is still represented in the questionnaire by the subscale "coaching and assessment". Items from the subscale "attendings' role" were divided into the new scales "educational atmosphere" and "accessibility of supervisors".

The literature also regards clinical teachers as being primarily responsible for creating an atmosphere, in which learners can comfortably identify and address their limitations[24], which is represented in the questionnaire by the factor "educational atmosphere". The subscale "supervision" has been incorporated in the new subscale "accessibility of supervisors". Similarly, accessibility of timely and appropriate supervision has been regarded to be an important factor in both educational and patient outcomes.[25] In conclusion, we believe that the two new subscales are theoretically representative of the contributing factors to the learning climate in residency.

The new 9-factor structure was supported by the CFA. The SRMR and the RMSEA, indicating how well the a priori model reproduces the sample data, showed a good fit. In contrast, the CFI and TLI, measuring the improvement of our model when compared to a restricted model, showed a slightly lower albeit acceptable performance, especially at the department level. Nevertheless, rather than evaluating single fit indices to accept or reject a model, it has been recommended to consider the fit indices in a combined manner instead.[26] As such, the overall good fit of the new 9-factor model on the resident level has been demonstrated by the acceptable values of the incremental fit indices (CFI and TLI) and the good fit of the absolute fit indices (SRMR and RMSEA). When compared with the fit of the original 11-factor structure in our sample, the 9-factor model showed an improvement in the CFI and TLI indices. At the department level, the incremental fit indices were slightly below the acceptable cutoff points, while the absolute indices indicate good fit. It can be concluded that the 9-factor model showed an acceptable fit at the department level. The applicability of the 9-factor structure at the resident and department levels was further supported by the item-total correlations >0.30 that indicated that each item contributed to the measurement of the concept learning climate, and inter-scale correlations <0.70 that indicated that D-RECT comprised nine sub-constructs.

With regard to the reliability, internal consistency (Cronbach's α) was satisfactory at the resident as well as the department level. Generalizability analysis, where residents were nested within departments, showed that a minimum of eight resident evaluations was required for the 9-factor questionnaire, whereas previously 11 resident evaluations were needed. The overall climate could still be reliably evaluated with only three residents. Between-resident differences accounted for two to three times more variation in scores than between-department differences.

In choosing our statistical approach, we aimed to retain a number of factors that would maximize the instrument's explanatory power. We considered this especially important since those that use D-RECT have appreciated its multidimensionality when used in quality improvement activities.[14] In line with this reasoning, the two-item scale (patient sign-out) was retained to support patient safety and residency training.[27]

With regard to the content of the subscales, the constructs of the D-RECT fit within climate frameworks.[28] Ostroff has researched the broad concept of climate and have organized climate perceptions into three higher order facets: the affective, the cognitive and the instrumental facet.[29] In the D-RECT, the affective facet is accounted for by the overall feeling of the atmosphere (constructive educational atmosphere), how well residents work together (resident peer collaboration) and how well the inter-professional team works together (teamwork). The cognitive facet is accounted for by the focus on how the supervisor helps the resident reflect on performance (coaching and assessment), to what extent the resident is involved in the patient hand-over (patient sign-out) and to what extent the work of the resident is adapted to the level of experience of the resident (work is adapted to residents' competence). The D-RECT takes the instrumental facet into account by evaluating planned education (formal education), what the involvement of the formal educator is (role of the specialty tutor) and to what extent the supervisors are involved (accessibility of supervision). Furthermore, the D-RECT has a more applied grounding. Since the D-RECT is an instrument to measure the learning climate in work-based GME, it is intuitive that the instrument does not only focus on educational activities that are linked to GME, but also to the patient-care related aspects.

Strengths and limitations of the study

In addition to contributing to the validity evidence of the D-RECT, this study adds to the literature by using the unit of analysis – the department. It would have been preferable to perform an EFA on the department level to identify the best fitting structure separately for this level. However, the aggregated dataset with only 291 departments was rather small to perform both an EFA and a CFA.[16] Since the number of resident evaluations was much higher compared to the number of departments in our sample, we, therefore, chose an alternative route by first exploring the structure of the D-RECT on the resident level and then confirming it on the department level. A larger sample is more likely to generate factor loadings that closely reflect the population and are therefore less variable in repeated testing.[30] In our study, the EFA is based on a sample exceeding the 10:1 ratio, thereby contributing to the stability of the findings.[16] The number of departments was considered to be sufficient for a CFA, especially since every department score is composited from multiple resident evaluations.

Furthermore, the multicenter setting, with both academic and nonacademic hospitals covering various specialties across the Netherlands, affirms the soundness of the results in the Netherlands. However, since learning climate reflects the perceptions of the residents, cultural differences in residents' expectations may occur.[31] As a result, residents from other cultures may emphasize aspects of training that are not covered by the D-RECT.

Implications for practice and future research

For those who use the D-RECT we hope to bring trust in the structure of the instrument, whilst providing slight nuances in the arrangement of subscales. A significant improvement for practice is the rigorous reduction in the number of items, which may improve response rate and truthfulness of residents' responses.[12] Another improvement is that D-RECT requires eight instead of 11 residents for evaluation of all scales. This may encourage smaller departments to make use of the D-RECT in the future. In departments with fewer than eight but at least three residents, the overall learning climate can still be reliably evaluated.

While this study furthers the evidence of D-RECT's internal structure, future research could continue to provide validity evidence, including the response process and relationships with other variables. Specifically, attention should be paid to the formulation of the items, including use of double negatives. Reformulation of such items might improve the responses of the residents. Finally, as already mentioned by Boor et al., investigating the effects of the D-RECT on practice would be a useful step forward in contributing to the discussion on how to improve the learning climate as part of quality assurance programs.[13]

Conclusion

In conclusion, after analyzing the reliability and internal validity of the D-RECT we propose an updated structure, which holds on both resident and department levels and can be used to evaluate the learning climate of departments provided a minimum number of residents per department. With 35 items divided into nine subscales, the instrument is now shorter and may therefore be more suitable for practice and research. However, ongoing evaluation of the applicability of the D-RECT items to practice is needed.

Practice points

- We propose a new structure of D-RECT that can reliably evaluate the learning climate at the intended level of use using a shorter questionnaire and fewer residents.
- The D-RECT provides feedback on affective, cognitive and instrumental aspects of the learning climate, covering both educational and patient care related facets of training.
- Evaluation of the learning climate using the detailed feedback provided by D-RECT can serve as a starting point for quality improvement initiatives in training programs.

Glossary

Educational Climate: the way in which students perceive their educational environment and teaching practices.[5]

References:

1. Postgraduate Medical Education: WFME global standards for quality improvement. Copenhagen: World Federation for Medical Education; 2003.
2. Weiss KB, Wagner R, Nasca TJ. Development, testing, and implementation of the ACGME Clinical Learning Environment Review (CLER) program. *J Grad Med Educ*. 2012;4(3):396-398.
3. Weiss KB, Bagian JP, Nasca TJ. The clinical learning environment: the foundation of graduate medical education. *JAMA*. 2013;309(16):1687-1688.
4. Roff S, McAleer S. What is educational climate? *Med Teach*. 2001;23(4):333-334.
5. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. *Med Teach*. 2001;23(5):445-454.
6. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One*. 2014;9(1):e86512.
7. Shimizu T, Tsugawa Y, Tanoue Y, et al. The hospital educational environment and performance of residents in the General Medicine In-Training Examination: a multicenter study in Japan. *Int J Gen Med*. 2013;6:637-640.
8. Delva MD, Kirby J, Schultz K, Godwin M. Assessing the relationship of learning approaches to workplace climate in clerkship and residency. *Acad Med*. 2004;79(11):1120-1126.
9. Llera J, Durante E. Correlation between the educational environment and burn-out syndrome in residency programs at a university hospital. *Arch Argent Pediatr*. 2014;112(1):6-11.
10. Cross V, Hicks C, Parle J, Field S. Perceptions of the learning environment in higher specialist training of doctors: implications for recruitment and retention. *Med Educ*. 2006;40(2):121-128.
11. Soemantri D, Herrera C, Riquelme A. Measuring the educational environment in health professions studies: a systematic review. *Med Teach*. 2010;32(12):947-952.
12. Colbert-Getz JM, Kim S, Goode VH, Shochet RB, Wright SM. Assessing medical students' and residents' perceptions of the learning environment: exploring validity evidence for the interpretation of scores from existing tools. *Acad Med*. 2014;89(12):1687-1693.
13. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011;33(10):820-827.
14. Pinnock R, Welch P, Taylor-Evans H, Quirk F. Using the DRECT to assess the intern learning environment in Australia. *Med Teach*. 2013;35(8):699.
15. Schneider B, Ehrhart MG, Macey WH. Perspectives on organizational climate and culture. APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization. Washington, DC: American Psychological Association; 2011:373-414.
16. Wetzel AP. Factor analysis methods and validity evidence: a review of instrument development across the medical education continuum. *Acad Med*. 2012;87(8):1060-1069.
17. Hatcher L, Stepanski EJ. A step-by-step approach to using the SAS system for univariate and multivariate statistics. Carry, NC: SAS Institute; 1994.
18. Schonrock-Adema J, Heijne-Penninga M, Van Hell EA, Cohen-Schotanus J. Necessary steps in factor analysis: enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Med Teach*. 2009;31(6):e226-232.
19. Brown TA. Confirmatory factor analysis for applied research. New York: Guilford; 2006.
20. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297-334.
21. Arah OA, Hoekstra JB, Bos AP, Lombarts KM. New tools for systematic evaluation of teaching qualities of medical faculty: results of an ongoing multi-center survey. *PLoS One*. 2011;6(10):e25983.
22. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34(11):960-992.
23. Yarger JB, James TA, Ashikaga T, et al. Characteristics in response rates for surveys administered to surgery residents. *Surgery*. 2013;154(1):38-45.
24. Ramani S. Patient-centered care or patient data-centered care: a tale of 2 admissions. *J Grad Med Educ*. 2013;5(3):362-364.

25. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad Med*. 2012;87(4):428-442.
26. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1-55.
27. Myers JS, Bellini LM. Resident handoffs: appreciating them as a critical competency. *J Gen Intern Med*. 2012;27(3):270-272.
28. Schonrock-Adema J, Bouwkamp-Timmer T, van Hell EA, Cohen-Schotanus J. Key elements in assessing the educational environment: where is the theory? *Adv Health Sci Educ Theory Pract*. 2012;17(5):727-742.
29. Ostroff C. The effects of climate and personal influences on individual behavior and attitudes in organizations. *Organ Behav Hum Decis Process*. 1993;.56(1):pp.
30. MacCallum RC, Widaman KF, Zhang SB, Hong SH. Sample size in factor analysis. *Psychol Methods*. 1999;4(1):84-99.
31. Wong AK. Culture in medical education: comparing a Thai and a Canadian residency programme. *Med Educ*. 2011;45(12):1209-1219.

Appendix

Table A1. Items removed during the exploratory factor analysis (EFA) due to factor loading <0.4

Original subscale	Code	Removed item	Cause for removal
Supervision	DS3	The guidelines clearly outline when to request input from a supervisor.	Content of the questions is unclear. May be difficult for the resident to answer.
	DS4	The amount of supervision I receive is appropriate for my level of experience.	Difficult question to assess. Item not in line with other items in the scale Supervision.
Coaching and assessment	DBT1	I am asked on a regular basis to provide a rationale for my management decisions and actions.	Content of the question is unclear.
	DBT2	My attendings coach me on how to communicate with difficult patients.	Unclear what is meant with “difficult patients”.
	DBT3	My attendings take the initiative to explain their actions.	Item not in line with other items in the scale Coaching and assessment.
Teamwork	DWT4	Teamwork is an integral part of my training.	Difficult question to assess. Content of the question is unclear.
Work is adapted to residents’ competence	DAA4	There is enough time in the schedule for me to learn new skills.	Information is covered in other items.
Attendings’ role	DAS1	My attendings take time to explain things when asked for advice.	Conditional question.
	DAS2	My attendings are happy to discuss patient care.	Formulation of “happy” is unclear for the purpose of this question.
	DAS4	My attendings treat me as an individual.	Content of the question is unclear.
	DAS6	My attendings are all in their own way positive role models.	The use of “all” makes it a complex question to answer. The question was also flagged by respondents in written comments as a difficult question to assess.
Patient sign out	DO1	When there is criticism of a management plan I have developed in consultation with my attending physician, I know the attending physician will back me up.	Conditional question.
	DO2	Sign out takes place in a safe climate.	Item not in line with other items in the scale Patient sign out.

Table A2: Item and scale characteristics, reliability coefficients and item-total correlations at resident and department level

Scale (N respondents needed*)	Code	Item	Corrected item-total correlation resident (department)	Cronbach's α resident (department)
Educational atmosphere (6)	DSO1	Continuity of care is not affected by differences of opinion between attendings.	0.67 (0.81)	0.83 (0.91)
	DSO2	Differences of opinion between attendings about patient management are discussed in such a manner that is instructive to others present.	0.62 (0.77)	
	DSO3	Differences of opinion are not such that they have a negative impact on the work climate.	0.71 (0.84)	
	DAS3	There is (are) NO attending physician(s) who have a negative impact on the educational climate.	0.65 (0.83)	
	DAS5	My attendings treat me with respect.	0.54 (0.72)	
Teamwork (6)	DWT1	Attendings, nursing staff, other allied health professionals and residents work together as a team.	0.58 (0.65)	0.79 (0.84)
	DWT2	Nursing staff and other allied health professionals make a positive contribution to my training.	0.70 (0.77)	
	DWT3	Nursing staff and other allied health professionals are willing to reflect with me on the delivery of patient care.	0.64 (0.76)	
Role of specialty tutor (5)	DFO1	The specialty tutor monitors the progress of my training.	0.59 (0.69)	0.86 (0.91)
	DFO2	The specialty tutor provides guidance to other attendings when needed.	0.67 (0.80)	
	DFO3	The specialty tutor is actively involved in improving the quality of education and training.	0.68 (0.78)	
	DFO4	In this rotation evaluations are useful discussions about my performance.	0.75 (0.83)	
	DFO5	My plans for the future are part of the discussion.	0.64 (0.72)	
	DFO6	During evaluations, input from several attendings is considered.	0.58 (0.66)	
Coaching and assessment (6)	DBT4	My attendings take the initiative to evaluate my performance.	0.63 (0.75)	0.82 (0.88)
	DBT5	My attendings take the initiative to evaluate difficult situations I have been involved in.	0.65 (0.79)	
	DBT6	My attendings evaluate whether my performance in patient care is commensurate with my level of training.	0.60 (0.72)	
	DBT7	My attendings occasionally observe me taking a history.	0.47 (0.53)	
	DBT8	My attendings assess not only my medical expertise but also other skills such as teamwork, organization or professional behavior.	0.63 (0.75)	
	DFB3	My attendings give regular feedback on my strengths and weaknesses	0.58 (0.71)	

Scale (N respondents needed*)	Code	Item	Corrected item-total correlation resident (department)	Cronbach's α resident (department)
Formal education (6)	DGO1	Residents are generally able to attend scheduled educational activities.	0.50 (0.61)	0.79 (0.86)
	DGO2	Educational activities take place as scheduled.	0.64 (0.72)	
	DGO3	Attendings contribute actively to the delivery of high-quality formal education.	0.63 (0.74)	
	DGO4	Formal education and training activities are appropriate to my needs.	0.64 (0.75)	
Resident peer collaboration (8)	DSP1	Residents work well together.	0.73 (0.83)	0.84 (0.91)
	DSP2	Residents, as a group, make sure the day's work gets done.	0.73 (0.83)	
	DSP3	Within our group of residents it is easy to find someone to cover or exchange a call.	0.66 (0.79)	
Work is adapted to residents' competence (5)	DAA1	The work I am doing is commensurate with my level of experience.	0.60 (0.72)	0.71 (0.80)
	DAA2	The work I am doing suits my learning objectives at this stage of my training.	0.61 (0.70)	
	DAA3	It is possible to do follow up with patients.	0.41 (0.56)	
Accessibility of supervisors (5)	DAS7	When I need a attending, I can always contact one.	0.61 (0.74)	0.71 (0.81)
	DAS8	When I need to consult an attending, they are readily available.	0.56 (0.68)	
	DS5	It is clear which attending supervises me.	0.44 (0.58)	
Patient sign-out (8)	DO3	Sign-out is used as a teaching opportunity.	0.64 (0.81)	0.78 (0.89)
	DO4	Attendings encourage residents to join in the discussion during sign-out.	0.64 (0.81)	

* Based on results of generalizability analysis.

Table A3: Mean (SD), median, minimum and maximum of the new D-RECT subscales and mean score

	Educational atmosphere	Teamwork	Role of specialty tutor	Coaching and assessment	Formal education	Resident peer collaboration	Work is adapted to residents' competence	Accessibility of supervisors	Patient sign-out	Mean score
Mean (SD)	3.86 (0.71)	3.92 (0.69)	4.04 (0.62)	3.33 (0.68)	3.84 (0.68)	4.27 (0.67)	4.06 (0.61)	4.30 (0.59)	3.82 (0.81)	3.89 (0.46)
Median	4.00	4.00	4.00	3.33	4.00	4.33	4.00	4.33	3.89	3.91
Min	1.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.74
Max	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00

CHAPTER 3

The association between learning climate and adverse obstetrical outcomes in 16 non-tertiary obstetrics and gynecology departments in the Netherlands

A. Smirnova
A.C.J. Ravelli
R.E. Stalmeijer
O.A. Arah
M.J. Heineman
C.P.M. van der Vleuten
J.A.M van der Post
K.M.J.M.H. Lombarts

Academic Medicine, 2017 Sep 26; Published Ahead of Print

Abstract

Purpose
To investigate the association between learning climate and adverse perinatal and maternal outcomes in obstetrics– gynecology departments.

Method
The authors analyzed 23629 births and 103 learning climate evaluations from 16 nontertiary obstetrics–gynecology departments in the Netherlands in 2013. Multilevel logistic regressions were used to calculate the odds of adverse perinatal and maternal outcomes, by learning climate score tertile, adjusting for maternal and department characteristics. Adverse perinatal outcomes included fetal or early neonatal mortality, five-minute Apgar score <7, or neonatal intensive care unit admission for ≥24 hours. Adverse maternal outcomes included postpartum hemorrhage and/or transfusion, death, uterine rupture, or third- or fourth-degree perineal laceration. Bias analyses were conducted to quantify the sensitivity of the results to uncontrolled confounding and selection bias.

Results
Learning climate scores were significantly associated with increased odds of adverse perinatal outcomes (aOR 2.06, 95% CI 1.14–3.72). Compared with the lowest tertile, departments in the middle tertile had 46% greater odds of adverse perinatal outcomes (aOR 1.46, 95% CI 1.09–1.94); departments in the highest tertile had 69% greater odds (aOR 1.69, 95% CI 1.24–2.30). Learning climate was not associated with adverse maternal outcomes (middle vs. lowest tertile: OR 1.04, 95% CI 0.93–1.16; highest vs. lowest tertile: OR 0.98, 95% CI 0.88–1.10).

Conclusions
Learning climate was associated with significantly increased odds of adverse perinatal, but not maternal, outcomes. Research in similar clinical contexts is needed to replicate these findings and explore potential mechanisms behind these associations.

Introduction

The primary aim of residency is to shape high-performing physicians in a clinical learning environment that ensures safe and effective care for patients and high-quality learning, safety, and well-being for residents.[1] Asch and colleagues[2] found a correlation between the location of physicians' obstetrics–gynecology residency training and the quality of care they deliver as specialists after graduation, though initial performance and years of experience also contributed to the safety of the care they provided after training.[3] Similarly, location of residency training was associated with graduates' future ability to practice conservatively[4] and with their spending patterns.[5] As a result, efforts to improve the quality of graduate medical education training have focused on improving the clinical learning environment for residents while simultaneously ensuring the safety of the care delivered to patients.[6,7]

Recent efforts to optimize the clinical learning environment for education and patient care, including duty hours reform efforts by the Accreditation Council for Graduate Medical Education and more recently the Clinical Learning Environment Review program,[8] have increasingly incorporated residents' perceptions of their learning environments.[9-11] In the Netherlands, quality improvement efforts in graduate medical education have focused specifically on improving learning climates for residents.[12,13] The learning climate incorporates residents' shared perceptions of their learning environment, including formal and informal aspects of their training, as well as the relevant policies, practices, and procedures that affect that learning environment.[14,15] Previous studies found associations between residents' positive perceptions of their learning environment and higher performance on in-training examinations,[16] better quality of life, fewer symptoms of burnout,[17] and a lower likelihood of leaving practice.[18] In addition to contributing to residents' professional development, the learning climate also can contribute to patient safety. In particular, residents' perceptions of a supportive and judgment-free environment can reduce medical errors, as residents are more likely to consult an on-call physician for assistance.[19] Moreover, residents who perceive low levels of support in their learning environment are less likely to report medical errors, which can potentially affect patient outcomes.[20] Because these studies did not investigate actual patient care outcomes, the effects of the learning climate on patients can only be implied.

Although the clinical learning environment should foster both effective resident learning and safe patient care, no study to date has linked residency learning climate to patient outcomes. To better support residency programs in integrating high-quality residency training with patient safety, a better understanding of how the learning climate relates to patient outcomes in individual departments is needed. The purpose of this study was to investigate the association between the learning climate in obstetrics–gynecology departments and adverse perinatal and maternal outcomes.

Method

Setting and design

This study was set in obstetrics–gynecology departments in community teaching hospitals in the Netherlands. Community teaching hospitals are university-affiliated, nontertiary perinatal centers providing obstetrical care to women with lower-risk pregnancies who do not meet the criteria for transfer to a tertiary perinatal center. They also provide (1) training and supervision for interns and residents in their first two years of graduate medical education, during which time trainees are expected to learn low-risk obstetrics and basic surgery skills, as well as (2) one- to two year differentiation rotations for residents in the last two years of training.[21] We chose this setting because residents at community teaching hospitals perform low-risk deliveries in a relatively young and healthy patient population, and also because previous research has attributed the variation in labour outcomes in part to the variation in obstetrical care delivery and organization.[22–24]

We retrospectively analyzed learning climate data from the Dutch Residency Educational Climate Test (D-RECT) and clinical registration data from the Netherlands Perinatal Registry (PRN), which is a national database that combines anonymized pooled data about the mother, pregnancy, childbirth, child, and process of obstetrical care for 96% of all births in the Netherlands.[25] To ensure the largest sample possible, we studied data from the year in which the most departments evaluated their learning climate (January–December 2013). On the basis of the existing literature, we hypothesized that a more positive learning climate would be associated with lower odds of adverse obstetrical outcomes in obstetrics–gynecology departments in community teaching hospitals.

Data collection

Trainees (residents, fellows, and interns) were invited to evaluate the learning climate of their most recent obstetrics–gynecology department rotation using either an online or a paper-based D-RECT questionnaire, depending on hospital policy. We obtained permission to analyze anonymized D-RECT survey results from the participating departments that used the Web-based D-RECT platform and from the regional educational committees for those departments that used the paper-based D-RECT questionnaire. Departments using the online questionnaire invited trainees using an e-mail with a link to an anonymous questionnaire, sending up to three reminders to non-respondents. Trainees usually had one month to complete the questionnaire. Because trainees spend one to two years in a single department, most completed the questionnaire during the rotation that they were evaluating. The four departments that used the paper-based questionnaire made it available in their offices, asking trainees to evaluate the department where they had spent the most time in the last three months, which may not have been their current department. In all cases, trainees were ensured that their participation was voluntary and their responses anonymous.

We accessed anonymized patient data with special permission from the PRN (PRN-14.20). We excluded any deliveries meeting the criteria for transfer to a tertiary centre (i.e., <32 weeks gestational age, weight <1250 grams at birth, or serious maternal complications). Additionally, we excluded any births with congenital abnormalities due to increased risk of perinatal mortality.[22]

The institutional ethical review board of the Academic Medical Centre of the University of Amsterdam confirmed that the Medical Research Involving Human Subjects Act did not apply to this study because the study involved the use of existing data collected for quality improvement purposes and provided a waiver of ethical approval for the overall study design.

Measures

Residency learning climate. The D-RECT questionnaire was developed in the Netherlands based on qualitative research, expert opinion, and a Delphi panel.[26] It has since been extensively validated nationally[27] and increasingly used internationally.[28] The questionnaire consists of 35 items that assess nine domains of the residency learning climate: educational atmosphere, accessibility of supervisors, coaching and assessment, teamwork, role of the specialty tutor, formal education, peer collaboration, adaptation of work to residents' competence, and patient sign-out (see Appendix 1, Table A1.1 for more about the D-RECT questionnaire).[27] Responses are given on a five-point Likert scale (1 = totally disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = totally agree) with an additional option of "cannot evaluate." Higher scores indicate a more favourable learning climate. On the basis of previous large-scale validation studies, we included departments with at least three completed D-RECT questionnaires to ensure reliable overall learning climate scores.[26,27]

Adverse obstetrical outcomes. We divided obstetrical outcomes into perinatal outcomes and maternal outcomes. In the absence of a perinatal adverse outcomes index, which is still in development for the Netherlands, we used quality measures for inpatient obstetrics to define the relevant outcomes.[29] Adverse perinatal outcomes then included any combination of the following: fetal and early neonatal mortality <7 days after birth, five-minute Apgar score <7, or admission of the child to a neonatal intensive care unit in a tertiary center for ≥24 hours. Adverse maternal outcomes included postpartum hemorrhage ≥1000 ml and/or the need for a transfusion during or shortly after delivery, maternal death, uterine rupture, or third- or fourth-degree perineal laceration.

Covariates. We accounted for any known or potential covariates and confounders. Known covariates included maternal age,[22] parity (nulliparous/multiparous),[22] ethnicity (Western/

non-Western),[22,24] multiple gestations (multiple births/singletons), and maternal socioeconomic status.[30] We used socioeconomic status scores that were calculated by the Netherlands Institute for Social Research in 2006 using four-digit postal codes, which were then linked to the mean income level and percentage of households with low income or high unemployment in that area. Each score was categorized as low (<25th percentile), middle (25th–75th percentile), or high (>75th percentile).[31]

In addition to these known covariates, we also included two potential confounders: previous D-RECT evaluations and hospital volume. The number of previous D-RECT evaluations (1, 2–3, ≥4) refers to the number of times the department had evaluated its learning climate in the past. It was included because learning climate scores tend to improve over time.[13] Hospital volume was included because it had been associated with delivery complication rates[32,33] and because it can affect the learning climate. Hospital volume was defined as the total number of deliveries at ≥22 weeks of gestation in the department in the study year, categorized in tertiles as low (<1500 deliveries), average (1500–1750 deliveries), or high (>1750 deliveries).

Statistical and bias analyses

In total, we included 16 nontertiary obstetrics–gynecology departments that evaluated their learning climate between January and December 2013 in our study. We retrieved adverse outcomes data from the PRN for these departments for the same period (23629 births). We retrieved responses to the D-RECT questionnaires for the selected obstetrics–gynecology departments from a larger data set containing 559 evaluations of obstetrics–gynecology departments between 2009 and 2014. These data had been cleaned and imputed using an expectation maximization algorithm because the proportion of missing data (<5%) was considered ignorable.[34]

We determined each department's learning climate score by calculating the average D-RECT score for that department. To assess the reliability of the scores, we calculated the intraclass correlation coefficient (*2, k*), which estimates the reliability of multiple measurements of a group mean based on average group size (*k*) and can be interpreted as the proportion of the total variance in mean scores that can be explained by the differences between the departments.[35]

Because the included adverse outcomes were infrequent, we grouped departments in tertiles based on their overall learning climate scores and used multilevel logistic regressions to test whether the odds of adverse maternal and perinatal outcomes differed between the three groups. We chose multilevel logistic regression because it accounts for the clustering of patients within departments and minimizes between-department differences due to unmeasured department- or hospital-level characteristics.[36,37]

To assess the effects of the covariates, we built three separate models. In the first model, we estimated the odds of adverse perinatal and maternal outcomes given the department's learning climate score tertile, without adjusting for clustering or other factors. In the second model, we used a random intercept multilevel binary logistic model that controlled for maternal or pregnancy characteristics. In the final model, we added hospital volume and previous D-RECT evaluations as covariates to the second model. We calculated unadjusted and adjusted odds ratios (ORs and aORs, respectively) and their 95% confidence intervals (CIs).

We conducted sensitivity analyses to test the robustness of our results against the possible effects of grouping the variables differently and of selecting a different study population. First, we tested whether grouping the scores in quartiles instead of tertiles affected our results. Second, we hypothesized that a department's learning climate could affect only the deliveries performed by residents. We therefore repeated the analysis for only the deliveries performed by residents. Next, to minimize possible bias from our sampling strategy, we conducted subgroup analyses excluding multiple births and fetal deaths before the onset of labor (antepartum death). Finally, because we could not possibly account for all potential confounders, we conducted a bias analysis to quantify whether part or all of our results could be explained by uncontrolled confounding. Bias analysis is an innovative tool for adjusting observational study data for potential residual confounding due to unknown or unmeasured confounders.[38] We also investigated whether varying responses to the D-RECT questionnaire introduced selection bias and affected our results. We conducted the additional selection bias analysis on top of the uncontrolled confounding analysis as part of a multiple-bias modeling exercise to check the sensitivity of our results.[39] We implemented the bias adjustment using Monte Carlo methods.[38] See Appendix 2 for more about these additional analyses.

We used IBM SPSS Statistics for Windows, version 21 (IBM Corp., Armonk, New York) for imputing the missing D-RECT data. Multilevel logistic regressions were performed using the mle4 package in R statistical software version 3.2.0. The remaining analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina).

Results

Of 171 potential respondents (82% female), 103 (85% female) completed a D-RECT questionnaire (60% response rate, mean 6.4 questionnaires per department). The four departments using paper-based questionnaires allowed their trainees to evaluate the rotation where they spent the most time in the last three months, potentially giving these departments a response rate greater than 100%. Therefore, we estimated the overall response rate based on responses in departments using the online questionnaire (12 departments). The departments' learning climate scores ranged from 3.7 to 4.4 (mean 4.0, standard deviation 0.3) (see Figure 1). The intraclass correlation coefficient was 0.69, which means that 69% of the variance in departments' overall learning climate scores can be explained by differences between the departments. Table 1 includes the characteristics of the departments by learning climate score tertile, and Table 2 presents the included perinatal and maternal characteristics as well as the adverse outcomes, also by learning climate score tertile.

The lowest percentage of adverse perinatal outcomes (1.5%) was found in departments in the lowest learning climate score tertile, while the highest percentage of adverse perinatal outcomes (2.4%) was found in departments in the highest tertile. In the multilevel logistic regression analyses using overall learning climate score as a continuous variable, higher learning climate scores were associated with significantly greater odds of an adverse perinatal outcome (aOR 2.06, 95% CI 1.14–3.72) (see Table 3). Compared with departments in the lowest tertile, departments in the middle tertile had 46% greater odds of an adverse perinatal outcome (aOR 1.46, 95% CI 1.09–1.94), while departments in the highest tertile had 69% greater odds (aOR 1.69, 95% CI 1.24–2.30) (see Table 4).

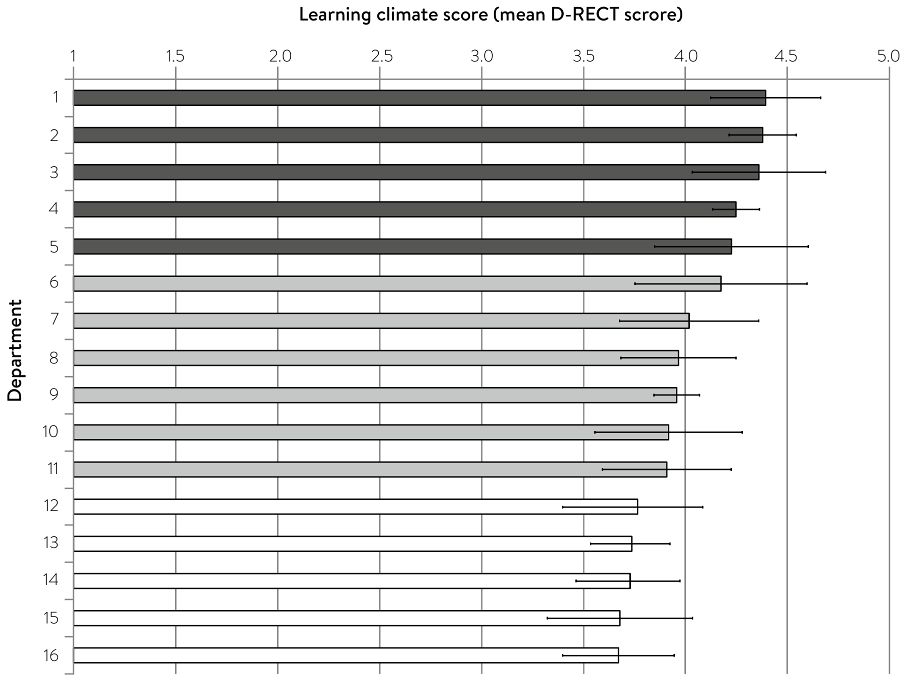


Figure 1. Mean learning climate scores for 16 nontertiary obstetrics–gynecology departments in the Netherlands that participated in a study of the association between learning climate and adverse obstetrical outcomes, 2013. Each bar represents the average Dutch Residency Educational Climate Test (D-RECT) score for one department on a scale of 1 to 5 with whiskers indicating the standard deviation. The dark gray bars indicate the departments in the highest learning climate score tertile (mean D-RECT score >4.17), the light gray bars indicate the departments in the middle tertile (mean D-RECT score 3.91–4.17), and the white bars indicate the departments in the lowest tertile (mean D-RECT score <3.91)

Table 1. Characteristics of learning climate evaluations and obstetrical practice at 16 nontertiary obstetrics and gynecology departments in the Netherlands, by learning climate score tertile, 2013

Characteristic	Overall learning climate score Lowest tertile (D-RECT score <3.91)	Middle tertile (D-RECT score 3.91-4.17)	Highest tertile (D-RECT score >4.17)	All
No. of departments	**5**	**6**	**5**	**16**
D-RECT questionnaire characteristics				
Completed questionnaires per department, mean (SD)	6.2 (2.3)	7.8 (2.7)	5.0 (2.3)	6.4 (2.6)
Online response rate (12 departments), mean (min-max)	66% (36%-90%)	66% (40%-91%)	48% (33%-80%)	60% (33%-91%)
Female respondents, % (SD)	92.0 (17.9)	79.2 (14.2)	86.1 (14.6)	85.4 (15.6)
Previous D-RECT evaluations, mean (SD)	2.2 (0.8)	3.0 (1.4)	2.2 (1.3)	2.5 (1.2)
Department characteristics				
Deliveries performed by a resident, no. (%)	1741 (44.2)[a]	4190 (45.2)	4220 (56.5)	10151 (49.1)[b]
Spontaneous vaginal deliveries, no. (%)	4822 (70.1)	6508 (70.1)	4723 (63.2)	16053 (67.9)
Assisted deliveries, no. (%)	697 (10.1)	886 (9.5)	916 (12.3)	2499 (10.6)
Primary caesarean section deliveries, no. (%)	711 (10.3)	881 (9.49)	811 (10.9)	2403 (10.2)
Secondary caesarean section deliveries, no. (%)	637 (9.3)	944 (10.2)	967 (13.0)	2548 (10.8)
Other types of deliveries, no. (%)	14 (0.2)	61 (0.7)	51 (0.7)	126 (0.5)
Deliveries in low volume hospitals (<1500 deliveries), no. (%)	3237 (47.0)	2617 (28.2)	3612 (48.4)	9466 (40.1)
Deliveries in high volume hospitals (>1750 deliveries), no. (%)	2149 (31.2)	5055 (54.5)	2387 (32.0)	9591 (40.6)

Abbreviations: D-RECT indicates Dutch Residency Educational Climate Test; SD, standard deviation.

[a] Based on 3 departments (n = 3935 births).

[b] Based on 14 departments (n = 20683 births).

Table 2. Maternal and perinatal characteristics and adverse outcomes of included births at 16 nontertiary obstetrics and gynecology departments in the Netherlands, by learning climate score tertile, 2013

	Overall learning climate score			
Characteristic	**Lowest tertile (D-RECT score <3.91)**	**Middle tertile (D-RECT score 3.91-4.17)**	**Highest tertile (D-RECT score >4.17)**	**All**
No. of included births	**6881**	**9280**	**7468**	**23629**
Maternal characteristics				
Maternal age, mean (SD)	31.2 (4.9)	30.2 (5.2)	31.9 (5.0)	31.0 (5.1)
Nulliparous, no. (%)	3697 (53.7)	4621 (49.8)	4001 (53.6)	12319 (52.1)
Non-Western ethnicity, no. (%)	1817 (26.4)	3274 (35.3)	1883 (25.2)	6974 (29.5)
Low socioeconomic status, no. (%)	1992 (29.0)	2884 (31.1)	1512 (20.25)	6388 (27.0)
High socioeconomic status, no. (%)	1344 (19.5)	2366 (25.5)	1979 (26.5)	5689 (24.1)
Pregnancy characteristics				
Gestational age at birth, mean in weeks/days (SD in days)	39/0 (2)	39/0 (2)	38/6 (2)	39/0 (2)
Multiple births, no. (%)	214 (3.1)	346 (3.7)	321 (4.3)	881 (3.7)
Child characteristics				
Birth weight in grams, mean (SD)	3387 (652)	3346 (549)	3377 (565)	3368 (586)
Deliveries with an adverse perinatal outcome, no. (%)	**101 (1.5)**	**200 (2.2)**	**178 (2.4)**	**479 (2.0)**
Fetal and early neonatal mortality <7 days, no. (%)	19 (0.3)	24 (0.3)	8 (0.1)	51 (0.2)
Admission to NICU ≥24 hours, no. (%)	5 (0.1)	6 (0.1)	11 (0.2)	22 (0.1)
Five-minute Apgar score <7, no. (%)	78 (1.1)	172 (1.9)	161 (2.2)	411 (1.7)
Deliveries with an adverse maternal outcome, no. (%)	**610 (8.9)**	**851 (9.2)**	**651 (8.7)**	**2112 (8.9)**
Postpartum hemorrhage ≥1000 ml and/or blood transfusion during or immediately after delivery, no. (%)	443 (6.4)	643 (6.93)	512 (6.9)	1598 (6.8)
Maternal death, no. (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Uterine rupture, no. (%)	0 (0.0)	2 (<0.1)	1 (<0.1)	3 (<0.1)
Third- of fourth-degree perineal laceration, no. (%)	195 (2.8)	226 (2.4)	159 (2.1)	580 (2.5)

Abbreviations: D-RECT indicates Dutch Residency Educational Climate Test; SD, standard deviation; NICU, neonatal intensive care unit.

Table 3. Unadjusted and adjusted odds ratios for the association between learning climate score as a continuous variable and covariates with adverse perinatal outcomes in 16 nontertiary obstetrics and gynecology departments in the Netherlands, 2013

Characteristic	Unadjusted OR (CI)	Model 1, adjusted OR (CI)[a]	Model 2, adjusted OR (CI)[b]
Overall learning climate score	1.68 (1.17 - 2.43)[c]	1.81 (1.00 - 3.26)[c]	2.06 (1.14 - 3.72)[c]
Covariates			
Nulliparity		1.31 (1.08 - 1.58)[c]	1.29 (1.07 - 1.57)[c]
Maternal age		1.02 (1.00 - 1.04)[c]	1.02 (1.00 - 1.04)[c]
Multiple births		2.35 (1.68 - 3.29)[c]	2.35 (1.68 - 3.29)[c]
Non-Western ethnicity		1.23 (1.00 - 1.51)[c]	1.21 (0.99 - 1.48)
High socioeconomic status (vs. middle)		0.74 (0.57 - 0.95)[c]	0.70 (0.54 - 0.91)[c]
Low socioeconomic status (vs. middle)		1.07 (0.86 - 1.34)	1.04 (0.82 - 1.31)
Two to three previous D-RECT evaluations (vs. 1)			1.09 (0.76 - 1.57)
Four or more previous D-RECT evaluations (vs. 1)			1.11 (0.78 - 1.59)
Low hospital volume, <1500 deliveries (vs. 1500-1750)			0.72 (0.49 - 1.06)
High hospital volume, >1750 deliveries (vs. 1500-1750)			0.93 (0.67 - 1.29)

Abbreviations: OR indicates odds ratio; CI, confidence interval; D-RECT, Dutch Residency Educational Climate Test.

[a] Model 1 includes nulliparity, maternal age, multiple births, ethnicity, and socioeconomic status.

[b] Model 2 includes all Model 1 covariates plus number of previous D-RECT evaluations and hospital volume.

[c] $P < 0.05$.

Table 4. Unadjusted and adjusted odds ratios for the association between learning climate score category and covariates with adverse perinatal outcomes in 16 nontertiary obstetrics and gynecology departments in the Netherlands, 2013

Characteristic	**Unadjusted OR (CI)**	**Model 1, adjusted OR (CI)[a]**	**Model 2, adjusted OR (CI)[b]**
Primary predictor			
Lowest learning climate score tertile	Reference	Reference	Reference
Middle learning climate score tertile	1.49 (1.17 - 1.89)[c]	1.54 (1.14 - 2.09)[c]	1.46 (1.09 - 1.94)[c]
Highest learning climate score tertile	1.64 (1.28 - 2.10)[c]	1.64 (1.20 - 2.26)[c]	1.69 (1.24 - 2.30)[c]
Covariates			
Nulliparity		1.31 (1.08 - 1.59)[c]	1.30 (1.07 - 1.57)[c]
Maternal age		1.02 (1.00 - 1.04)[c]	1.02 (1.00 - 1.04)[c]
Multiple births		2.34 (1.67 - 3.27)[c]	2.33 (1.67 - 3.27)[c]
Non-Western ethnicity		1.22 (1.00 - 1.49)[c]	1.20 (0.98 - 1.47)
High socioeconomic status (vs. middle)		0.72 (0.56 - 0.93)[c]	0.68 (0.52 - 0.89)[c]
Low socioeconomic status (vs. middle)		1.07 (0.85 - 1.33)	1.02 (0.81 - 1.29)
Two to three previous D-RECT evaluations (vs. 1)			1.08 (0.81 - 1.46)
Four or more previous D-RECT evaluations (vs. 1)			1.07 (0.79 - 1.43)
Low hospital volume, <1500 deliveries (vs. 1500-1750)			0.78 (0.56 - 1.07)
High hospital volume, >1750 deliveries (vs. 1500-1750)			0.93 (0.70 - 1.22)

Abbreviations: OR indicates odds ratio; CI, confidence interval; D-RECT, Dutch Residency Educational Climate Test.
[a] Model 1 includes nulliparity, maternal age, multiple births, ethnicity, and socioeconomic status.
[b] Model 2 includes all Model 1 covariates plus number of previous D-RECT evaluations and hospital volume.
[c] $P < 0.05$.

Covariates that significantly contributed to adverse perinatal outcomes included multiple births, nulliparity, maternal age, and socioeconomic status. Neither previous number of D-RECT evaluations nor hospital volume significantly contributed to adverse perinatal outcomes.

The highest percentage of deliveries with an adverse maternal outcome (9.2%) was found in departments in the middle tertile of learning climate scores. However, in multilevel logistic regression analyses, we did not find a significant association between learning climate score and odds of an adverse maternal outcome (middle vs. lowest tertile: 9.2% vs. 8.9%; OR 1.04, 95% CI 0.93–1.16; highest vs. lowest tertile: 8.7% vs. 8.9%; OR 0.98, 95% CI 0.88–1.10).

The results of the sensitivity and bias analyses are reported in Appendix 1 and 2, respectively. In the sensitivity analyses, the protective effect of lower learning climate scores against adverse perinatal outcomes remained when D-RECT scores were grouped into quartiles, with the two highest quartiles having significantly greater odds of an adverse perinatal outcome and no difference in the odds of an adverse maternal outcome (see Appendix 1, Table A1.2). The pattern did not change when we performed the subgroup analyses (in tertiles) of deliveries performed by residents only, nor when we performed the subgroup analyses of deliveries excluding multiple births and stillbirths (see Appendix 1, Tables A1.3–A1.5).

The results of the bias analysis showed that controlling for an unmeasured confounder set that increased the learning climate score but decreased the odds of adverse outcomes (or vice versa) would increase the strength of the observed positive association between learning climate and adverse perinatal outcomes (see Appendix 2, Figures A2.1 and A2.2). Controlling for an unmeasured confounder set that positively (or negatively) affected both learning climate scores and the odds of adverse perinatal outcomes would result in a less positive or even reversed association, implying that higher learning climate scores could be associated with fewer adverse perinatal outcomes (see Appendix 2, Figures A2.3 and A2.4). The multiple bias analysis that also adjusted for selection bias (due to biasing response rates) revealed similar conclusions (see Appendix 2, Figures A2.5–A2.8).

Discussion

Explanation of findings

Our study is the first to our knowledge to investigate the association between residency learning climate and adverse perinatal and maternal outcomes. We found that departments' learning climate scores are associated with increased odds of adverse perinatal outcomes but have no association with the odds of adverse maternal outcomes.

In their commentary on evaluating residency programs on the basis of patient outcomes, Asch and colleagues[40] posited that the strength of the association between residency programs and patient outcomes may vary across medical conditions. In other words, a program may provide excellent care for one condition but not for another. Our study supports this argument, as we observed differences in the associations between learning climate and maternal and perinatal outcomes. Herein, our choice of outcomes measures guided the associations we observed. By studying the odds of any predefined adverse maternal or perinatal outcome, instead of the odds of specific outcomes, we could have masked existing associations with specific complications. In our sample, for instance, as the learning climate score tertile increased, the percentage of severe perineal lacerations decreased, while the percentage of severe maternal blood loss cases increased. As a result, these mixed trends could have diluted the overall association we observed.

On the other hand, the association between learning climate score and adverse perinatal outcomes seemed to be driven solely by low Apgar scores, as fetal or early neonatal mortality and neonatal intensive care unit admissions were extremely rare. In term infants, a low five-minute Apgar score (<7) is related to a higher risk of death[41], neurological disability including cerebral palsy, and long-term cognitive impairment in adulthood.[42,43] While the etiology of a low five-minute Apgar score is multifactorial, in term infants, low Apgar scores have been associated with high-risk interventions during labour[44] as well as with the quality of neonatal resuscitation.[42] The difference in adverse perinatal outcomes between the highest and lowest learning climate score tertiles was 0.9% in absolute terms, which is a small, but significant, difference in potentially preventable perinatal complications, especially when considered over time and across multiple departments.

The association between residency learning climate and adverse perinatal outcomes conflicted with our initial hypothesis and with findings from previous studies linking residents' perceptions with an increased likelihood to consult with a physician on call[19] or to report medical errors.[20] Notably, previous studies focused on individual learning climate perceptions, while our study focused on the learning climate as a department-level characteristic. Our results echoed those from previous studies that found weak and even negative associations between measures of quality on a department or hospital level and individual patient outcomes, even when there was a demonstrable positive association at the individual level.[45–47]

According to Finney and colleagues,[47] aggregated variables reflect department-level characteristics, which in turn are influenced by a different set of confounders than individual-level variables. Analyzing the subgroup of deliveries performed by residents only did not change our results, which reinforces our view that the overall learning climate is not just a measure of individual residents' participation in deliveries but also reflects attributes of the whole department. Hence, the positive association we observed between a department's overall learning climate and the odds of adverse perinatal outcomes is likely explained by department processes that can affect both patient safety and the learning climate in the department. Clinical environments are not specifically designed to facilitate learning and face unique challenges in balancing the demands of patient care and those of resident training.[48,49] It is, therefore, possible that department processes that simultaneously contribute to the learning climate and patient outcomes compete for the department's (limited) resources, potentially leading to tensions between the two functions.

Among the covariates we included in our analysis, maternal age, nulliparity, multiple births, and lower socioeconomic status were significant, which corresponded to previous findings in the literature.[22,24,30] Number of previous D-RECT evaluations and hospital volume were not significant; however, we kept them in the model because they were considered potential confounding factors. Our bias analyses showed that uncontrolled confounding could strengthen or weaken the observed associations depending on the direction and strength of the relationship between the unmeasured confounder and the learning climate and adverse perinatal outcome (see Appendix 2 for details). In the bias analyses, including the selection bias analyses, we found that this relationship would be weakened if the unmeasured confounder set simultaneously increased (or decreased) the learning climate score (a desirable result) and the odds of an adverse outcome (an undesirable result). However, we cannot identify a confounder that would cause such a change in practice.

Limitations

Given the observational nature of our study, we cannot make any claims of causality. While we have taken great care to identify relevant outcomes using a high-quality national database and a validated instrument for measuring the learning climate to maximize validity, by using data from quality improvement databases, we could not account for all potential sources of bias, including measurement error. Because this work is a hypothesis-generating study, our results may not be generalizable to other time periods, clinical outcomes, or clinical contexts. In particular, this study did not include academic hospitals, which may better integrate learners in clinical care. Furthermore, because of the de-identified nature of the patient data and anonymized D-RECT questionnaires, we could not study how individual residents' perceptions of their learning climate, or their characteristics such as performance and previous experience, may have affected patient outcomes. Therefore, we could not compare our results with those from other studies examining outcomes of care delivered by individual residents or obstetricians.[2,3,50,51]

Future research

Future research is needed to confirm our findings in other clinical contexts, including academic and nonacademic hospitals, as well as to explore the relationship between resident- and department-level learning climate characteristics and other metrics of quality and safety, such as patient safety climate. Learning climate measurements should be standardized, and efforts should optimize response rates to increase the validity of results and enable the analysis of learning climate subscales. In addition, the moderating role of learning climate strength (the agreement between residents on given scores) should be investigated. Investigators also could focus on exploring potential effect mechanisms, modifiers, and other explanations, by better characterizing lower- and higher-scoring learning environments that systematically report better perinatal outcomes. Particularly, the effects of contributing factors, such as residents' experience, composition of the resident group, supervisory activities, and the role of supervisors' site of training, should be investigated. Assessing the perceptions of formal educators, staff, and program directors could provide additional information about the clinical learning environment as it relates to the residency learning climate as well as to perinatal outcomes.[52] Longitudinal studies should specifically explore the long-term outcomes of physicians trained in different learning climates. Finally, extensive (multiple-) bias analysis to detect possible systematic error due to uncontrolled confounding, selection bias, and measurement error should be part of future studies that are large enough to make random sampling error negligible.[39,53-55]

Conclusions

We found a positive association between learning climate and the odds of adverse perinatal, but not maternal, outcomes. These findings should be considered when optimizing clinical learning environments—for example, by emphasizing patient safety and quality improvement in graduate medical education. Research in similar clinical contexts involving informed covariates and examining department-specific effect modification is needed to replicate our findings and to explore potential mechanisms behind the associations between learning climate and patient outcomes in clinical teaching departments.

References:

1. Leach DC, Philibert I. High-quality learning for high-quality health care: getting it right. *JAMA*. 2006;296(9):1132-1134.
2. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA*. 2009;302(12):1277-1283.
3. Epstein AJ, Srinivas SK, Nicholson S, Herrin J, Asch DA. Association between physicians' experience after training and maternal obstetrical outcomes: cohort study. *BMJ*. 2013;346:f1596.
4. Sirovich BE, Lipner RS, Johnston M, Holmboe ES. The association between residency training and internists' ability to practice conservatively. *JAMA Intern Med*. 2014;174(10):1640-1648.
5. Chen C, Petterson S, Phillips R, Bazemore A, Mullan F. Spending patterns in region of residency training and subsequent expenditures for care provided by practicing physicians for Medicare beneficiaries. *JAMA*. 2014;312(22):2385-2393.
6. Weiss KB, Bagian JP, Nasca TJ. The clinical learning environment: the foundation of graduate medical education. *JAMA*. 2013;309(16):1687-1688.
7. Nasca TJ, Weiss KB, Bagian JP. Improving clinical learning environments for tomorrow's physicians. *N Engl J Med*. 2014;370(11):991-993.
8. Weiss KB, Wagner R, Nasca TJ. Development, testing, and implementation of the ACGME Clinical Learning Environment Review (CLER) program. *J Grad Med Educ*. 2012;4(3):396-398.
9. Borman KR, Jones AT, Shea JA. Duty hours, quality of care, and patient safety: general surgery resident perceptions. *J Am Coll Surg*. 2012;215(1):70-79.
10. Holt KD, Miller RS, Philibert I, Heard JK, Nasca TJ. Residents' perspectives on the learning environment: data from the Accreditation Council for Graduate Medical Education resident survey. *Acad Med*. 2010;85(3):512-518.
11. Philibert I. Satisfiers and hygiene factors: residents' perceptions of strengths and limitations of their learning environment. *J Grad Med Educ*. 2012;4(1):122-127.
12. The Royal Dutch Medical Association (KNMG). Directive of the Central College of Medical Specialists (in Dutch). https://www.knmg.nl/opleiding-herregistratie-carriere/cgs/regelgeving/oude-regelgeving/oude-regelgeving-ccms.htm. Published November 11, 2015. Accessed November 7, 2017.
13. Silkens ME, Arah OA, Scherpbier AJ, Heineman MJ, Lombarts KM. Focus on quality: investigating residents' learning climate perceptions. *PLoS One*. 2016;11(1):e0147108.
14. Roff S, McAleer S. What is educational climate? *Med Teach*. 2001;23(4):333-334.
15. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education - a unifying perspective. *Med Teach*. 2001;23(5):445-454.
16. Shimizu T, Tsugawa Y, Tanoue Y, et al. The hospital educational environment and performance of residents in the General Medicine In-Training Examination: a multicenter study in Japan. *Int J Gen Med*. 2013;6:637-640.
17. van Vendeloo SN, Brand PL, Verheyen CC. Burnout and quality of life among orthopaedic trainees in a modern educational programme: importance of the learning climate. *Bone Joint J.* 2014;96-B(8):1133-1138.
18. Degen C, Weigl M, Glaser J, Li J, Angerer P. The impact of training and working conditions on junior doctors' intention to leave clinical practice. *BMC Med Educ*. 2014;14:119.
19. Naveh E, Katz-Navon T, Stern Z. Resident physicians' clinical training and error rate: the roles of autonomy, consultation, and familiarity with the literature. *Adv Health Sci Educ Theory Pract.* 2015;20(1):59-71.
20. National training survey 2014: concerns about patient safety. Manchester, UK: General Medical Council; 2014.
21. Scheele FC, N.; van Luijk, S.; den Rooyen, C.; van Loon, K. Better Education for Obstetrics and Gynecology: Dutch National Competency Based Curriculum for Obstetrics & Gynaecology (NL). Utrecht: Dutch Association for Obstetrics and Gynecology;2013.
22. Poeran J, Borsboom GJ, de Graaf JP, Birnie E, Steegers EA, Bonsel GJ. Population attributable risks of patient, child and organizational risk factors for perinatal mortality in hospital births. *Matern Child Health J.* 2015;19(4):764-775.
23. de Graaf JP, Ravelli AC, Visser GH, et al. Increased adverse perinatal outcome of hospital delivery at night. *BJOG*. 2010;117(9):1098-1107.
24. Vos AA, Denktas S, Borsboom GJ, Bonsel GJ, Steegers EA. Differences in perinatal morbidity and mortality on the neighbourhood level in Dutch municipalities: a population based cohort study. *BMC Pregnancy Childbirth*. 2015;15(1):201.

25. The Netherlands Perinatal Registry Trends 1999-2012 (in Dutch). Utrecht: The Foundation of The Netherlands Perinatal Registry; 2013.
26. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011;33(10):820-827.
27. Silkens ME, Smirnova A, Stalmeijer RE, et al. Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. *Med Teach*. 2016;38(5):476-481.
28. Piek J, Bossart M, Boor K, et al. The work place educational climate in gynecological oncology fellowships across Europe: the impact of accreditation. *Int J Gynecol Cancer*. 2015;25(1):180-190.
29. Bailit JL. Measuring the quality of inpatient obstetrical care. *Obstet Gynecol Surv.* 2007;62(3):207-213.
30. Ravelli AC, Jager KJ, de Groot MH, et al. Travel time from home to hospital and adverse perinatal outcomes in women at term in the Netherlands. *BJOG*. 2011;118(4):457-465.
31. de Graaf JP, Ravelli AC, de Haan MA, Steegers EA, Bonsel GJ. Living in deprived urban districts increases perinatal health inequalities. *J Matern Fetal Neonatal Med*. 2013;26(5):473-481.
32. Kyser KL, Lu X, Santillan DA, et al. The association between hospital obstetrical volume and maternal postpartum complications. *Am J Obstet Gynecol*. 2012;207(1):42.e41-17.
33. Sebastiao YV, Womack LS, Lopez Castillo H, et al. Hospital variations in unexpected complications among term newborns. *Pediatrics*. 2017.
34. Dong Y, Peng CY. Principled missing data methods for researchers. *Springerplus*. 2013;2(1):222.
35. Bliese PD. Within-group agreement, non-independence, and reliability: implications for data aggregation and analysis. In: Klein KJ, Kozlowski SWJ, ed. Multilevel theory, research, and methods in organizations: foundations, extensions, and new directions. San Francisco: Jossey-Bass Inc.; 2000.
36. Brumback BA, Dailey AB, Brumback LC, Livingston MD, He Z. Adjusting for confounding by cluster using generalized linear mixed models. *Stat Probab Lett*. 2010;80(21–22):1650-1654.
37. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press; 2007.
38. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol*. 2008;18(8):637-646.
39. Thompson CA, Arah OA. Selection bias modeling using observed data augmented with imputed record-level probabilities. *Ann Epidemiol*. 2014;24(10):747-753.
40. Asch DA. Evaluating residency programs by whether they produce good doctors. *LDI Issue Brief*. 2009;15(1):1-4.
41. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med*. 2001;344(7):467-471.
42. Ehrenstein V. Association of Apgar scores with death and neurologic disability. *Clin Epidemiol*. 2009;1:45-53.
43. Ehrenstein V, Pedersen L, Grijota M, Nielsen GL, Rothman KJ, Sorensen HT. Association of Apgar score at five minutes with long-term neurologic disability and cognitive function in a prevalence study of Danish conscripts. *BMC Pregnancy Childbirth*. 2009;9:14.
44. Lai S, Flatley C, Kumar S. Perinatal risk factors for low and moderate five-minute Apgar scores at term. *Eur J Obstet Gynecol Reprod Biol*. 2017;210:251-256.
45. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006;296(22):2694-2702.
46. Bradley EH, Herrin J, Elbel B, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. *JAMA*. 2006;296(1):72-78.
47. Finney JW, Humphreys K, Kivlahan DR, Harris AH. Why health care process performance measures can have different relationships to outcomes for patients and hospitals: understanding the ecological fallacy. *Am J Public Health*. 2011;101(9):1635-1642.
48. Hoffman KG, Donaldson JF. Contextual tensions of the clinical environment and their influence on teaching and learning. *Med Educ*. 2004;38(4):448-454.
49. Dornan T. Workplace learning. *Perspect Med Educ*. 2012;1(1):15-23.
50. Aiken CE, Aiken AR, Park H, Brockelsby JC, Prentice A. Factors associated with adverse clinical outcomes among obstetrics trainees. *Med Educ*. 2015;49(7):674-683.
51. van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med*. 2012;10:65.

52. Roth LM, Severson RK, Probst JC, et al. Exploring physician and staff perceptions of the learning environment in ambulatory residency clinics. *Fam Med*. 2006;38(3):177-184.
53. Rothman KJ, Lash TL, Greenland S. Modern epidemiology. 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
54. Arah OA, Sudan M, Olsen J, Kheifets L. Marginal structural models, doubly robust estimation, and bias analysis in perinatal and paediatric epidemiology. *Paediatr Perinat Epidemiol*. 2013;27(3):263-265.
55. Arah OA. Bias analysis for uncontrolled confounding in the health sciences. *Annu Rev Public Health*. 2017;38:23-38.

Appendix 1

Table A1.1. Items in the Dutch Residency Educational Climate Test (D-RECT) used to measure the learning climate in 16 non-tertiary obstetrics and gynecology departments in the Netherlands, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

Educational atmosphere

1. Continuity of care is not affected by differences of opinion between attendings.
2. Differences of opinion between attendings about patient management are discussed in such a manner that is instructive to others present.
3. Differences of opinion are not such that they have a negative impact on the work climate.
4. There is (are) NO attending physician(s) who have a negative impact on the educational climate.
5. My attendings treat me with respect.

Teamwork

6. Attendings, nursing staff, other allied health professionals and residents work together as a team.
7. Nursing staff and other allied health professionals make a positive contribution to my training.
8. Nursing staff and other allied health professionals are willing to reflect with me on the delivery of patient care.

Role of specialty tutor

9. The specialty tutor monitors the progress of my training.
10. The specialty tutor provides guidance to other attendings when needed.
11. The specialty tutor is actively involved in improving the quality of education and training.
12. In this rotation evaluations are useful discussions about my performance.
13. My plans for the future are part of the discussion.
14. During evaluations, input from several attendings is considered.

Coaching and assessment

15. My attendings take the initiative to evaluate my performance.
16. My attendings take the initiative to evaluate difficult situations I have been involved in.
17. My attendings evaluate whether my performance in patient care is commensurate with my level of training
18. My attendings occasionally observe me taking a history.
19. My attendings assess not only my medical expertise but also other skills such as teamwork, organization or professional behavior.
20. My attendings give regular feedback on my strengths and weaknesses.

Formal education

21. Residents are generally able to attend scheduled educational activities.
22. Educational activities take place as scheduled.
23. Attendings contribute actively to the delivery of high-quality formal education.
24. Formal education and training activities are appropriate to my needs.

Resident peer collaboration

25. Residents work well together.
26. Residents, as a group, make sure the day's work gets done.
27. Within our group of residents it is easy to find someone to cover or exchange a call.

Work is adapted to residents' competence

28. The work I am doing is commensurate with my level of experience.
29. The work I am doing suits my learning objectives at this stage of my training.
30. It is possible to do follow up with patients.

Accessibility of supervisors

31. When I need an attending, I can always contact one.
32. When I need to consult an attending, they are readily available.
33. It is clear which attending supervises me.

Patient sign-out

34. Sign-out is used as a teaching opportunity.
35. Attendings encourage residents to join in the discussion during sign-out.

Table A1.2. Unadjusted odds ratios (and 95% confidence intervals) for the association between learning climate score category in quartiles and adverse perinatal and maternal outcome (n=23629 deliveries) in 16 non-tertiary obstetrics and gynecology departments in the Netherlands, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

Learning climate score category (quartile)	**Total number of deliveries, n (%)**	**Total deliveries with an adverse perinatal outcome, n (%)**	**Total deliveries with an adverse maternal outcome, n (%)**	**Crude Odds Ratio for adverse peri-natal outcome**	**Crude Odds Ratio for adverse maternal outcome**
Low Q1(<3.75)	6084 (25.8)	98 (1.6)	536 (8.8)	Reference	Reference
Q2 (3.75 – 3.96)	5741 (24.3)	108 (1.9)	516 (9.0)	1.17 (0.89 – 1.54)	1.02 (0.90 – 1.16)
Q3 (3.97-4.24)	6723 (28.5)	162 (2.4)	630 (9.4)	1.51 (1.17 – 1.94)*	1.07 (0.95 – 1.21)
High Q4 (>4.24)	5081 (21.5)	111 (2.2)	430 (8.5)	1.36 (1.04 – 1.79)*	0.96 (0.84 – 1.09)

*P < 0.05.

Table A1.3. Unadjusted odds ratios (and 95% confidence intervals) for the association between learning climate score category and adverse perinatal and maternal outcome for deliveries performed by residents only (n=10151 deliveries based on 14 hospitals using available information) in 16 non-tertiary obstetrics and gynecology departments in the Netherlands, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

Learning climate score category (tertile)	**Crude Odds Ratio for adverse perinatal outcome**	**Crude Odds Ratio for adverse maternal outcome**
Lowest learning climate tertile	Reference	Reference
Middle learning climate tertile	1.49 (0.94 – 2.38)	1.07 (0.86 – 1.32)
Highest learning climate tertile	1.67 (1.05 – 2.64)*	1.13 (0.91 – 1.40)

*P < 0.05.

Table A1.4. Unadjusted odds ratios (and 95% confidence intervals) for the association between learning climate score category and adverse perinatal and maternal outcome for singletons only (n=22748 deliveries) in 16 non-tertiary obstetrics and gynecology departments in the Netherlands, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

Learning climate score category (tertile)	**Crude Odds Ratio for adverse perinatal outcome**	**Crude Odds Ratio for adverse maternal outcome**
Lowest learning climate tertile	Reference	Reference
Middle learning climate tertile	1.37 (1.08 – 1.76)*	1.01 (0.90 – 1.13)
Highest learning climate tertile	1.43 (1.11 – 1.84)*	0.96 (0.86 – 1.09)

*P < 0.05.

Table A1.5. Unadjusted odds ratios (and 95% confidence intervals) for the association between learning climate score category and adverse perinatal and maternal outcome for subgroup excluding stillbirths (n=23592 deliveries) in 16 non-tertiary obstetrics and gynecology departments in the Netherlands, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

Learning climate score category (tertile)	**Crude Odds Ratio for adverse peri-natal outcome**	**Crude Odds Ratio for adverse maternal outcome**
Lowest learning climate tertile	Reference	Reference
Middle learning climate tertile	1.54 (1.20 – 2.00)*	1.04 (0.93 – 1.16)
Highest learning climate tertile	1.82 (1.40 – 2.36)*	0.98 (0.87 – 1.10)

*P < 0.05.

Appendix 2

Bias Analysis for Uncontrolled Confounding of the Relation Between Learning Climate and Adverse Obstetrical Outcomes

Like in all observational studies that investigate association or causation, this study can suffer from bias and uncertainty due to unmeasured confounder(s)[1-3] of the relation of between learning climate and adverse obstetrical outcomes, after controlling for measured confounders. To say our study is threatened by an unmeasured confounder *U* after controlling for measured confounders, we mean one of the following possible scenarios: (a) *U* is an unmeasured cause of adverse obstetric care outcome that is also a cause of learning climate; (b) *U* is an unmeasured cause of adverse obstetric care outcome but is only associated (i.e. shares another unmeasured common cause) with learning climate; (c) *U* is an unmeasured cause of learning climate but is only associated (i.e. shares another unmeasured common cause) with adverse obstetrical care outcome. The difference between scenario (a) and scenarios (b) and (c) is that *U* has a causal relationship with both variables. Although all of these situations could be possible in our study, we decided to focus our bias analysis on scenario (a) where *U* is the unmeasured cause of both adverse obstetrical care outcome as well as the learning climate because its results are applicable to all other scenarios.

Directed acyclic graphs (DAGs) have been widely used in other fields such as epidemiology, computer science, social sciences, and recently medical education research to illustrate causal relations between variables.[4,5] DAGs are used to explicate causal assumptions about relations between variables by drawing arrows in the direction of the causation.[4] The arrows can be augmented to indicate either a positive causation (+) or a negative causation (–). We constructed four DAGs for the following scenarios: (i) where the unmeasured confounder *U* has a positive effect on the learning climate and a negative effect on adverse obstetrical outcome (Figure A2.1); (ii) where the unmeasured confounder set *U* has a negative effect on learning climate and a positive effect on adverse obstetrical outcome (Figure A2.2); (iii) *U* has a positive effect on the learning climate and adverse obstetrical outcome (Figure A2.3); (iv) *U* has a negative effect on the learning climate and adverse obstetrical outcome (Figure A2.4).

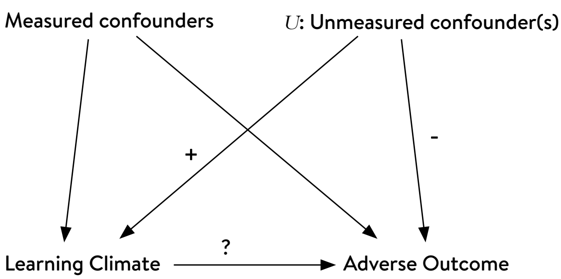


Figure A2.1. Directed acyclic graph (DAG) depicting uncontrolled confounding due to an unmeasured common cause *U* of both the learning climate and adverse outcome. In this diagram *U* positively affects the learning climate while decreasing the adverse outcome.

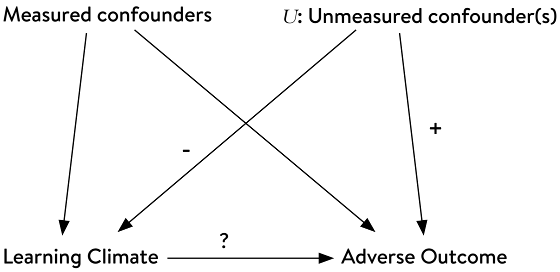


Figure A2.2. Directed acyclic graph (DAG) depicting uncontrolled confounding due to an unmeasured common cause *U* of both learning climate and adverse outcome. In this diagram *U* negatively affects learning climate but positively affects adverse outcome.

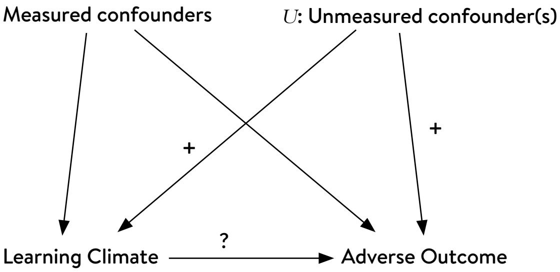


Figure A2.3. Directed acyclic graph (DAG) depicting uncontrolled confounding due to an unmeasured common cause *U* of learning climate and adverse outcome. Here, *U* increases both learning climate and adverse outcome.

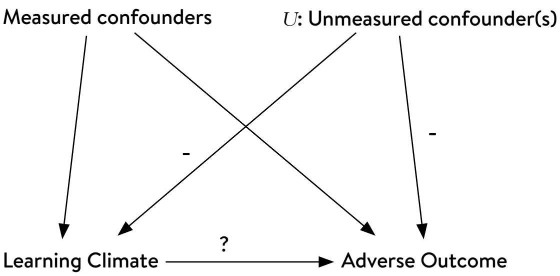


Figure A2.4. Directed acyclic graph (DAG) depicting uncontrolled confounding due to an unmeasured common cause *U* of both learning climate and adverse outcome. Here, *U* decreases both learning climate and adverse outcome.

In Figure A2.1, *U* positively affects the learning climate and negatively affects the adverse outcome, then accounting for this confounder would result in a net negative effect on the relation between learning climate and adverse outcomes. Therefore, controlling for this type of confounder would only strengthen the observed positive relationship between learning climate and adverse obstetrical outcomes. In Figure A2.2, *U* negatively affects the learning climate but positively affects the adverse obstetrical outcome; therefore, we would also expect a net negative uncontrolled confounding of the relation between learning climate and adverse outcomes due to unmeasured *U*. This implies that the potentially biased association between learning climate and adverse outcomes (not adjusted for *U*) could be attenuated by such negative confounding. Thus, adjusting for *U* in in Figures A2.1 and A2.2 would strengthen or increase the association. We believe the scenarios depicted in the Figures A2.1 and A2.2 are more plausible than those of A2.3 and A2.4 because it is more likely that an unmeasured factor that would increase (or decrease) learning climate would also be decrease (or increase) adverse outcomes.

In Figures A2.3 and A2.4, *U* affects both learning climate and adverse outcome in the same direction (either by increasing or decreasing both); therefore, we would expect a net positive uncontrolled confounding of the relation between learning climate and adverse obstetrical outcomes. Therefore, controlling for this type of confounding will attenuate the observed positive association between learning climate and adverse outcomes. We take this scenario to be less plausible because it is hard to imagine the type of factors or confounders that would simultaneously increase (or decrease) learning climate (a desirable result) and adverse outcomes (an undesirable result).

Using bias formulas to adjust the observed odds ratios for uncontrolled confounding due to unmeasured confounder *U* in each DAG, we estimated new odds ratios (and their corresponding 95% confidence intervals) after adjustment for the unmeasured confounder *U*.[6-10] Table A2.1 reports the bias adjusted odds ratios after bias analysis for uncontrolled confounding due to unmeasured confounder *U* in each of the four possible scenarios of uncontrolled confounding described above.

Table A2.1. Bias adjusted odds ratios (95% confidence intervals) for the relation between learning climate (*LC*) and adverse obstetric outcomes (*Y*) in 16 non-tertiary obstetrics and gynecology departments in the Netherlands, after bias analysis for uncontrolled confounding due to unmeasured confounder *U**, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

	Odds ratio relating unmeasured confounder *U* to adverse outcome *Y*, conditional on measured confounders			
	Negative relation (-)		**Positive relation (+)**	
Regression coefficient relating U to LC	**0.25**	**0.50**	**2**	**4**
Positive (+)	**DAG 1 (Figure A2.1)**		**DAG 3 (Figure A2.3)**	
0.10	4.50 (1.57 – 12.91)	3.20 (1.17 – 8.76)	1.17 (0.50 – 2.76)	0.72 (0.33 – 1.57)
0.25	8.99 (3.46 – 23.36)	4.80 (1.93 – 11.97)	0.78 (0.35 – 1.72)	0.33 (0.16 – 0.68)
0.50	9.92 (4.37 – 22.57)	4.96 (2.28 – 10.83)	0.81 (0.40 – 1.62)	0.34 (0.17 – 0.66)
Negative (-)	**DAG 4 (Figure A2.4)**		**DAG 2 (Figure A2.2)**	
-0.10	0.84 (0.29 – 2.42)	1.20 (0.44 – 3.28)	3.41 (1.45 – 8.06)	5.75 (2.65 – 12.52)
-0.25	0.43 (0.17 – 1.13)	0.82 (0.33 – 2.04)	5.21 (2.37 – 11.49)	12.83 (6.20 – 26.60)
-0.50	0.41 (0.18 – 0.93)	0.82 (0.38 – 1.79)	5.15 (2.57 – 10.33)	12.45 (6.41 – 24.22)

*Prevalence of the *U* is set at P(*U* = 1) = 0.4 for the probabilistic bias analysis. Similar result patterns are obtained at different *U* prevalence values or distributions.

Results of the bias analysis indicate that adjusting for an unmeasured confounder in scenarios 1 and 2 (Fig. A2.1 and Fig. A2.2) would result in larger ORs and, as a result, strengthening of the observed association between learning climate and higher risk of adverse perinatal outcome. However, in situations where the unmeasured confounder has a positive or negative effect on both learning climate and adverse perinatal outcomes in scenarios 3 and 4 (Fig. A2.3 and Fig. A2.4), the association will be weaker and may even become reversed if the unmeasured confounder is strongly associated with both the learning climate and risk of adverse perinatal outcome.

In conclusion, unmeasured confounders may have substantial effects on the relationship between department's learning climate and risk of adverse perinatal outcome depending on the nature and strength of their relationship to both the learning climate and adverse perinatal outcome. Thus, controlling for variables that increase the learning climate of a department but decrease the risk of the perinatal outcome (or vice versa), the strength of the association between learning climate and risk of adverse perinatal outcomes would increase. On the other hand, after controlling for variables that either increase or decrease both the learning climate and the risk of adverse perinatal outcome, then a more positive learning climate would have a protective effect on adverse perinatal outcomes.

Selection Bias Analysis of the Relation Between Learning Climate and Adverse Obstetrical Outcomes

Our study results could have potentially been biased due to self-selection of respondents on the D-RECT questionnaire that was used to evaluate the departments' learning climate. To say that our results could potentially be affected by selection bias means that the learning climate (or cause of the learning climate) and the adverse outcome both directly or indirectly affect selection into or participation in the study.[11] In our case, it was possible that selection ($S = 1$) would be caused by the learning climate, an unmeasured confounder (U), measured confounders and the adverse outcomes. We investigated possible scenarios of multiple-bias due to selection and uncontrolled confounding by building on the DAGs in Figures A2.1 to A2.4. The new DAGs in Figures A2.5 to A2.8 respectively built on Figures A2.1 to A2.4 by allowing for selection bias to be dependent on learning climate, adverse outcomes, unmeasured confounder U, and measured confounders. If selection bias was caused only by the learning climate and the measured confounders, then our multivariable adjustment of those measured confounding variables will suffice to remove the selection bias in our primary analysis, the selection bias.[8,11] However, there would be residual selection bias if selective participation or response was additionally caused by an unmeasured confounder (or risk factor) and/or the adverse outcomes under study. The selection bias analysis reported here focuses on this more challenging residual selection bias, possibly in the presence of uncontrolled confounding.[1,3,6,8-11] Although we focused on the scenarios in which learning climate and U decreased response but adverse outcomes increased response directly, we explored other unreported scenarios. Those unreported scenarios yielded similar conclusions as the ones reported here.

We used inverse probability of selection weighting and investigator-specified obtained selection bias parameters to adjust the observed odds ratios for selection bias in each DAG.[11] Using published bias analytical methods described elsewhere,[6,8,10,11] we estimated new multiple-bias adjusted odds ratios and their corresponding multiple-bias adjusted 95% confidence intervals after adjustment for both uncontrolled confounding[10] and selection bias.[11] Table A2.2 reports the multiple-bias adjusted odds ratios for the DAGs in Figures A2.5 to A2.8.

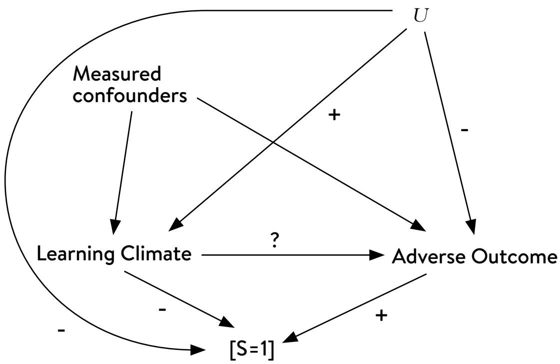


Figure A2.5. Directed acyclic graph (DAG) depicting selection or response (S = 1) bias caused by the adverse outcome (positively), the learning climate (negatively) and an unmeasured confounder U (negatively). Here, U decreases study response (S = 1) and adverse outcome but increases learning climate. [Arrow from measured confounders only omitted for visual reasons but can be assumed without loss of generality.]

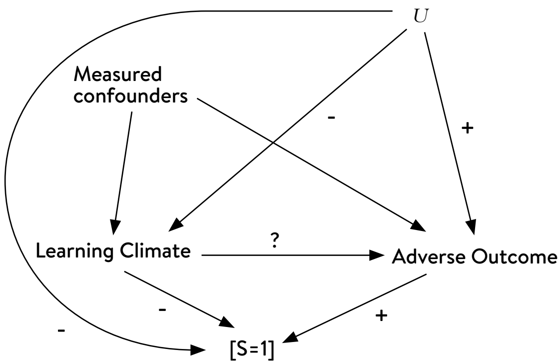


Figure A2.6. Directed acyclic graph (DAG) depicting selection or response (S = 1) bias caused by the adverse outcome (positively), the learning climate (negatively) and an unmeasured confounder U (negatively). Here, U decreases study response (S = 1) and learning climate but increases adverse outcome. [Arrow from measured confounders only omitted for visual reasons but can be assumed without loss of generality.]

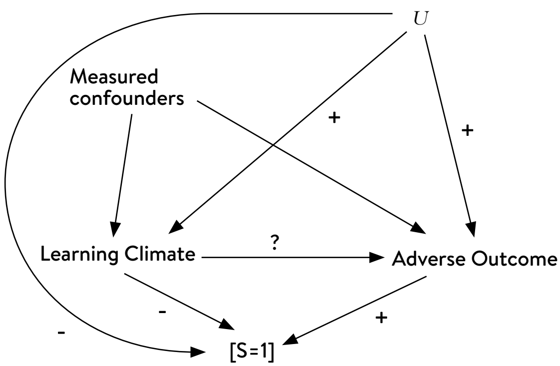


Figure A2.7. Directed acyclic graph (DAG) depicting selection or response (S = 1) bias caused by the adverse outcome (positively), the learning climate (negatively) and an unmeasured confounder U (negatively). Here, U decreases study response (S = 1) but increases learning climate and adverse outcome. [Arrow from measured confounders only omitted for visual reasons but can be assumed without loss of generality.]

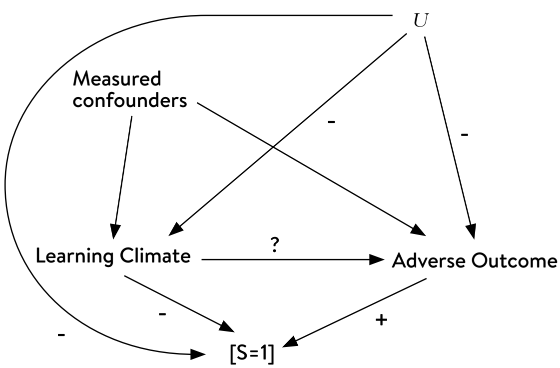


Figure A2.8. Directed acyclic graph (DAG) depicting selection or response (S = 1) bias caused by the adverse outcome (positively), the learning climate (negatively) and an unmeasured confounder U (negatively). Here, U decreases study response (S = 1), learning climate, and adverse outcome. [Arrow from measured confounders only omitted for visual reasons but can be assumed without loss of generality.]

Table A2.2 Multiple-bias adjusted odds ratios (95% confidence intervals) for the relation between learning climate (LC) and adverse obstetric outcome (Y), after bias adjustment for uncontrolled confounding due to unmeasured confounder U and for selection bias (selective response)*, From Study on Learning Climate and Adverse Obstetrical Outcomes, 2013

	Odds ratio relating unmeasured confounder *U* to adverse outcome *Y*, conditional on measured confounders			
	Negative relation (-)		Positive relation (+)	
Regression coefficient relating U to LC	**0.25**	**0.50**	**2**	**4**
Positive (+)	**Figure A2.5**		**Figure A2.7**	
0.10	5.06 (1.81 – 14.13)	3.58 (1.34 – 9.57)	1.30 (0.56 – 3.02)	0.81 (0.38 – 1.74)
0.25	10.49 (4.14 – 26.65)	5.57 (2.29 – 13.60)	0.90 (0.42 – 1.96)	0.38 (0.19 – 0.78)
0.50	11.51 (5.16 – 25.74)	5.75 (2.68 – 12.35)	0.94 (0.47 – 1.86)	0.40 (0.21 – 0.77)
Negative (-)	**Figure A2.8**		**Figure A2.6**	
-0.10	0.85 (0.31 – 2.37)	1.21 (0.45 – 3.24)	3.46 (1.49 – 8.01)	5.80 (2.71 – 12.46)
-0.25	0.42 (0.17 – 1.08)	0.80 (0.33 – 1.95)	5.08 (2.34 – 11.05)	12.45 (6.08 – 25.55)
-0.50	0.40 (0.18 – 0.90)	0.80 (0.37 – 1.73)	5.03 (2.53 – 10.03)	12.15 (6.70 – 23.51)

*Prevalence of U is set at P(U = 1) = 0.4 for the probabilistic bias analysis. Similar result patterns are obtained at different U prevalence values or distributions. The odds ratios relating learning climate, unmeasured confounder set U, and adverse outcome, adjusted for measured confounders, to study response or participation were respectively set to 0.75, 0.75 and 5.

References Appendix:

1. Pearl J. Causality: models, reasoning and inference. 2nd Edition. New York: Cambridge University Press, 2009.
2. Pearl J, Glymour M, Newell NP. Causal inference in statistics: a primer. Hoboken, NJ: Wiley, 2016.
3. Helmich E, Boerebach BC, Arah OA, Lingard L. Beyond limitations: Improving how we handle uncertainty in health professions education research. *Med Teach*. 2015;37(11):1043-50.
4. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48.
5. Boerebach BC, Lombarts KM, Scherpbier AJ, Arah OA. The teacher, the physician and the person: exploring causal connections between teaching performance and role model types using directed acyclic graphs. *PLoS One*. 2013;8(7):e69449.
6. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol*. 2008;18(8):637-46.
7. VanderWeele TJ, Arah OA. Bias expressions for external adjustment and sensitivity analysis for general outcomes, treatments, and unmeasured confounding variables. *Epidemiology.* 2011; 22(1):42-52.
8. Rothman KJ, Lash TL, Greenland S. Modern Epidemiology. 3rd Edition. Philadelphia: Lippincott Williams & Wilkins, 2008.
9. Arah OA, Sudan M, Olsen J, Kheifets L. Marginal structural models, doubly robust estimation, and bias analysis in perinatal and paediatric epidemiology. *Paediatr Perinat Epidemiol*. 2013;27(3):263-5.
10. Arah OA. Bias analysis for uncontrolled confounding in the health sciences. *Annu Rev Public Health*. 2017;38.
11. Thompson CA, Arah OA. Selection bias modeling using observed data augmented with imputed record-level probabilities. *Ann Epidemiol*. 2014;24(10):747-753.

CHAPTER 4

Closing the patient experience chasm: a two-level validation of the Consumer Quality Index Inpatient Hospital Care

A. Smirnova
K.M.J.M.H. Lombarts
O.A. Arah
C.P.M. van der Vleuten

Health Expectations, 2017; 00:1-8

Abstract

Background
Evaluation of patients' health care experiences is central to measuring patient-centred care. However, different instruments tend to be used at the hospital or departmental level but rarely both, leading to a lack of standardization of patient experience measures.

Objective
To validate the Consumer Quality Index (CQI) Inpatient Hospital Care for use on both department and hospital levels.

Design
Using cross-sectional observational data, we investigated the internal validity of the questionnaire using confirmatory factor analyses (CFA), and the generalizability of the questionnaire for use at the department and hospital levels using generalizability theory.

Setting and participants
22924 adults hospitalized for ≥24 hours between 1 January 2013 and 31 December 2014 in 23 Dutch hospitals (515 department evaluations).

Main variable
CQI Inpatient Hospital Care questionnaire.

Results
CFA results showed a good fit on individual level (CFI = 0.96, TLI = 0.95, RMSEA = 0.04), which was comparable between specialties. When scores were aggregated to the department level, the fit was less desirable (CFI = 0.83, TLI = 0.81, RMSEA = 0.06), and there was a significant overlap between communication with doctors and explanation of treatment subscales. Departments and hospitals explained ≤5% of total variance in subscale scores. In total, 4-8 departments and 50 respondents per department are needed to reliably evaluate subscales rated on a 4-point scale, and 10 departments with 100-150 respondents per department for binary subscales.

Discussion and conclusions
The CQI Inpatient Hospital Care is a valid and reliable questionnaire to evaluate inpatient experiences in Dutch hospitals provided sufficient sampling is done. Results can facilitate meaningful comparisons and guide quality improvement activities in individual departments and hospitals.

Introduction

Evaluation of patients' health care experiences has become central to measuring quality in health care and, as a result, health care providers are more often held responsible for monitoring and improving patients' care experiences.[1] Patient care experiences reflect the degree to which care is patient-centred (i.e. care that is respectful and responsive to patients' preferences, needs and values).[2] In addition to its intrinsic value as an indicator of quality, a growing body of evidence points to the positive associations between positive patient experiences and clinical processes of care[3,4] as well as better patient adherence to treatment, improved clinical outcomes and decreased utilization of health care services.[5]

Even though improving patient care experiences is increasingly being incorporated in both local and global health agendas,[6] patient feedback remains largely underutilized in local hospital improvement plans.[7] One of the main reasons for this is lack of specific and timely feedback that is easily translatable to improvements on the frontline.[8,9] Current instruments used to collect patient experience data mostly collect data on hospital-wide level for identification of larger national trends and contracting of hospital services. In order to bridge the gap between external reporting and internal quality assurance, some have recommended to use different instruments for different purposes.[9,10] This is, however, not desirable due to lack of standardization of measures, a lack of common language and possible disconnect between local improvement efforts and hospital-wide measurements. Implementation of instruments is also costly and can potentially lead to duplication of work and unnecessary use of valuable resources.

An alternative approach is to adapt existing instruments to reflect their multiple purposes. In this study, we attempt to address these problems using the Dutch version of the American Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, which was imported into the Netherlands in 2006 by Arah et al. for use within the Dutch health care system.[11] This has led to the development of nationally used standardized questionnaires and protocols called the Consumer Quality Index (CQI), wherein the Dutch HCAHPS is known as the CQI Inpatient Hospital Care.[12] Efforts to adapt this questionnaire for multiple purposes, including external accountability and internal quality assurance, have resulted in different versions of the questionnaire to be produced.[13,14] However, no extensive validation of the CQI Inpatient Hospital Care has occurred since the original validation study by Arah et al.[11] As the results are consequentially used by patients, hospital staff, health insurers, the inspectorate and researchers for different purposes, it is imperative that the questionnaire can evaluate and differentiate patient care experiences across hospitals, specialties as well as departments reliably and validly. We aimed to assess the internal validity and reliability of the CQI Inpatient Hospital Care on both hospital and department levels. Additionally, we investigated whether the questionnaire measured similar domains of patient experiences across four specialties, namely surgery, obstetrics and gynecology, internal medicine and cardiology.

Methods

Setting and study population

We analyzed CQI Inpatient Hospital Care questionnaire data from 23 Dutch hospitals including four academic centres, 515 department evaluations in 17 specialties (nine surgical and eight medical) collected between 1 January 2013 and 31 December 2014. Eligible patients 16 years or older who were hospitalized for at least 24 hours with a discharge within the previous 12 months were identified using hospital admission lists. Eligible participants were invited to evaluate their experiences of hospitalization using either online or paper-based CQI Inpatient Hospital Care (Appendix, Table A1). Evaluations collected in 2013 were used for national benchmarking among 43 hospitals in four specialties, namely surgery, internal medicine, cardiology, obstetrics and gynecology. Therefore, we focused on these specialties in this study. The hospitals and clinical departments that re-evaluated their inpatient hospital care using the same questionnaire in 2014 for own internal quality assurance purposes were considered to be independent evaluations and were, therefore, also included in the analysis. We analyzed the results both for 2013 and 2014 together and separately, and if there was no change, reported the combined results only.

As retrospective research does not fall under the Dutch Medical Research Involving Human Subjects Act (WMO), an official ethical review was not required for this study. Nonetheless, we obtained permissions from individual hospitals to use anonymized questionnaire data for research purposes. Furthermore, we consulted a privacy officer at our institution to ensure that the data provided for this research complied with Dutch Personal Data Protection Act. Participating hospitals were recruited through the Miletus Foundation (www.stichtingmiletus.nl), a coordinating body of all CQI evaluations within the Netherlands. A detailed research proposal was sent to all hospitals and subsequently discussed at the general meeting. Hospitals interested in participating in the study gave informed consent either via the Miletus Foundation or by directly contacting the primary researcher (AS). MediQuest (home.mediquest.nl), a company that processes patient evaluation data from these evaluations, provided the final data set for the study.

CQI Inpatient Hospital Care questionnaire

The CQI Inpatient Hospital Care questionnaire has been developed in co-operation with patient and consumer organizations based on three existing instruments used to measure patient care experiences: the CAHPS Hospital Care questionnaire, the Dutch Hospital Association inpatient satisfaction questionnaire and the Hospital Comparison questionnaire from the Netherlands Institute for Health Services Research and the Consumers' Association.[13,15] The CQI Inpatient Hospital Care consists of a total of 50 items: 38 items about patient experiences and 12 items asking background information. An earlier exploratory factor analysis[14] identified nine domains of patient experience, namely admission (Q4a-j), communication with nurses (Q6-8), communication with doctors (Q9-10), own contribution (Q13-15, 17, 25), explanation of treatment (Q18-20), pain management (Q21-22), communication about medication (Q23-24), feeling of safety (Q27-29) and discharge information (Q30-34). Admission and information at discharge were assessed on a 2-point scale (yes = 1, no = 0). Other scales were assessed on a 4-point Likert scale ranging from 1 (Never) to 4 (Always). Building on this previously identified work, we used this structure to test the internal validity, reliability and generalizability of the CQI Inpatient Hospital Care questionnaire.

Statistical analysis

First, respondents and non-respondents were described using descriptive statistics. Questionnaires were excluded if they had a negative or no response to the question whether or not the patient had a hospital admission within the last 12 months or if less than half of core items were completed. Evaluations with missing data were imputed using multiple imputation technique to create 10 complete data sets.[16] Multiple imputation was preferable to single-imputation methods such as maximum-likelihood approaches because it better reflected the inherent uncertainty due to missing data in the sample.[17] Convergence of the imputations was assessed by examining trace plots and calculating the Rhat statistic.[18] In order to maximize convergence, we increased the number of maximum iterations to 200. We then calculated the subscale scores for each imputed data set by averaging the scores for the items within each subscale.

The internal validity of the questionnaire was evaluated by assessing the fit of the pre-identified 9-factor structure of the questionnaire. In order to assess the overall fit of the model, we performed a confirmatory factor analysis (CFA) on all imputed data sets and combined the final results using Rubin's rules. For categorical variables, weighted least squares with mean and variance adjusted (WLSMV) estimator was preferred to account for the categorical nature of the answers. The WLSMV estimator is a robust estimator that does not assume normally distributed variables and is preferred for modelling categorical or ordered data.[19] We assessed the global model fit using the comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA).[20] The following cut-off values indicated a good fit: CFI ≥0.95, TLI ≥0.95 and RMSEA ≤0.06.[19] The overall fit was deemed acceptable if at least two of the three criteria of fit indices were met.[21] In order to establish whether the questionnaire measured similar patient experiences across various medical specialties, CFA was then repeated in four subgroups: surgery, obstetrics and gynecology, internal medicine and cardiology. These specialties were chosen because these specialties were included in the national benchmark. Same cut-off points were used to evaluate the fit of the factor structure as for the overall sample. Finally, we repeated the CFA on the department level by aggregating the scores of each variable to the department level.

Internal consistency of the subscales was evaluated by calculating Cronbach's α statistic for individual questionnaires and the department in each imputed data set, and averaging it across imputed data sets. Overall Cronbach's α ≥0.70 was deemed acceptable. The degree to which the subscales measured distinct concepts was assessed by calculating inter-scale correlations, which were also calculated for individual scores and scores aggregated to the department. A correlation of <0.70 indicated that there was no significant overlap between the subscales. Construct validity was assessed by examining the relative importance of the subscales with two global ratings, namely overall evaluation of the department (Q36, scale 0-10) and hospital (Q35, scale 0-10) using multiple linear regression and accounting for respondents' age, sex, education, self-rated physical health, self-rated psychological health, country of origin and the number of admissions in the previous 12 months.

Generalizability analysis was conducted to estimate the minimum number of respondents needed to reliably evaluate each subscale on both department and hospital levels. For department-level evaluations, we estimated a model where the number of items was considered as fixed, with department (*d*) as the unit of analysis, where respondents (*p*) were nested within departments (*p:d*). The resulting design was unbalanced single-facet nested design.[22] For hospital-level analyses, we similarly regarded the number of items as fixed; however, this time we regarded hospital as the unit of analysis and respondents (*p*) to be nested within departments (*d*), which were, in turn, nested within hospitals (*h*), resulting in an multifacet unbalanced nested design (*p:d:h*). We averaged variance components, including variance across the departments (*Sd*) and respondents nested within departments (*Sp:d*) and respondents nested within departments and hospitals (*Sp:d:h*), across imputed data sets. Then, we estimated the proportion of the total variance in scores that are due to differences between departments or hospitals. In a D-study, we estimated the G coefficient and the standard error of measurement (SEM) associated with varying number of respondents within departments and departments within hospitals for mean subscale scores. For seven scales evaluated on a 4-point scale, we used 0.4 units as an admissible level of "noise," representing SEM <0.10 ($1.96 \times 0.10 \times 2 \approx 0.4$) as the maximum value for 95% confidence interval interpretation. For dichotomous scales, we used 0.1 on a scale of 0-1 as an admissible level of noise, representing SEM <0.025 ($1.96 \times 0.025 \times 2 \approx 0.1$).

Missing data were imputed using the mice package (version 2.25) in R statistical software version 3.2.3.[23,24] The confirmatory factor analyses on imputed data sets were performed using the semTools package (version 0.4-11) and on aggregated data sets using the lavaan package (version 0.5-20) in R version 3.2.3.[25] Inter-scale correlations, Cronbach's α, variance components calculations and multiple linear regression analyses were performed using SPSS version 23.0.0.2 (IBM Corp., Armonk, New York).

Results

Of the distributed 74090 questionnaires, 23476 were returned (gross response rate 31.7%). Table 1 reports characteristics of respondents and non-respondents. In total, 552 questionnaires were excluded due to negative or no response to the question whether or not they had a hospital admission within the last 12 months or less than half of core items completed. The resulting sample size was 22924 (net response rate 30.9%), including 23 hospitals, 17 different specialties and 515 department evaluations. Table 2 further describes the demographic characteristics of the included respondents. As the results did not differ between 2013 and 2014, we report only combined results below.

Table 1. Characteristics of respondents and non-respondents of the CQI Inpatient Hospital Care questionnaire

Characteristic	Respondents N (%) (n=23476)	Non-respondents N (%) (n=50614)	Total N (%) (n=74090)
Gender			
Male	11255 (47.9)	21802 (43.1)	33057 (44.6)
Female	12221 (52.1)	28812 (56.9)	41033 (55.4)
Age (years)			
16-24	486 (2.1)	3623 (7.2)	4109 (5.5)
25-34	1580 (6.7)	6999 (13.8)	8579 (11.6)
35-44	1833 (7.8)	6356 (12.6)	8189 (11.1)
45-54	3062 (13.0)	7246 (14.3)	10308 (13.9)
55-64	5195 (22.1)	8224 (16.2)	13419 (18.1)
65-74	5492 (23.4)	9720 (19.2)	15212 (20.5)
75-79	2737 (11.7)	3088 (6.1)	5825 (7.9)
80+	3091 (13.2)	5358 (10.6)	8449 (11.4)
Type of questionnaire			
Online	17922 (76.3)	-	-
Mail	5554 (23.7)	-	-

Table 2. Characteristics of the respondents included in validation of the CQI Inpatient Hospital Care questionnaire

Characteristic	N (Total=22924)	(%)
Gender		
Male	10992	(47.9)
Female	11932	(52.1)
Age (years)		
16-24	486	(2.1)
25-34	1572	(6.9)
35-44	1828	(8.0)
45-54	3053	(13.3)
55-64	5170	(22.6)
65-74	5462	(23.8)
75-79	2535	(11.1)
80+	2818	(12.3)
Level of education		
Lower Secondary or less	6561	(28.6)
Upper Secondary	10511	(45.9)
Tertiary	5852	(25.5)
Self-reported health		
Excellent	1389	(6.1)
Very good	2962	(21.9)
Good	10673	(46.6)
Average	6694	(29.2)
Bad	1206	(5.3)
Self-reported psychological health		
Excellent	4149	(18.1)
Very good	5460	(23.8)
Good	10968	(47.8)
Average	2130	(9.3)
Bad	217	(0.9)
Country of origin		
The Netherlands	21152	(92.3)
Germany	156	(0.7)
(Former) Netherlands Antilles/Aruba/Suriname	293	(1.3)
Indonesia/Netherlands Indies	281	(1.2)
Morocco/Turkey	194	(0.8)
Other	738	(3.2)
Missing	110	(0.5)
Number of admissions in the previous 12 months including current one		
1	13283	(57.9)
2	5947	(25.9)
3	2119	(9.2)
4+	1464	(6.4)
Missing	111	(0.5)
Specialty		
Surgical:	11344	(49.5)
General surgery	3225	(14.1)
Orthopedic surgery	2502	(10.9)

Characteristic	N (Total=22924)	(%)
Urology	1773	(7.7)
Cardiothoracic surgery	895	(3.9)
Neurosurgery	822	(3.6)
Otolaryngology	743	(3.2)
Obstetrics and gynecology	643	(2.8)
Plastic surgery	607	(2.6)
Ophthalmology	134	(0.6)
Medical:	8000	(34.9)
Cardiology	2697	(11.8)
Internal medicine	1984	(8.7)
Pulmonology	1877	(8.2)
Neurology	1262	(5.5)
Rheumatology	67	(0.3)
Geriatrics	54	(0.2)
Dermatology	38	(0.2)
Anesthesiology	21	(0.1)
Missing	3580	(15.6)

Psychometric properties

CFA showed a good fit for surgical, obstetrics and gynecology, internal medicine, cardiology specialties and all specialties combined (Table 3). When the scores were aggregated to the department level, the incremental fit indices decreased to CFI = 0.83 and TLI = 0.81. Internal consistency of the scales was acceptable, except for subscales own contribution (0.69), communication about medication (0.68) and feeling of safety (0.64). On the department level, all subscales demonstrated acceptable Cronbach's α, except for feeling of safety (0.64) (Table 4). Inter-scale correlations showed that on the department level, the subscales communication with doctors with explanation of treatment overlapped substantially (Pearson's $r = 0.72$) (Table 4). Communication of treatment did not predict global ratings of either the hospital or the department, while explanation of treatment was a significant predictor of the rating of the hospital but not the global rating of the department (Table 4).

Five percent or less of total variance in scores was attributable to the department or the hospital (Table 5). Results of the generalizability analysis showed that a minimum of 50 respondents is needed to reliably evaluate subscales of patient experience scored 1-4 in a department (Appendix, Table A2). For subscales evaluated on Yes/No (0-1) scale (admission and discharge information), 100 and 150 patient evaluations were needed, respectively, for department-level evaluations. For hospital-level evaluations, subscales rated 1-4 can reliably be evaluated with 4-8 departments with at least 50 patient evaluations each. For admission and discharge information, at least 10 departments with 100 patient evaluations are needed.

Table 3. Fit indices for surgery, cardiology, internal medicine, and obstetrics and gynecology, and all specialties on individual (patient) level and department level. Department level scores were obtained by calculating the means for every item per department across all imputed datasets

	Surgery (n=3225) Individual level	Cardiology (n=2697) Individual level	Internal medicine (n=1984) Individual level	Obstetrics & gynecology (n=643) Individual level	All specialties (n=22924) Individual level	All specialties (n=515) Department level
CFI (≥0.95)	0.98	0.97	0.98	0.98	0.96	0.83
TLI (≥0.95)	0.98	0.96	0.98	0.97	0.95	0.81
RMSEA (≤0.06)	0.03	0.03	0.03	0.02	0.04	0.06

Table 4. Scale means with standard deviations (SD), reliability coefficients (Cronbach's α) and inter-scale correlations, for individual (above the diagonal) and aggregated department (below the diagonal) evaluations, and estimates of multiple linear regression analyses with 95% confidence intervals examining associations with global department and hospital ratings corrected for respondents' age, sex, level of education, self-rated physical and psychological health, number of admissions in the previous 12 months, and place of birth (*P < 0.05, ** P < 0.001)

Subscale (scoring range)	Mean (SD)	Cronbach's α individual (department)	1	2	3	4	5	6	7	8	9	Global rating department	Global rating hospital
1. Admission (0-1)	0.6 (0.25)	0.77 (0.81)	1	0.30	0.29	0.31	0.39	0.25	0.38	0.35	0.42	0.14 (0.05-0.23)*	0.20 (0.12-0.27)**
2. Communication with nurses (1-4)	3.4 (0.61)	0.83 (0.87)	0.36	1	0.56	0.49	0.51	0.55	0.47	0.46	0.35	1.00 (0.97-1.04)**	0.59 (0.56-0.62)**
3. Communication with doctors (1-4)	3.4 (0.71)	0.81 (0.84)	0.41	0.56	1	0.42	0.56	0.41	0.43	0.37	0.33	0.08 (0.05-0.11)**	0.27 (0.25-0.30)**
4. Own contribution (1-4)	3.0 (0.66)	0.69 (0.80)	0.31	0.50	0.47	1	0.44	0.38	0.46	0.39	0.31	0.31 (0.28-0.34)**	0.27 (0.25-0.30)**
5. Explanation of treatment (1-4)	3.5 (0.67)	0.81 (0.89)	0.50	0.59	0.72	0.52	1	0.45	0.57	0.43	0.43	-0.03 (-0.06-0.00)	0.08 (0.05-0.11)**
6. Pain management (1-4)	3.5 (0.62)	0.79 (0.86)	0.48	0.68	0.52	0.42	0.61	1	0.42	0.41	0.32	0.34 (0.31-0.38)**	0.26 (0.22-0.29)**
7. Communication about medication (1-4)	3.0 (0.91)	0.68 (0.85)	0.41	0.60	0.60	0.55	0.67	0.52	1	0.47	0.45	-0.01 (-0.03-0.02)	0.004 (-0.02-0.03)
8. Feeling of safety (1-4)	3.4 (0.68)	0.64 (0.64)	0.47	0.47	0.48	0.30	0.50	0.61	0.52	1	0.38	0.21 (0.18-0.24)**	0.18 (0.15-0.21)**
9. Information at discharge (0-1)	0.7 (0.31)	0.76 (0.82)	0.60	0.49	0.50	0.43	0.60	0.50	0.53	0.42	1	0.54 (0.48-0.60)**	0.52 (0.47-0.57)**

Table 5. Variance components for departments, hospitals and residual variance

	Residual variance	Between-department variance (total variance)	Between-hospital variance (total variance)	Hospital variance vs. hospital and department variance
1. Admission	0.059	0.003 (5)	0.000 (0)	0.0
2. Communication with nurses	0.360	0.005 (1)	0.004 (1)	0.44
3. Communication with doctors	0.490	0.006 (1)	0.004 (1)	0.40
4. Own contribution	0.404	0.014 (3)	0.020 (5)	0.59
5. Explanation of treatment	0.435	0.012 (3)	0.003 (1)	0.20
6. Pain management	0.376	0.008 (2)	0.002 (1)	0.20
7. Communication about medication	0.805	0.012 (1)	0.008 (1)	0.40
8. Feeling of safety	0.446	0.010 (2)	0.002 (0)	0.17
9. Information at discharge	0.089	0.005 (5)	0.000 (0)	0.0

Discussion

To our knowledge, this is the first study to validate an inpatient experience questionnaire for multiple purposes, namely on the level of the hospital and the department. The CFA results showed a good overall fit, which was comparable between specialties. On the department level, however, the CFA showed a less desirable fit with a significant overlap on the department level between the subscales communication with doctors and explanation of treatment. Differences between departments and hospitals explained only a small proportion of total variance in patient experience scores, with the hospital and the department varying in importance depending on the subscales. A total of 4-8 departments and 50 respondents per department are needed to reliably evaluate most subscales on both department and hospital levels. For binary subscales, such as admission and discharge information, a minimum of 100-150 patients per department and 10 departments are needed.

The overall good fit provides evidence of validity for the internal structure of the CQI Inpatient Hospital Care questionnaire on the level that it was first designed for, that is the patient. The goodness-of-fit indices for surgery, obstetrics and gynecology, cardiology and internal medicine specialties were similarly good, suggesting that patients experience similar aspects of care in different specialties, allowing for comparisons of patient experiences between specialties to be made. Previous research has demonstrated that, even though aspects of patient experience may be comparable across specialties, their importance can differ substantially by type of hospitalization.[26] Although we did not research the relative importance of these aspects for different specialties, departments or hospitals will need to take this into account when choosing priorities for areas of quality improvement.

The internal consistency of the scales was acceptable except for three subscales: own contribution, communication about medication and feeling of safety. The same subscales also demonstrated a lower internal consistency in a previous pilot validation study.[14] Furthermore, our study found that the subscale communication about medication did not significantly contribute to the global ratings of the department or the hospital, which may indicate a need for improvement in external validity of this scale. Alternatively, global ratings may not be a good indicator of overall health care quality and should, therefore, not be used in external validation, as research by Krol et al. has shown it may be measuring a different concept.[27] Similar to other studies[11,26,28], we found that communication with nurses was the strongest predictor of overall ratings of the department as well as the hospital. This is not surprising as nurses are the primary providers of care in the hospital environment. Furthermore, research has shown that factors related to nursing work such as nursing work environment, nurse-to-patient ratios[28] and missed nursing care[29] and nurse-patient interaction[30] can influence patient satisfaction ratings. A new finding, however, is that higher scores on the subscale discharge information significantly contributed to patients' global ratings of both the hospital and the department. This is different from the findings by Elliott et al.[26], in which discharge information was one of the least valued aspects of in-patient care and was important for only half of hospitalization types. This is not surprising as there appears to be a gap in communication between patients and providers at discharge. A survey of hospitalized patients showed that more than half of patients 70 years or older did not receive instructions about how to care for themselves after hospitalization.[31] Our findings suggest that discharge information may be more important than previously thought and that hospitals and departments may improve the overall patient experience by improving how they handle discharges. Yet, as De Boer et al. demonstrated, although global ratings represent experiences regarding priorities, experiences with the important elements of care may still have inconsistent relationships with global ratings.[32]

On the department level, fit indices did not demonstrate an acceptable fit based on the incremental fit indices (CFI = 0.83 and TLI = 0.81), while RMSEA was within acceptable bounds at 0.06 (≤0.06 acceptable). As two of the three criteria do not meet the cut-off criteria, we conclude that the current model is not a good representation of the latent constructs on the department level. Combined with the significant overlap between subscales explanation of treatment and communication with doctors, these results point that on the department level a different structure would provide a better fit of the data. Another reason for a poor fit of the structure on the department level could be the use of aggregated scores, which does not consider the variability of the scores within each department. This may have unnecessarily distorted the data. As the patients are naturally nested within departments and hospitals, confirmation of the fit using multilevel CFA is desirable.

The results of the variance component analysis showed that the department and the hospital each account for 5% or less in total variability of the subscale scores. This corresponds with previous research that has found limited influence of the department and the hospital on variability of patient experience scores.[15,33] Generalizability analysis found that it is possible to reliably evaluate patients' experience using subscales with the scoring scale 1-4 with 50 respondents (in 4-8 departments for hospital-level evaluations), and with 100-150 respondents (in 10 departments) for the two subscales with the Yes/No (0-1) scoring scale. More respondents are needed for binary subscales because of the small range of possible scores, leading to higher precision and reliability needed to detect small changes. Compared with other instruments,[5,10] this study shows an improvement in the number of respondents that are needed for reliable evaluation of patient experiences of a single department. Similar size samples are required to reliably evaluate all subscales on the hospital level using our criteria. However, different cut-off criteria may be chosen depending on whether the results of the CQI Inpatient Hospital Care are to be used by departments for their own quality improvement purposes, or by health insurance companies and health-care authorities to make summative judgements about the quality of care.[34] We, therefore, recommend using the generalizability results of this study (shown in Appendix, Table A2) to adjust the cut-off criteria based on the proposed use of the questionnaire.

Limitations

In interpreting the results, several limitations should be mentioned. Patient surveys suffer from low response rates. Our response rate of 31% was similar to those previously seen in this setting.[14] Reasons for non-response were not collected during the original data collection process, which made a non-responder analysis impossible. Although we tried to account for non-respondents by including sex and age as covariate in regression analyses, this may not have been sufficient because respondents and non-respondents may also vary based on other characteristics that we have not been able to account for, such as country of origin, language spoken at home or level of education. For example, we did not have any data on how many patients were invited to fill out online or paper-based questionnaire. Furthermore, in this study we aggregated the individual scores to the level of the department, because this is how typically the scores may be used. Other methods can be tried, such as using median or factor scores, but these may be difficult to interpret. Also, we did not test alternative models on the department level or factor equivalency between different specialties or respondents groups. Finally, we did not investigate the external validity of this questionnaire by studying the relationship between aspects of inpatient hospital care and other important process or outcome measures. Nonetheless, this study also has several strengths. One strength of this study is its use of more than 22000 patient evaluations and over 500 department evaluations from multiple specialties in multiple hospitals including academic and non-academic centres, which supports the generalizability of our results. Another strength in this study is the use of multiple imputation for handing missing data, which accounts for the uncertainty associated with imputation of missing data.[17]

Implications

With this study, we contribute evidence for validity of the CQI Inpatient Hospital Care questionnaire and its utility for use in different settings and for both quality assurance and summative purposes. We recommend that stakeholders including hospitals, clinical departments and health insurers using this questionnaire use appropriate sample sizes based on its purpose and level of use. Considering the response rate is 31%, much larger samples may be required to arrive at recommended numbers of evaluations. Low response rates have become worrisomely common in survey research[35], with many studies now reporting rates as low as or lower than ours.[36] Low response rates may indicate low levels of receptivity of the instrument by patients. Improvements in response rates, for example by identifying and addressing reasons for non-response, are needed to ensure optimal use of resources as well as appropriate sample sizes. Although this questionnaire has originally been imported to facilitate standardization of the instrument for international comparisons[11], at this point, both the CQI Inpatient Hospital Care and the American HCAHPS, on which the CQI Inpatient Hospital Care is based, have changed substantially such that any international comparisons can only be made based on a collection of limited number of questions that are present in both questionnaires. Future research can investigate whether patient experiences of hospital care improve over time with continuous measurement. Like Zuidgeest et al.[37] and Damman et al.[38], we recommend using multilevel models for longitudinal and hierarchical data analyses, rather than using average department or hospital scores.

Conclusion

In conclusion, the CQI Inpatient Hospital Care questionnaire can provide valid and reliable data on patient experiences of inpatient hospital care on both department and hospital. The resulting data can be used to facilitate meaningful comparisons and guide quality improvement activities. Future research can focus on improving reliability of the scales, wording of the individual items to reflect specific provider or clinical settings better, and validating the structure on the department level and for different specialties.

References:

1. Lazar EJ, Fleischut P, Regan BK. Quality measurement in healthcare. *Annu Rev Med*. 2013;64:485-496.
2. Crossing the quality chasm: a new health system for the 21st century. The National Academies Press; 2001.
3. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perception of hospital care in the United States. *N Engl J Med*. 2008;359(18):1921-1931.
4. Arah OA, Roset B, Delnoij DM, Klazinga NS, Stronks K. Associations between technical quality of diabetes care and patient experience. *Health Expect*. 2011;16(4):e136-145.
5. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev*. 2014;71(5):522-554.
6. Scott KW, Jha AK. Putting quality on the global health agenda. *N Engl J Med*. 2014;371(1):3-5.
7. Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. *BMJ*. 2014;348:g2225.
8. Flott KM, Graham C, Darzi A, Mayer E. Can we use patient-reported feedback to drive change? The challenges of using patient-reported feedback and how they might be addressed. *BMJ Qual Saf*. 2016(0):1-6.
9. Rozenblum R, Lisby M, Hockey PM, et al. The patient satisfaction chasm: the gap between hospital management and frontline clinicians. *BMJ Qual Saf*. 2013;22(3):242-250.
10. Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst Rev*. 2015;4:97.
11. Arah OA, ten Asbroek AH, Delnoij DM, et al. Psychometric properties of the Dutch version of the Hospital-level Consumer Assessment of Health Plans Survey instrument. *Health Serv Res*. 2006;41(1):284-301.
12. The CQ-index. https://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/de+cq-index. Accessed June 21, 2016.
13. Sixma H, Zuidgeest M, Rademakers J. CQ-index Hospitalization: instrument development (in Dutch). Utrecht: Netherlands Institute for Health Services Research (NIVEL); 2009.
14. Batterink M. Analysis Report: further development of the CQI Hospitalization 2011 (in Dutch). Barnveld: Significant;2011.
15. Delnoij DM, Rademakers JJ, Groenewegen PP. The Dutch consumer quality index: an example of stakeholder involvement in indicator development. *BMC Health Serv Res*. 2010;10:88.
16. Stuart EA, Azur M, Frangakis C, Leaf P. Multiple imputation with large data sets: a case study of the Children's Mental Health Initiative. *Am. J Epidemiol*. 2009;169(9):1133-1139.
17. Dong Y, Peng CY. Principled missing data methods for researchers. *Springerplus*. 2013;2(1):222.
18. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press; 2007.
19. Brown TA. Confirmatory factor analysis for applied research. New York: Guilford; 2006.
20. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Res*. 2006;99(6):323-338.
21. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1-55.
22. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34(11):960-992.
23. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):67.
24. R Core Team. R: A language and environment for statistical computing. 2015; https://www.R-project.org/.
25. semTools Contributors. semTools: Useful tools for structural equation modeling. R package version 0.4-11. http://cran.r-project.org/package=semTools.
26. Elliott MN, Kanouse DE, Edwards CA, Hilborne LH. Components of care vary in importance for overall patient-reported experience by type of hospitalization. *Med Care*. 2009;47(8):842-849.
27. Krol MW, de Boer D, Rademakers JJ, Delnoij DM. Overall scores as an alternative to global ratings in patient experience surveys; a comparison of four methods. *BMC Health Serv Res*. 2013;13:479.

28. Kutney-Lee A, McHugh MD, Sloane DM, et al. Nursing: a key to patient satisfaction. *Health Aff (Millwood)*. 2009;28(4):w669-677.
29. Lake ET, Germack HD, Viscardi MK. Missed nursing care is linked to patient satisfaction: a cross-sectional study of US hospitals. *BMJ Qual Saf*. 2015(0):1-9.
30. Aiello A, Garman A, Morris SB. Patient satisfaction with nursing care: a multilevel analysis. *Qual Manag Health Care*. 2003;12(3):187-190.
31. Flacker J, Park W, Sims A. Hospital discharge information and older patients: do they get what they need? *J Hosp Med*. 2007;2(5):291-296.
32. de Boer D, Delnoij D, Rademakers J. Do patient experiences on priority aspects of health care predict their global rating of quality of care? A study in five patient groups. *Health Expect*. 2010;13(3):285-297.
33. Krol MW, De Boer D, Sixma H, Van Der Hoek L, Rademakers JJ, Delnoij DM. Patient experiences of inpatient hospital care: a department matter and a hospital matter. *Int J Qual Health Care*. 2015;27(1):17-25.
34. Crossley J, Russell J, Jolly B, et al. 'I'm pickin' up good regressions': the governance of generalisability analyses. *Med Educ*. 2007;41(10):926-934.
35. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol*. 2007;17(9):643-653.
36. Baird M, Daugherty L, Kumar KB, Arifkhanova A. Regional and gender differences and trends in the anesthesiologist workforce. *Anesthesiology*. 2015;123(5):997-1012.
37. Zuidgeest M, Delnoij DM, Luijkx KG, de Boer D, Westert GP. Patients' experiences of the quality of long-term care among the elderly: comparing scores over time. *BMC Health Serv Res*. 2012;12:26.
38. Damman OC, Stubbe JH, Hendriks M, et al. Using multilevel modeling to assess case-mix adjusters in consumer experience surveys in health care. *Med Care*. 2009;47(4):496-503.

Appendix

Table A1. Original domains and associated items in the CQI Inpatient Hospital Care questionnaire (version 1.3)

Domain	Item No.	Description	Response categories
Hospital accessibility	2	Has the accessibility of the hospital using your own means of transport been a problem?	Major problem; Minor problem; No problem; Not applicable
	3	Have the number of parking spaces at the hospital been a problem?	Ditto
Admission		Were the following items discussed with you on your admission to the hospital?	
	4a	Your rights as patient (complaint procedure, etc.)	No; Yes; Do not remember
	4b	What will happen during this hospitalization	Ditto
	4c	A card or film on patient safety	Ditto
	4d	The person in the hospital whom you can contact if you have questions	Ditto
	4e	What medications you are taking	Ditto
	4f	If you would like to be resuscitated	Ditto
	4g	Any dietary and nutritional requirements	Ditto
	4h	Any hypersensitivity to substances/medication	Ditto
	4i	Your provisional discharge date	Ditto
	4j	Your personal needs during the hospitalization	Ditto
	5	Did the healthcare providers/ staff, upon arrival to the ward, have enough time for you?	No; Yes
Nurses' care during your hospitalization	6	Did the nurses listen carefully to you?	Never; Sometimes; Mostly; Always
	7	Did the nurses have enough time for you?	Ditto
	8	Did the nurses explain things to you in an understandable way?	Ditto
Doctors' care	9	Did the doctors have enough time for you?	Ditto
	10	Did the doctors explain things to you in an understandable way?	Ditto
	11	Did the doctors give you conflicting information?	Ditto
Your stay at the hospital	12	Were the toilet, the shower and the bathroom in or near your room clean?	Ditto
	13	Did you have enough privacy during your personal care?	Ditto
	14	Could receive visitors at times that you wanted?	Ditto
	15	Could you retire to a quiet place if you wanted to?	Ditto
	16	Was the food tasty?	Ditto
	17	Could you eat at the times that you wanted to?	Ditto

Domain	Item No.	Description	Response categories
Communication around treatment	18	Did the doctors or nurses tell you beforehand why the treatment, examination or intervention was needed?	Ditto
	19	Did the doctors or nurses tell you beforehand what the treatment, examination or intervention was?	Ditto
	20	Did the doctors or nurses explain to you in an understandable way about any possible side effects or consequences of a treatment, examination or intervention?	Ditto
	21	Did the staff at the hospital react quickly when you indicated you had pain?	Never; Sometimes; Mostly; Always; Not applicable, I never indicated to have had pain > go to question 23
	22	Was your pain well controlled?	Never; Sometimes; Mostly; Always
	23	Before you received a new medication or before your medication was changed, did you get an explanation of what was the purpose of the new medication?	Never; Sometimes; Mostly; Always; Not applicable, I do not use any medications > go to question 25
	24	Before you received a new medication or before your medication was changed, did you get an explanation of the possible complications in an understandable way?	Never; Sometimes; Mostly; Always;
	25	During admission to the hospital how often did you have a say in matters that were important to you?	Ditto
	26	Was the coordination of work between staff a problem?	A big problem; A small problem; No problem
Safety in this hospital	27	When medication was provided, did staff check whether the medication was intended for you, for example by asking your name or checking your wristband?	Never; Sometimes; Mostly; Always; Not applicable, I did not use any medications
	28	Before a treatment, examination, or intervention began, did staff check that you were the right person, for example by asking your date of birth?	Never; Sometimes; Mostly; Always; I do not know (any more)
	29	Did the staff of the hospital pay enough attention to unsafe situations?	Never; Sometimes; Mostly; Always

Domain	Item No.	Description	Response categories
Discharge from the hospital	30	Upon discharge from the hospital, did you receive written and verbal information about the use of new medications in combination with medication that you were already using?	No; Yes; I do not know (any more); Not applicable, I did not use any medications before my hospitalization
	31	Upon discharge from the hospital, did you receive information about submitting any complaints or health problems that you were supposed to be on the lookout for?	No; Yes; I do not know (any more)
	32	Upon discharge from the hospital, did you receive information about which activities you could or could not do?	Ditto
	33	Before you were discharged from the hospital, did you speak to the hospital staff about the help you might need after your discharge?	Ditto
	34	Upon discharge from the hospital, did you get information about what to do if problems occurred after your discharge?	Ditto

Table A2. G coefficients and standard error of measurement (SEM) for varying numbers of respondents for department-level evaluations, and for varying numbers of departments and respondents per department for hospital-level evaluations

	Department evaluations			Hospital evaluations			
Subscale	N respondents	G coeff	SEM	N departments	N respondents /department	G coeff	SEM
1. Admission	10	0.37	0.077	2	20	0.08	0.056
	20	0.54	0.054	4	20	0.15	0.040
	30	0.64	0.044	6	20	0.20	0.032
	40	0.70	0.038	8	20	0.26	0.028
	50	0.75	0.034	10	20	0.30	0.025
	60	0.78	0.031	2	50	0.11	0.048
	70	0.81	0.029	4	50	0.19	0.034
	80	0.83	0.027	6	50	0.27	0.027
	90	0.84	0.026	8	50	0.32	0.024
	100	0.86	0.024	10	50	0.38	0.021
	150	0.90	0.020	10	100	0.41	0.020
2. Communication with nurses	10	0.19	0.190	2	20	0.27	0.106
	20	0.32	0.134	4	20	0.43	0.075
	30	0.41	0.110	6	20	0.53	0.061
	40	0.49	0.095	8	20	0.60	0.053
	50	0.54	0.085	10	20	0.65	0.047
	60	0.59	0.078	2	50	0.42	0.077
	70	0.62	0.072	4	50	0.59	0.054
	80	0.65	0.067	6	50	0.68	0.044
	90	0.68	0.063	8	50	0.74	0.038
	100	0.70	0.060	10	50	0.78	0.034

	Department evaluations			Hospital evaluations			
Subscale	N respondents	G coeff	SEM	N departments	N respondents /department	G coeff	SEM
3. Communication with doctors	10	0.17	0.221	2	20	0.19	0.125
	20	0.29	0.157	4	20	0.32	0.088
	30	0.38	0.128	6	20	0.41	0.072
	40	0.45	0.111	8	20	0.49	0.062
	50	0.50	0.099	10	20	0.54	0.056
	60	0.55	0.090	2	50	0.31	0.090
	70	0.59	0.084	4	50	0.47	0.064
	80	0.62	0.078	6	50	0.57	0.052
	90	0.65	0.073	8	50	0.64	0.045
	100	0.67	0.070	10	50	0.69	0.040
4. Own contribution	10	0.45	0.201	2	20	0.54	0.131
	20	0.62	0.142	4	20	0.70	0.092
	30	0.71	0.116	6	20	0.78	0.075
	40	0.77	0.101	8	20	0.83	0.065
	50	0.80	0.090	10	20	0.86	0.058
	60	0.83	0.082	2	50	0.65	0.105
	70	0.85	0.076	4	50	0.79	0.074
	80	0.87	0.071	6	50	0.85	0.061
	90	0.88	0.067	8	50	0.88	0.052
	100	0.89	0.064	10	50	0.90	0.047
5. Explanation of treatment	10	0.26	0.21	2	20	0.17	0.130
	20	0.41	0.15	4	20	0.29	0.092
	30	0.51	0.12	6	20	0.38	0.075
	40	0.58	0.1	8	20	0.45	0.065
	50	0.63	0.09	10	20	0.50	0.058
	60	0.68	0.09	2	50	0.25	0.102
	70	0.71	0.08	4	50	0.40	0.072
	80	0.74	0.07	6	50	0.50	0.059
	90	0.76	0.07	8	50	0.57	0.051
	100	0.78	0.07	10	50	0.62	0.045
6. Pain management	10	0.21	0.194	2	20	0.14	0.115
	20	0.34	0.137	4	20	0.25	0.081
	30	0.44	0.112	6	20	0.33	0.067
	40	0.51	0.097	8	20	0.39	0.058
	50	0.57	0.087	10	20	0.45	0.052
	60	0.61	0.079	2	50	0.22	0.087
	70	0.65	0.073	4	50	0.36	0.062
	80	0.68	0.069	6	50	0.46	0.050
	90	0.70	0.065	8	50	0.53	0.044
	100	0.72	0.061	10	50	0.59	0.039

	Department evaluations			Hospital evaluations			
Subscale	N respondents	G coeff	SEM	N departments	N respondents /department	G coeff	SEM
7. Communication about medication	10	0.19	0.284	2	20	0.22	0.162
	20	0.32	0.201	4	20	0.37	0.114
	30	0.42	0.164	6	20	0.47	0.093
	40	0.49	0.142	8	20	0.54	0.081
	50	0.54	0.127	10	20	0.59	0.072
	60	0.59	0.116	2	50	0.35	0.119
	70	0.63	0.107	4	50	0.52	0.084
	80	0.66	0.100	6	50	0.62	0.068
	90	0.68	0.095	8	50	0.68	0.059
	100	0.71	0.090	10	50	0.73	0.053
8. Feeling of safety	10	0.22	0.211	2	20	0.12	0.127
	20	0.36	0.149	4	20	0.21	0.090
	30	0.45	0.122	6	20	0.29	0.074
	40	0.52	0.106	8	20	0.35	0.064
	50	0.58	0.094	10	20	0.40	0.057
	60	0.62	0.086	2	50	0.19	0.098
	70	0.66	0.080	4	50	0.32	0.069
	80	0.69	0.075	6	50	0.41	0.056
	90	0.71	0.070	8	50	0.48	0.049
	100	0.73	0.067	10	50	0.54	0.044
9. Information at discharge	10	0.37	0.094	2	20	0.09	0.068
	20	0.54	0.067	4	20	0.17	0.048
	30	0.64	0.054	6	20	0.23	0.039
	40	0.70	0.047	8	20	0.29	0.034
	50	0.75	0.042	10	20	0.33	0.030
	60	0.78	0.038	2	50	0.12	0.057
	70	0.81	0.036	4	50	0.22	0.041
	80	0.83	0.033	6	50	0.30	0.033
	90	0.84	0.031	8	50	0.36	0.029
	100	0.86	0.030	10	50	0.41	0.026
	150	0.90	0.024	10	100	0.45	0.024

CHAPTER 5

The residency learning climate and inpatient care experience in clinical teaching departments

A. Smirnova
O.A. Arah
R.E. Stalmeijer
K.M.J.M.H. Lombarts
C.P.M. van der Vleuten

Under Editorial Review

Abstract

Purpose

To study the association of departments' residency learning climate with patients' experiences of hospitalization.

Method

The authors analyzed 1201 evaluations of learning climate (evaluated using the Dutch Residency Educational Climate Test) and 6718 patients' hospitalization experiences (evaluated using the Consumer Quality Index Inpatient Hospital Care) in 86 departments across 15 specialties in 18 hospitals between 2013 and 2014. Authors used multilevel linear regressions to study the associations between departments' overall and subscale learning climate scores and patients' experiences of hospitalization, controlling for respondent- and department-level characteristics, year and correcting for multiple testing.

Results

Overall residency learning climate was positively associated with communication with doctors and feeling of safety in clinical teaching departments. The learning climate subscale coaching and assessment was positively associated with communication with doctors (b = 0.22, 95% CI 0.08 to 0.37) and explanation of treatment (b = 0.22, 95% CI 0.08 to 0.36). In departments with higher scores on formal education, patients gave lower scores on pain management (b = -0.16, 95% CI -0.26 to -0.05). However, in departments with higher scores on resident peer collaboration, patients gave higher scores on pain management (b = 0.14, 95% CI 0.03 to 0.24). Other subscales showed no significant associations.

Conclusions

Residency learning climate showed small but significant associations with specific aspects of inpatient care experience in clinical teaching departments. In particular, coaching and assessment, peer collaboration and formal education subscales of the learning climate seem to play a role, albeit not all positively associated.

Introduction

Recent efforts in quality improvement in graduate medical education (GME) have been aimed at improving the clinical learning environment in order to support residents' learning and well-being and ensure patient care safety and quality of care.[1] Residency programs are increasingly turning to residents' perception-based tools in order to assess the effectiveness of interventions to improve their clinical learning environments and to identify further areas for improvement.[2-4] Residents' collective perceptions of both formal and informal aspects of their training, including educational atmosphere and how department's educational policies and procedures are enacted, constitute a department's learning climate[5], for which a number of assessment tools have been developed.[2,6] Past research has demonstrated that optimizing the learning climate for residents in GME benefits residents' learning[7], exam performance[8], as well as their professional development[9] and well-being.[10] Ultimately, efforts to improve residents' learning climates should consider how it might affect quality of patient care in clinical learning environments. Although patient care should ideally benefit from educational improvements, however, only one study to our knowledge has researched this using patient outcomes.[11] A better understanding of the relationship between residency learning climate and healthcare quality in clinical teaching departments is needed across various specialties using relevant patient-centered outcome measures.

One of the relevant healthcare outcomes identified as an important area for quality improvement is patient-centered care, defined as care that is "respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions."[12] Measuring patients' experiences, as opposed to patient satisfaction, has become the preferred method to capture patient-centered care in practice because patient experiences better reflect quality of care received during hospitalization or an outpatient visit, including the interpersonal aspects of care as well as the extent to which the patient's needs were met during this encounter.[13,14] In addition to their association with patient satisfaction measures[15,16], better patient experiences have also been associated with a higher adherence to treatment and prevention guidelines, better clinical outcomes, better patient safety within hospitals and less care utilization.[14] Finally, since patient complaints usually revolve around the interpersonal aspects of care, such as communication and professionalism, improving patient care experience can also play a role in reducing patient complaints.[17] In the United States, patient experience has been included in the Medicare's hospital value-based purchasing program used for payment purposes and to evaluate quality of healthcare.[18]

Previous studies of organizational climates in healthcare, which relate how employees experience their organization, linked positive perceptions of organizational climate with improved employee functioning[19], quality of care[20] and patient care outcomes.[21] In particular, organizational climates supporting health care professionals' welfare, autonomy and development, have been found to positively impact patient satisfaction with healthcare services.[22] Teaching hospitals, however, present a unique situation. Carvajal and colleagues found that patient experiences differ in teaching vs. non-teaching hospitals.[23] Since in teaching departments residents tend to be viewed by patients as their primary caregivers during hospitalization[24], we hypothesized that it is conceivable that the residency learning climate, which reflects residents' perceptions of their learning environment, might also impact patient care experience. Since both the learning climate and patient care experience are multi-dimensional concepts, it is not known which aspects of the learning climate are more important in predicting patient care experiences, and which aspects of patient experiences are most impacted by the residency learning climate in clinical teaching departments.

In this study we investigated the associations of residency learning climate and inpatient care experiences in both academic and nonacademic teaching hospitals. Additionally, in order to support targeted efforts to improve the residency learning climate in relation to patient care experiences, we aimed to identify which facets of the learning climate are most likely to be associated with patient care experiences. Therefore, we addressed the following research questions: (1) is there a relationship between overall learning climate and patients' experience of their care following hospitalization?; (2) which facets of the learning climate are associated with patient care experience?

Method

Setting and data collection

The study took place in clinical teaching departments of 18 hospitals in the Netherlands that evaluated their learning climate and inpatient hospital care between January 2013 and December 2014. The Netherlands provides 28 hospital-based residency training programs between four to six years in duration. Prior to entering a residency program, a period of internship (in the same or different specialty) is common. Residency programs are organized in academic and nonacademic hospitals, where residents spend up to two years. Depending on the program, residents may also spend at least three months in a different specialty. In the Netherlands, all teaching departments, including clinical supervisors and program directors, are responsible for monitoring and improving the learning climate in residency training.[25] As part of their GME internal quality improvement activities, departments asked trainees (interns, residents, and fellows), currently rotating or who had recently completed their rotation to evaluate the learning climate using either online or paper-based questionnaires depending on the department's preferences. In the same year, patients who were 16 years or older and who spent at least 24 hours in a hospital within the previous

12 months, were invited to fill out the Consumer Quality Index (CQI) Inpatient Hospital Care according to a prescribed protocol, which included up to three reminders. Patients could respond via email or use a paper-based questionnaire. Respondents were assured their participation was voluntary and anonymous.

The institutional ethical review board of the Academic Medical Centre of the University of Amsterdam, The Netherlands, confirmed that the Medical Research Involving Human Subjects Act (WMO) did not apply to this study and provided a waiver of ethical approval for the overall study design (W14_065 #14.17.0090). However, written permissions were asked and granted from the departments using the web-based Dutch Residency Educational Climate Test (D-RECT) platform and from hospitals using paper D-RECT questionnaires. The authors obtained permissions from individual hospitals to use anonymized questionnaire data for this research and consulted a privacy officer at our institution to ensure that the data used in the study complied with Dutch Personal Data Protection Act. Participating hospitals were recruited through the Miletus Foundation (www.stichtingmiletus.nl), a coordinating body of all Consumer Quality Index (CQI) evaluations within the Netherlands. A detailed research proposal was sent to all hospitals and subsequently discussed at the general meeting. Hospitals interested in participating in the study gave informed consent either via the Miletus Foundation or by directly contacting the primary researcher (AS). MediQuest (home.mediquest.nl), a company that processes patient evaluation data from these evaluations, provided the final data set for the study.

Measures

Residency learning climate. The D-RECT is a questionnaire that was developed on the basis of interviews, a Delphi panel and literature review.[26] The questionnaire consists of 35 questions evaluated on a 5-point Likert scale (1 = totally disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = totally agree) with an additional "not applicable" option. These questions cover nine subscales of the learning climate namely 1) educational atmosphere (regarding level of respect among team members and the impact of differences of opinion between attendings on work and educational climate, as well as care), 2) teamwork (regarding ability of other healthcare team members to work well together and contribute to residents' learning), 3) role of specialty tutor (regarding involvement of the program director in monitoring and guiding residents' performance and quality of the training), 4) coaching and assessment (regarding involvement and initiative of attendings to assess and provide feedback on residents' performance), 5) formal education (regarding organization, relevance and attendance of formal education sessions), 6) resident peer collaboration (regarding teamwork among residents as a group), 7) work is adapted to residents' competence (regarding adaptation of clinical work to residents' needs and opportunities to follow up with patients), 8) accessibility of supervisors (regarding guidelines for contacting the supervisor and availability of supervisors when needed), and 9) patient sign-out (regarding the educational value of patient handover and residents' role during discussions) (see Appendix, Table A1).[5] Since its development the D-RECT has been validated and widely used internationally in the context of quality improvement in GME.[27,28] Based on a previous large-scale validation study, at least three completed questionnaires were required in order to reliably evaluate the overall learning climate, and five to eight questionnaires for reliable evaluation of the separate climate subscales.[5] We followed the recommendations for minimum numbers of completed questionnaires to calculate the subscale and overall learning climate scores for each department.

Inpatient Care Experience. Departments evaluated patient experiences of their care during hospitalization using the Consumer Quality Index (CQI) Inpatient Hospital Care, a standardized questionnaire that was developed for purposes of internal quality improvement, national benchmarking and for informing patients, regulators and insurance companies.[29] CQI Inpatient Hospital Care is a validated questionnaire measuring experiences with inpatient hospital care using seven subscales rated on a scale of 1 (Never) to 4 (Always) (see Appendix, Table A2): communication with nurses (whether nurses took time to listen and explain things to the patient) (3 items), communication with doctors (whether doctors took time to explain things to the patient) (2 items), patient's contribution (whether patient had a say in important matters around his/her care) (5 items), explanation of treatment(whether doctors and nurses explained the treatment, its purpose and possible side effects) (3 items), pain management (whether the staff reacted quickly if the patient had pain and whether the pain was well-controlled) (2 items), communication about medication (whether the purpose and possible side effects of new or changed medication were explained) (2 items), and feeling of safety (whether staff checked patient's identity before giving medication or treatment and paid sufficient attention to unsafe situations) (3 items), and 2 subscales measured on a dichotomous scale (Yes/No): admission (regarding completeness of information provided during admission to the hospital) (10 items) and discharge information (regarding completeness of information provided at discharge and quality of discharge planning) (5 items).[30] We used a 10-point rating as the overall rating of the department: "How do you rate the department where you were admitted?"

Covariates. We tested several covariates for statistical significance to all outcome measures and chose those with a statistically significant contribution. Patient covariates included in the model were patient age (16-24, 25-34, 45-54, 55-64, 65-74, 75-79, 80 years or older), sex, education level (lower secondary or less, upper secondary, tertiary education), and self-rated overall physical health (bad, moderate, good, very good, excellent), mental health (bad, moderate, good, very good, excellent), number of previous admissions to within the previous 12 months (1, 2, 3, 4 or more admissions), whether the patient needed help with filling out the questionnaire and whether the patient was born outside the Netherlands. Department covariates included total number of previous D-RECT evaluations at the department[31] as well as the gender mix of the respondents on the D-RECT survey. The average post-graduate year of the respondents on the D-RECT questionnaire or the type of hospital (academic, nonacademic or top-clinical) were not significantly associated with the outcome measures, therefore we did not include them in the final model.

Data analysis

Since the percentage of missing data in D-RECT evaluations was considered ignorable (7.4% overall), we imputed it using the expectation maximization (EM) algorithm.[32] We then calculated the mean overall learning climate score for each department as well as the subscale scores, if the minimum number of evaluations per department for the subscale was met.[5] The dataset containing CQI Inpatient Hospital Care evaluations was first analyzed for non-response (sex and age category). Significant variables were included in all subsequent analyses. Participant data were then excluded if they did not answer affirmative to the question whether they had been hospitalized in the previous 12 months or if they had more than half of the items missing. Since some variables had a relatively high percentage of missing values (up to 80% for a single variable, overall 14%) we chose multiple imputation as the appropriate missing data imputation method.[32] We imputed multiple datasets that were analyzed separately, and their parameter estimates and standard errors were subsequently pooled using Rubin's rules.[32] In order to account for potential drop in statistical power due to relatively high percentage of missing values in one variable, we opted for a higher number of imputed datasets (m = 80, maxit = 20) using patient age and gender as predictors.[33] We checked the convergence of the imputation by calculating the Rhat statistic.[34] Since our study had nested design, only departments with multiple CQI Inpatient Hospital Care evaluations were included (no departments had single evaluations). In the linked datasets we calculated intraclass correlation coefficient ICC(1) for patient evaluations of each subscale of inpatient care experience, and ICC(2) for evaluations of the departmental learning climate to assess the reliability of means of the department-level overall learning climate and the subscales.[35]

We conducted 2-level linear hierarchical panel analyses to take into account the nesting of patients within departments as well as the year of evaluation. To answer the first question we regressed the patients' overall rating of the department and each patient experience subscale on the department's overall learning climate score including covariates. To answer the second research question we regressed each subscale of patient experience on the departments' overall learning climate and each learning climate subscale with mutual adjustment for the other subscales of the learning climate, also including the covariates. Since multiple comparisons of learning climate subscale scores could increase Type I error, we applied the false discovery rate (FDR) correction to the *P* values of the multiple comparisons.[36] We set the rate of false discoveries (or false positives) at 5%. Finally, we conducted the main analysis using Bartlett factor scores instead of means as sensitivity analysis of our results.[37] The advantage of using Bartlett factor scores over a simple mean is that factor scores take into account the strength of association of each item with the subscale, thereby minimizing potential bias that can be brought in by items that are only weakly associated with the subscales when using simple means.[38]

Missing data were imputed using the mice package (version 2.25) in R statistical software (version 3.2.3).[39,40] All other analyses were performed using SPSS version 23.0.0.2 (IBM Corp., Armonk, NY).

Results

In total 86 teaching departments in 18 hospitals and 15 specialties evaluated their learning climate and inpatient care experience between 2013-2014, of which 39 (45%) departments evaluated in both years. Of the total 1204 returned D-RECT evaluations, 3 had more than 50% items missing, resulting in a total of 1201 D-RECT questionnaires to be included in the final analysis (mean 9.6; SD 6.2; min = 3; max = 42 evaluations per department). Eighty-two percent of the departments used a web-based system for learning climate evaluations, yielding a 63.1% average response rate. The response rate for paper-based sample (18% of departments) could not be reliably calculated because the number of the invited trainees was not collected. However, we expected a similar or even higher response rate for the paper-based questionnaires based on existing literature.[41] Of the 6853 returned CQI Inpatient Hospital Care questionnaires (gross response rate 30.1%), 87 were excluded due to negative or no response to the question whether or not the patient had a hospital admission within the last 12 months and 48 questionnaires were excluded due to having more than 50% of items missing. The final sample included 6718 CQI Inpatient Hospital Care questionnaires (effective response rate 27.6%; mean 53.5; SD 37.1; min = 3; max = 192 evaluations per department). Table 1 describes characteristics of the study sample.

Table 2 reports the descriptive statistics of the learning climate and patient experience scores in the study sample. Differences between departments explained a large proportion of variance in departments' overall learning climate scores (ICC2 = 76%) and the subscale scores (ICC2 = 57 to 83%). In contrast, differences between departments explained between 2-6% of variance (ICC1) in patient experience scores.

Table 3 reports the adjusted significant associations of the learning climate and its subscales with inpatient hospital care experience. The overall department learning climate was positively associated with patients' evaluations of communication with doctors ($b = 0.11$, 95% confidence interval (CI) 0.02 to 0.20, $P = 0.016$) and feeling of safety ($b = 0.09$, 95% CI 0.01 to 0.17, $P = 0.032$). Among the learning climate subscales, coaching and assessment exhibited the strongest association with communication with doctors ($b = 0.22$, 95% CI 0.08

to 0.37, *P* = 0.003) and explanation of treatment (*b* = 0.22, 95% CI 0.08 to 0.36, *P* = 0.002). Peer collaboration was positively associated with patient's pain management (*b* = 0.14, 95% CI 0.03 to 0.24, *P* = 0.01). In contrast, higher scores on formal education were negatively associated with patients' experience of pain management (*b* = -0.16, 95% CI -0.26 to -0.05, *P* = 0.003). Adjusted estimates for all learning climate subscales are provided in Appendix, Table A3. Sensitivity analysis using Bartlett factor scores instead of means for CQI Inpatient Hospital Care questionnaire subscales gave the same results (Appendix, Table A4).

Table 1. Description of the study sample for the association study of residency learning climate and patients' hospital care experience, 2013-2014

Characteristic	**No.**	**(%)**
No. teaching hospitals	18	(100)
No. departments:	86	(100)
in surgical specialties[a]	49	(57.0)
in medical specialties[b]	37	(43.0)
D-RECT questionnaire	**1201**	**(100)**
Sex of respondents:		
Male	519	(43.2)
Female	674	(56.1)
Missing	8	(0.7)
Year of training:		
1	161	(13.4)
2	162	(13.5)
3	199	(16.6)
4	243	(20.2)
5	129	(10.7)
6	89	(7.4)
Intern	198	(16.5)
Fellow	20	(1.7)
CQI Inpatient Hospital Care questionnaire	**6689**	**(100)**
Patient age (years):		
16-24	131	(1.9)
25-34	322	(4.8)
35-44	496	(7.4)
45-54	950	(14.2)
55-64	1630	(24.4)
65-74	1608	(24.0)
75-79	709	(10.6)
80 and older	843	(12.6)
Patient sex:		
Female	3321	(49.6)
Level of education:		
Lower secondary or less	1931	(28.9)
Upper secondary	3061	(45.8)
Tertiary education	1697	(25.4)

Country of birth:		
Netherlands	6101	(91.2)
Other	561	(8.4)
Missing	27	(0.4)
Language spoken at home:		
Dutch/Fries/Dutch dialect	6280	(93.9)
Other	314	(4.7)
Missing	95	(1.4)
Self-reported overall health:		
Excellent	393	(5.9)
Very good	770)	(11.5
Good	3039	(45.4)
Moderate	2082	(31.1)
Bad	405	(6.1)
Self-reported psychological health:		
Excellent	1230	(18.4)
Very good	1565	(23.4)
Good	3161	(47.3)
Moderate	657	(9.8)
Bad	76	(1.1)
Total number of admissions in previous year:		
1	3716	(55.6)
2	1806	(27.0)
3	675	(10.1)
4 or more	464	(6.9)
Missing	28	(0.4)
Help with filling out the questionnaire:		
No	5954	(89.0)
Yes	714	(10.7)
Missing	21	(0.3)

Abbreviations: Dutch Residency Educational Climate Test (D-RECT), Consumer Quality Index (CQI)

[a] Surgical specialties: surgery, obstetrics & gynecology, orthopedic surgery, plastic surgery, urology, otolaryngology, neurosurgery, ophthalmology, cardiothoracic surgery

[b] Medical specialties: internal medicine, cardiology, neurology, pulmonology, anesthesiology, dermatology

Table 2. Descriptive statistics of the D-RECT and CQI Inpatient Hospital Care for the association study of residency learning climate and patients' hospital care experience, 2013-2014

	No. evaluations	Mean no. evaluations per dept	Scale	Mean	SD	Min.	Max.	ICC (1)[b]	ICC (2)[c]
D-RECT									
Overall learning climate	1201	9.6	1-5	3.87	0.29	2.58	4.57	0.25	0.76
Peer collaboration	904	13.5	1-5	4.25	0.36	3.44	4.92	0.25	0.82
Patient sign-out	904	13.5	1-5	3.75	0.39	2.81	4.56	0.25	0.82
Educational Atmosphere	1057	11.6	1-5	3.86	0.47	1.80	4.80	0.30	0.83
Teamwork	1057	11.6	1-5	3.92	0.36	3.00	4.76	0.21	0.76
Coaching and assessment	1057	11.6	1-5	3.24	0.35	2.08	4.03	0.17	0.70
Formal education	1057	11.6	1-5	3.80	0.37	2.56	4.54	0.22	0.77
Role of specialty tutor	1137	10.6	1-5	4.04	0.33	2.83	4.73	0.18	0.70
Adaptation of work to residents' competence	1137	10.6	1-5	3.97	0.35	2.47	4.78	0.14	0.63
Accessibility of supervisors	1137	10.6	1-5	4.23	0.29	3.00	4.89	0.11	0.57
CQI Inpatient Hospital Care[a]									
Admission	6689	-	0-1	0.57	0.25	0	1	0.05	-
Communication with nurses	6689	-	1-4	3.36	0.63	1	4	0.02	-
Communication with doctors	6689	-	1-4	3.32	0.74	1	4	0.02	-
Patient's contribution	6689	-	1-4	2.97	0.68	1	4	0.06	-
Explanation of treatment	6689	-	1-4	3.45	0.70	1	4	0.03	-
Pain management	6689	-	1-4	3.47	0.67	1	4	0.03	-
Communication about medication	6689	-	1-4	2.95	0.97	1	4	0.04	-
Feeling of safety	6689	-	1-4	3.39	0.68	1	4	0.03	-
Information at discharge	6689	-	0-1	0.72	0.31	0	1	0.06	-
Overall rating of the department	6689	-	0-10	7.96	1.65	0	10	0.02	-

Abbreviations: Dutch Residency Educational Climate Test (D-RECT), Consumer Quality Index (CQI), Intraclass Correlation (ICC)

[a] Pooled mean scores and average standard deviations (SD) across 80 multiply imputed datasets are reported in the Table.

[b] ICC(1) refers to the proportion of variance in individual learning climate and patient experience scores explained by between-departments differences. ICC(1) is the proportion of between-department variance to total variance (total variance = between-department variance + residual variance).

[c] ICC(2) refers to the proportion of variance in departments' learning climate scores explained by between-department differences. ICC(2) is calculated similarly to ICC(1), however the residual variance is divided by the average number of observations per department (reported in the last column).

Table 3. Adjusted unstandardized regression coefficients (*b*) and their 95% confidence intervals (CIs) for the significant associations ($P < 0.05$) between residency learning climate on inpatient hospital care experience, 2013-2014

Inpatient hospital care experience	*b* (95% CI)
Communication with doctors	
Overall learning climate	0.11 (0.02 – 0.20)
Coaching and assessment	0.22 (0.08 – 0.37)
Explanation of treatment	
Coaching and assessment	0.22 (0.08 – 0.36)
Pain management	
Peer collaboration	0.14 (0.03 – 0.24)
Formal education	-0.16 (-0.26 – -0.05)
Feeling of safety	
Overall learning climate	0.09 (0.01 – 0.17)

Discussion

While the overall residency learning climate is not associated with the overall rating of patient experience within the department, on inspection of specific dimensions of patient experience, we found small but significant associations of the overall learning climate with communication with doctors and feeling of safety. Furthermore, certain dimensions of the learning climate were also positively associated with certain dimensions of patient experience, including coaching and assessment with doctors' communication and explanation of treatment as well as peer collaboration with patient experience of pain management. Formal education, however, was negatively associated with patients' experiences of pain management.

Explanation of findings

In this study we found no association between a department's overall learning climate score and patients' average rating of the department. One potential explanation is related to the measures themselves. While the overall learning climate score is based on an average of the 35 items of the D-RECT questionnaire, the patients' global rating of the department was based on a mean of a single item rated 1-10 where patients were asked to globally evaluate their experience. Previous studies comparing global ratings to overall questionnaire scores have preferred the latter since global ratings show poor associations with the individual dimensions of experience.[42,43] A recent validation study of the CQI Inpatient Hospital Care questionnaire similarly showed that global ratings had inconsistent relationships with the individual dimensions of patient care experience.[30] Therefore, using global ratings may not provide an accurate evaluation of the overall inpatient experience in in this setting. As a result, the associations of the residency learning climate with the specific dimensions of patient experience may provide a better reflection of the nature of the association than using global measures alone.

Two domains of inpatient experience demonstrated small but significant positive associations with overall learning climate, namely communication with doctors and feeling of safety. Communication with doctors reflects patients' perceptions of whether doctors spent sufficient time with the patient and whether the doctors explained things to patients in an understandable way. Previous research shows that clinical supervisors facilitate residents' development of clinical, professional and communication competencies in the learning environment[44,45], and departments' learning climate is associated with better teaching qualities in clinical supervisors.[46] The results of this study are in line with these findings. The subscale coaching and assessment reflects the extent to which clinical supervisors observe and provide feedback on residents' clinical competence and performance. It is known that supervisors who actively engage in patient care together with the resident, such as direct observation of residents, can better balance patient care and residents' learning needs.[47-49] On the other hand, supervisors' involvement may not only contribute to a better overall learning climate, but also to patients' perceptions of safety in a teaching environment. In a recent study, Silkens and colleagues (*under editorial review*) found that residency learning climate was associated with departments' patient safety climate, which in turn was associated with better residents' patient safety behaviour and compliance. Therefore, the association between overall learning climate and patients' perceptions of safety could also be mediated by improved patient safety climate at teaching departments.

The relatively small effect sizes reported in this study could potentially be due to the answer scales used to evaluate both the learning climate and inpatient experience. Perception-based tools suffer from ceiling effects, limited number of answer categories and generally negatively skewed distributions, all of which make finding differences in performance difficult.[38] A previous study assessing change in the D-RECT scores over time found small but significant change[31], which was similar to studies using similar perception-based tools.[50] Furthermore, aggregation of the D-RECT scores to calculate the overall department residency climate score may have resulted in the loss of information of important associations.[38] It is, therefore, important to examine the individual D-RECT subscales in addition to the overall learning climate.

The results of the associations of the learning climate subscales showed that three of the nine subscales were significantly associated with inpatient care experience, while six out of nine learning climate subscales were not significantly associated with inpatient care experience. Some of these showed statistically significant associations, but were subsequently deemed potentially false discoveries after the FDR procedure. Additionally, the lack of association may indicate that these learning climate domains do not contribute to patient care experience or that there are other elements influencing patient satisfaction that may be at play (e.g. the role of the nurse) that may compensate for a poor learning environment or vice versa. Among the learning climate domains that were significantly associated with inpatient care experience, we found mixed results. Formal education, which reflects the degree of organization, relevance and attendance of formal education sessions by residents and staff, was negatively associated with patients' perceptions of pain management, which reflects the staff's ability to respond to pain quickly and manage the pain well. These findings echo our previous study's results of higher odds of adverse perinatal outcomes in nonacademic obstetrics and gynecology teaching departments.[11] On the other hand, peer collaboration had a positive association with pain management. Since in those departments where residents perceive working better together patients also reported better management of pain, we may infer that activities that are resident-focused may come in conflict with immediate patient care needs and patient-centered activities may offset these negative effects. These results add to the growing literature of the tensions between training and care in clinical teaching departments[51,52], and provide potential insights on how they can be managed.

Limitations

One limitation may include possible ceiling effects due to rating scales used in the questionnaires, making it more difficult to detect smaller changes in patient experiences especially if the ratings tend to be high. We also did not have information on non-responders. Bias due to non-response could be less of a problem with the D-RECT, which had a response rate of 63%. Non-response bias could be more important with CQI Inpatient Hospital Care questionnaire due to its low response rate (effective response rate 28%). We minimized the potential for selection bias in CQI questionnaire by adjusting for important non-responder characteristics (sex and age) in our analyses, however, we could not account for all potentially important characteristics.[53] Memory is another factor that may influence patients' perceptions of their hospitalization. There was no way to record the variability in time since their hospitalization that patients filled out the questionnaire, which may have affected the overlap between residents and patients' perceptions. Finally, due to the cross-sectional nature of the study, we cannot establish causal relationships between learning climate and inpatient experiences. Well-designed longitudinal studies are needed to provide additional evidence of causality.

Implications for practice and research

Programs could potentially increase both the residency learning climate and patient care experiences by emphasizing those aspects of the residents' clinical learning environment that are integrated with direct patient care, such as workplace coaching and assessment and collaboration between residents. Programs may need to critically assess those aspects of residency training that are not well integrated with daily clinical practice or direct patient care, such as formal education sessions, in order to identify potential tensions that may exist and mitigate potential negative effects on clinical practice. Future research should address the factors influencing the relationship between learning climate and patient care experiences, which can shed light on the mechanisms behind these associations. Qualitative studies could be particularly well suited for identifying potential mechanisms. Testing the effect of interactions between supervision and learning climate perceptions on patients' care experience can also provide insights into how learning climate and supervision can be optimized.

Conclusions

Optimizing the learning climate in residency is an important step to ensuring high quality of GME training as well as patient care in clinical teaching departments. This study adds to a growing body of evidence on how residency learning environment relates to patient care quality in clinical teaching departments. Future research is can explore mechanisms and better understand the potential tensions between aspects of the residency learning climate and patient care experiences, and how to best resolve them without compromising patient care or training quality.

References:

1. Nasca TJ, Weiss KB, Bagian JP. Improving clinical learning environments for tomorrow's physicians. *N Engl J Med*. 2014;370(11):991-993.
2. Colbert-Getz JM, Kim S, Goode VH, Shochet RB, Wright SM. Assessing medical students' and residents' perceptions of the learning environment: exploring validity evidence for the interpretation of scores from existing tools. *Acad Med*. 2014;89(12):1687-1693.
3. Philibert I. Satisfiers and hygiene factors: residents' perceptions of strengths and limitations of their learning environment. *J Grad Med Educ*. 2012;4(1):122-127.
4. Holt KD, Miller RS, Philibert I, Heard JK, Nasca TJ. Residents' perspectives on the learning environment: data from the Accreditation Council for Graduate Medical Education resident survey. *Acad Med*. 2010;85(3):512-518.
5. Silkens ME, Smirnova A, Stalmeijer RE, et al. Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. *Med Teach*. 2016;38(5):476-481.
6. Schonrock-Adema J, Bouwkamp-Timmer T, van Hell EA, Cohen-Schotanus J. Key elements in assessing the educational environment: where is the theory? *Adv Health Sci Educ Theory Pract*. 2012;17(5):727-742.
7. Delva MD, Kirby J, Schultz K, Godwin M. Assessing the relationship of learning approaches to workplace climate in clerkship and residency. *Acad Med*. 2004;79(11):1120-1126.
8. Shimizu T, Tsugawa Y, Tanoue Y, et al. The hospital educational environment and performance of residents in the General Medicine In-Training Examination: a multicenter study in Japan. *Int J Gen Med*. 2013;6:637-640.
9. Cross V, Hicks C, Parle J, Field S. Perceptions of the learning environment in higher specialist training of doctors: implications for recruitment and retention. *Med Educ*. 2006;40(2):121-128.
10. van Vendeloo SN, Brand PL, Verheyen CC. Burnout and quality of life among orthopaedic trainees in a modern educational programme: importance of the learning climate. *Bone Joint J*. 2014;96-B(8):1133-1138.
11. Smirnova A, Ravelli ACJ, Stalmeijer RE, et al. The association between learning climate and adverse obstetrical outcomes in 16 nontertiary obstetrics-gynecology departments in the Netherlands. *Acad Med*. 2017;Publish Ahead of Print.
12. Crossing the quality chasm: a new health system for the 21st century. The National Academies Press; 2001.
13. Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst Rev*. 2015;4:97.
14. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev*. 2014;71(5):522-554.
15. Batbaatar E, Dorjdagva J, Luvsannyam A, Savino MM, Amenta P. Determinants of patient satisfaction: a systematic review. *Perspect Public Health*. 2017;137(2):89-101.
16. Bjertnaes OA, Sjetne IS, Iversen HH. Overall patient satisfaction with hospitals: effects of patient-reported experiences and fulfilment of expectations. *BMJ Qual Saf*. 2012;21(1):39-46.
17. Wofford MM, Wofford JL, Bothra J, Kendrick SB, Smith A, Lichstein PR. Patient complaints about physician behaviors: a qualitative study. *Acad Med*. 2004;79(2):134-138.
18. Zhao M, Haley DR, Spaulding A, Balogh HA. Value-based purchasing, efficiency, and hospital performance. *Health Care Manag (Frederick)*. 2015;34(1):4-13.
19. Bronkhorst B, Tummers L, Steijn B, Vijverberg D. Organizational climate and employee mental health outcomes: A systematic review of studies in health care organizations. *Health Care Manage Rev*. 2015;40(3):254-271.
20. Benzer JK, Young G, Stolzmann K, et al. The relationship between organizational climate and quality of chronic disease management. *Health Serv Res*. 2011;46(3):691-711.
21. MacDavitt K, Chou SS, Stone PW. Organizational climate and health care outcomes. *Jt Comm J Qual Patient Saf*. 2007;33(11 Suppl):45-56.
22. Ancarani A, Di Mauro C, Giammanco MD. How are organisational climate models and patient satisfaction related? A competing value framework approach. *Soc Sci Med*. 2009;69(12):1813-1818.
23. Carvajal DN, Blank AE, Lechuga C, Schechter C, McKee MD. Do primary care patient experiences vary by teaching versus nonteaching facility? *J Am Board Fam Pract*. 2014;27(2):239-248.
24. Dalia S, Schiffman FJ. Who's my doctor? First-year residents and patient care: hospitalized patients' perception of their "main physician". *J Grad Med Educ*. 2010;2(2):201-205.

25. Directive of the Central College of Medical Specialists. Utrecht: Royal Dutch Medical Association [In Dutch]; 2009.
26. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011;33(10):820-827.
27. Piek J, Bossart M, Boor K, et al. The work place educational climate in gynecological oncology fellowships across Europe: the impact of accreditation. *Int J Gynecol Cancer*. 2015;25(1):180-190.
28. Bennett D, Dornan T, Bergin C, Horgan M. Postgraduate training in Ireland: expectations and experience. *Ir J Med Sci*. 2014;183(4):611-620.
29. Delnoij DM, Rademakers JJ, Groenewegen PP. The Dutch consumer quality index: an example of stakeholder involvement in indicator development. *BMC Health Serv Res*. 2010;10:88.
30. Smirnova A, Lombarts KM, Arah OA, van der Vleuten CP. Closing the patient experience chasm: A two-level validation of the Consumer Quality Index Inpatient Hospital Care. *Health Expect*. 2017;00:1-8.
31. Silkens ME, Arah OA, Scherpbier AJ, Heineman MJ, Lombarts KM. Focus on Quality: Investigating Residents' Learning Climate Perceptions. *PLoS One*. 2016;11(1):e0147108.
32. Dong Y, Peng CY. Principled missing data methods for researchers. *Springerplus*. 2013;2(1):222.
33. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci*. 2007;8(3):206-213.
34. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press; 2007.
35. Bliese PD. Within-group agreement, non-independence, and reliability: implications for data aggregation and analysis. In: Klein KJ, Kozlowski SWJ, ed. Multilevel theory, research, and methods in organizations: foundations, extensions, and new directions. San Francisco: Jossey-Bass Inc.; 2000.
36. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met*. 1995;57(1):289-300.
37. DiStefano CZ, M.; Mindrila, D. Understanding and using factor scores: considerations for the applied researcher. *Practical Assessment, Research & Evaluation*. 2009;14(20).
38. Boerebach BC, Arah OA, Heineman MJ, Lombarts KM. Embracing the complexity of valid assessments of clinicians' performance: a call for in-depth examination of methodological and statistical contexts that affect the measurement of change. *Acad Med*. 2016;91(2):215-220.
39. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate Iimputation by chained equations in R. *J Stat Softw*. 2011;45(3):67.
40. R Core Team. R: A language and environment for statistical computing. 2015; https://www.R-project.org/.
41. Yarger JB, James TA, Ashikaga T, et al. Characteristics in response rates for surveys administered to surgery residents. *Surgery*. 2013;154(1):38-45.
42. Krol MW, de Boer D, Rademakers JJ, Delnoij DM. Overall scores as an alternative to global ratings in patient experience surveys; a comparison of four methods. *BMC Health Serv Res*. 2013;13:479.
43. de Boer D, Delnoij D, Rademakers J. Do patient experiences on priority aspects of health care predict their global rating of quality of care? A study in five patient groups. *Health Expect*. 2010;13(3):285-297.
44. Subramaniam A, Silong AD, Uli J, Ismail IA. Effects of coaching supervision, mentoring supervision and abusive supervision on talent development among trainee doctors in public hospitals: moderating role of clinical learning environment. *BMC Med Educ*. 2015;15(1):1-9.
45. Bates J, Ellaway RH. Mapping the dark matter of context: a conceptual scoping review. *Med Educ*. 2016;50(8):807-816.
46. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One*. 2014;9(1):e86512.
47. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad Med*. 2012;87(4):428-442.
48. van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med*. 2012;10:65.
49. Piquette D, Moulton CA, LeBlanc VR. Model of interactive clinical supervision in acute care environments. Balancing patient care and teaching. *Ann Am Thorac Soc*. 2015;12(4):498-504.

50. Van Der Leeuw RM, Boerebach BC, Lombarts KM, Heineman MJ, Arah OA. Clinical teaching performance improvement of faculty in residency training: A prospective cohort study. *Med Teach*. 2015:1-7.
51. Cleland J, Roberts R, Kitto S, Strand P, Johnston P. Using paradox theory to understand responses to tensions between service and training in general surgery. *Med Educ.* 2017.
52. Hoffman KG, Donaldson JF. Contextual tensions of the clinical environment and their influence on teaching and learning. *Med Educ*. 2004;38(4):448-454.
53. Zaslavsky AM, Zaborski LB, Cleary PD. Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. *Med Care*. 2002;40(6):485-499.

Appendix

Table A1. Items in the Dutch Residency Educational Climate Test (D-RECT) (evaluated on a 5-point Likert scale)

Educational atmosphere

1. Continuity of care is not affected by differences of opinion between attendings.
2. Differences of opinion between attendings about patient management are discussed in such a manner that is instructive to others present.
3. Differences of opinion are not such that they have a negative impact on the work climate.
4. There is (are) NO attending physician(s) who have a negative impact on the educational climate.
5. My attendings treat me with respect.

Teamwork

6. Attendings, nursing staff, other allied health professionals and residents work together as a team.
7. Nursing staff and other allied health professionals make a positive contribution to my training.
8. Nursing staff and other allied health professionals are willing to reflect with me on the delivery of patient care.

Role of specialty tutor

9. The specialty tutor monitors the progress of my training.
10. The specialty tutor provides guidance to other attendings when needed.
11. The specialty tutor is actively involved in improving the quality of education and training.
12. In this rotation evaluations are useful discussions about my performance.
13. My plans for the future are part of the discussion.
14. During evaluations, input from several attendings is considered.

Coaching and assessment

15. My attendings take the initiative to evaluate my performance.
16. My attendings take the initiative to evaluate difficult situations I have been involved in.
17. My attendings evaluate whether my performance in patient care is commensurate with my level of training.
18. My attendings occasionally observe me taking a history.
19. My attendings assess not only my medical expertise but also other skills such as teamwork, organization or professional behavior.
20. My attendings give regular feedback on my strengths and weaknesses.

Formal education

21. Residents are generally able to attend scheduled educational activities.
22. Educational activities take place as scheduled.
23. Attendings contribute actively to the delivery of high-quality formal education.
24. Formal education and training activities are appropriate to my needs.

Resident peer collaboration

25. Residents work well together.
26. Residents, as a group, make sure the day's work gets done.
27. Within our group of residents it is easy to find someone to cover or exchange a call.

Work is adapted to residents' competence

28. The work I am doing is commensurate with my level of experience.
29. The work I am doing suits my learning objectives at this stage of my training.
30. It is possible to do follow up with patients.

Accessibility of supervisors

31. When I need an attending, I can always contact one.
32. When I need to consult an attending, they are readily available.
33. It is clear which attending supervises me.

Patient sign-out

34. Sign-out is used as a teaching opportunity.
35. Attendings encourage residents to join in the discussion during sign-out.

Table A2. Items and subscales in the shortened version of the Consumer Quality Index (CQI) Inpatient Hospital Care

Subscale	Item No.	Description	Response categories
Admission		Were the following items discussed with you on your admission to the hospital?	
	4a	Your rights as patient (complaint procedure, etc.)	No; Yes; Do not remember
	4b	What will happen during this hospitalization	Ditto
	4c	A card or film on patient safety	Ditto
	4d	The person in the hospital whom you can contact if you have questions	Ditto
	4e	What medications you are taking	Ditto
	4f	If you would like to be resuscitated	Ditto
	4g	Any dietary and nutritional requirements	Ditto
	4h	Any hypersensitivity to substances/medication	Ditto
	4i	Your provisional discharge date	Ditto
	4j	Your personal needs during the hospitalization	Ditto
Communication with nurses	6	Did the nurses listen carefully to you?	Never; Sometimes; Mostly; Always
	7	Did the nurses have enough time for you?	Ditto
	8	Did the nurses explain things to you in an understandable way?	Ditto
Communication with doctors	9	Did the doctors have enough time for you?	Ditto
	10	Did the doctors explain things to you in an understandable way?	Ditto
Patient's contri-bution	13	Did you have enough privacy during your personal care?	Ditto
	14	Could receive visitors at times that you wanted?	Ditto
	15	Could you retire to a quiet place if you wanted to?	Ditto
	17	Could you eat at the times that you wanted to?	Ditto
	25	During admission to the hospital how often did you have a say in matters that were important to you?	Ditto
Explanation of treatment	18	Did the doctors or nurses tell you beforehand why the treatment, examination or intervention was needed?	Ditto
	19	Did the doctors or nurses tell you beforehand what the treatment, examination or intervention was?	Ditto
	20	Did the doctors or nurses explain to you in an understandable way about any possible side effects or consequences of a treatment, examination or intervention?	Ditto
Pain manage-ment	21	Did the staff at the hospital react quickly when you indicated you had pain?	Never; Sometimes; Mostly; Always; Not applicable, I never indicated to have had pain > go to question 23
	22	Was your pain well controlled?	Never; Sometimes; Mostly; Always

Subscale	Item No.	Description	Response categories
Communication abound medication	23	Before you received a new medication or before your medication was changed, did you get an explanation of what was the purpose of the new medication?	Never; Sometimes; Mostly; Always; Not applicable, I do not use any medications > go to question 25
	24	Before you received a new medication or before your medication was changed, did you get an explanation of the possible complications in an understandable way?	Never; Sometimes; Mostly; Always
Feeling of safety	27	When medication was provided, did staff check whether the medication was intended for you, for example by asking your name or checking your wristband?	Never; Sometimes; Mostly; Always; I do not know (any more); Not applicable, I did not use any medications
	28	Before a treatment, examination, or intervention began, did staff check that you were the right person, for example by asking your date of birth?	Never; Sometimes; Mostly; Always; I do not know (any more)
	29	Did the staff of the hospital pay enough attention to unsafe situations?	Never; Sometimes; Mostly; Always;
Discharge information	30	Upon discharge from the hospital, did you receive written and verbal information about the use of new medications in combination with medication that you were already using?	No; Yes; I do not know (any more); Not applicable, I did not use any medications before my hospitalization
	31	Upon discharge from the hospital, did you receive information about submitting any complaints or health problems that you were supposed to be on the lookout for?	No; Yes; I do not know (any more)
	32	Upon discharge from the hospital, did you receive information about which activities you could or could not do?	Ditto
	33	Before you were discharged from the hospital, did you speak to the hospital staff about the help you might need after your discharge?	Ditto
	34	Upon discharge from the hospital, did you get information about what to do if problems occurred after your discharge?	Ditto

Table A3. Adjusted unstandardized regression coefficients using means for the associations between residency learning climate on patients' hospital care experience, 2013-2014[a]

	Learning climate									
Patient hospital care experience	**Overall learning climate (1-5)**	**Peer collaboration (1-5)**	**Patient sign-out (1-5)**	**Educational atmosphere (1-5)**	**Teamwork (1-5)**	**Coaching and assessment (1-5)**	**Formal education (1-5)**	**Role of specialty tutor (1-5)**	**Adaptation of work to residents' competence (1-5)**	**Accessibility of supervision (1-5)**
Admission (0-1)	0.02	0.03	0.02	-0.02	-0.02	0.06	-0.01	0.04	-0.04	-0.07
Communication with nurses (1-4)	0.01	0.06	0.009	0.02	-0.003	0.09	-0.08	-0.03	0.06	-0.07
Communication with doctors (1-4)	0.11[b]	0.005	0.02	-0.04	-0.08	0.22[b]	-0.02	-0.01	0.06	-0.12
Patient's contribution (1-4)	0.02	0.10	0.11	-0.02	-0.003	0.23	-0.11	-0.23	-0.04	-0.13
Explanation of treatment (1-4)	0.07	0.14	0.01	-0.02	-0.13	0.22[b]	-0.09	0.04	-0.05	-0.13
Pain management (1-4)	-0.003	0.14[b]	0.07	-0.07	0.001	0.07	-0.16[b]	0.06	0.04	-0.04
Communication about medicines (1-4)	0.10	0.05	0.06	0.08	-0.19	0.12	-0.03	-0.25	0.16	-0.36
Feeling of safety (1-4)	0.09[b]	0.07	0.05	-0.04	-0.08	0.08	-0.08	-0.003	0.002	0.07
Information at discharge (0-1)	0.04	0.04	0.04	-0.03	-0.04	0.07	-0.06	0.09	-0.03	-0.01
Overall rating of the department (1-10)	0.03	0.14	0.09	-0.10	0.26	0.26	-0.27	-0.18	0.19	-0.27

[a] Coefficients adjusted for patient's age, sex, education, general physical health, psychological health, non-Dutch birth, number of admissions in the previous 12 months, help provided in filling out the questionnaire, the number of previous Dutch Residency Educational Test (D-RECT) evaluations as well as gender mix of respondents on D-RECT and year.

[b] indicates $P < 0.05$, including False Discovery Rate (FDR) correction for multiple comparisons.

Table A4. Adjusted unstandardized regression coefficients for sensitivity analysis using Bartlett factor scores of the association between residency learning climate on patients' hospital care experience, 2013-2014a

	Learning climate									
Patient hospital care experience	**Overall learning climate (1-5)**	**Peer collaboration (1-5)**	**Patient sign-out (1-5)**	**Educational atmosphere (1-5)**	**Teamwork (1-5)**	**Coaching and assessment (1-5)**	**Formal education (1-5)**	**Role of specialty tutor (1-5)**	**Adaptation of work to residents' competence (1-5)**	**Accessibility of supervision (1-5)**
Admission (0-1)	0.09	0.10	0.10	-0.09	-0.13	0.27	-0.04	0.14	-0.19	-0.27
Communication with nurses (1-4)	0.02	0.09	0.02	0.02	-0.004	0.13	-0.13	-0.05	0.09	-0.11
Communication with doctors (1-4)	0.15[b]	0.007	0.03	-0.06	-0.11	0.30[b]	-0.03	-0.02	0.08	-0.17
Patient's contribution (1-4)	0.03	0.12	0.16	-0.02	-0.007	0.33	-0.16	-0.32	-0.06	-0.18
Explanation of treat-ment (1-4)	0.09	0.20	0.01	-0.03	-0.18	0.32[b]	-0.12	0.06	-0.08	-0.18
Pain management (1-4)	-0.004	0.21[b]	0.10	-0.11	0.002	0.11	-0.23[b]	0.10	0.06	-0.07
Communication about medicines (1-4)	0.10	0.06	0.06	0.09	-0.19	0.12	-0.04	-0.25	0.15	-0.37
Feeling of safety (1-4)	0.13[b]	0.11	0.08	-0.07	-0.12	0.13	-0.13	0.02	-0.01	0.11
Information at dis-charge (0-1)	0.11	0.11	0.13	-0.12	-0.11	0.24	-0.20	0.28	-0.07	-0.04

[a] Coefficients adjusted for patient's age, sex, education, general physical health, psychological health, non-Dutch birth, number of admissions in the previous 12 months, help provided in filling out the questionnaire, the number of previous D-RECT evaluations, gender mix of respondents on D-RECT and year.

[b] indicates $P < 0.05$, including False Discovery Rate (FDR) correction for multiple comparisons.

CHAPTER 6

Associations of anesthesiology faculty's teaching performance and role modeling with perioperative care quality

A. Smirnova
F.O. Kooij
O.A. Arah
R.E. Stalmeijer
M.J. Heineman
C.P.M. van der Vleuten
K.M.J.M.H. Lombarts

Under Editorial Review

Abstract

Background

To investigate associations of anesthesiology faculty's teaching performance and role modeling with perioperative quality of care (QoC).

Methods

We analyzed 757 residents' evaluations of 54 anesthesiology faculty's teaching performance and role modeling between 2010-2012 using the System for Evaluation of Teaching Qualities. Using hierarchical panel analyses, we investigated the associations of faculty's mean teaching performance and role modeling scores with perioperative QoC in cases with and without a resident present in the period six months following the teaching evaluation. QoC measures included (1) intra-operative temperature monitoring and normothermia, (2) post-operative pain scores, (3) neuromuscular monitoring using Train of Four (TOF) count/value prior to extubation in the operating room and achievement of TOF value >70 or >90, (4) postoperative nausea and vomiting prophylaxis.

Results

Of over 15000 patient encounters, 43% were together with a resident. In cases without a resident, faculty with higher teaching performance scores outperformed on neuromuscular monitoring: Train of Four (TOF) ratio/count measured (OR 1.67, 95% CI 1.17 – 2.39), TOF values >70 (OR 1.96, 95% CI 1.40 – 2.75) or >90 (OR1.94, 95% CI 1.40 – 2.68). After controlling for the anesthesiology assistant in sensitivity analysis, patients of faculty with higher role modeling scores were also more likely to have TOF value >70 (OR 1.33, 95% CI 1.02 – 1.73) or TOF value >90 (OR 1.39, 95% CI 1.08 – 1.79). However, these associations were attenuated in cases with a resident present.

Conclusions

Faculty clinical teaching and role modeling appear to be associated with better performance on some perioperative care measures when residents were absent.

Introduction

Clinical teaching is the foundation of graduate medical education (GME).[1] Clinical teachers transmit essential skills, knowledge and attitudes to residents, provide essential feedback and create a safe learning climate, in which residents can discuss their uncertainties about patients and discover their own strengths and weaknesses. Residents who rate their clinical teachers highly are also more likely to consider them to be role models.[2,3] Role-modeling constitutes another powerful teaching strategy because residents tend to emulate their role models' professional values, attitudes and clinical behaviour in their own practice.[4-6] Since residents observe and emulate their role models' behaviour in clinical practice, it is essential that faculty demonstrate excellent clinical competence in addition to having good teaching and interpersonal skills.[5,7]

Despite the importance of excellent clinical competence in clinical teachers, the existing evidence on the relationship between faculty's teaching qualities, role modeling and clinical performance is surprisingly sparse. Previous reviews of clinical teaching evaluations have not explicitly addressed the relationship of teaching performance and clinical performance of teaching faculty.[8-10] However, individual studies show mixed results. In one study, faculty who were involved in ongoing bedside clinical teaching of pre-clerkship students perceived that teaching had positive effects on their own clinical skills.[11] In another study, Tortolani et al. found that residents' ratings of general surgeons' interpersonal skills correlated inversely with surgeons' complication rate, and residents' overall ratings of general surgeons' teaching correlated inversely with patients' length of stay, even after adjustment for case-mix differences and differences in the amount of contact with residents.[12] On the other hand, Mourad et al. did not find a correlation between internists' teaching effectiveness scores and their patients' mean length of stay or readmission rates.[13] However, residents' involvement in surgeries and the team-based nature of internal medicine makes it difficult to attribute patient outcomes to an individual faculty member. In a study of role modeling, Kravet et al. found that faculty who were considered as excellent by their peers were more likely to be named as role models by trainees.[14] Yet this study relied on perceptions of clinical performance rather than clinical measures.

In this study we aimed to provide empirical evidence for the relationship of residents' perceptions of anesthesiology faculty's teaching quality and role modeling with faculty's clinical performance in an academic teaching hospital, using a valid and reliable questionnaire for the assessment of anesthesiologists' teaching performance that has been developed and used over a number of years.[15,16] In addition, recent developments in electronic patient data collection have made it possible to identify physician-attributable outcome measures, as well as record whether a patient encounter was conducted with or without a resident.[17,18] Hence, our main research question in this study was: Do clinical teaching evaluations and/or faculty's role modeling status predict faculty's clinical performance? Based on previous studies discussed earlier, we hypothesized that those faculty receiving higher clinical teaching scores and/or role modeling scores would also perform better on pre-defined clinical performance indicators. Since faculty are also involved in teaching residents, we additionally investigated the following research question: How does the presence of a resident influence the association between faculty's teaching quality/role modeling status and their clinical performance?

Materials and methods

Setting and data collection

We studied a retrospective cohort of anesthesiology faculty's teaching, role modeling and clinical performance in an Academic Medical Centre in the Netherlands between 2010-2012. The faculty in this teaching department received yearly evaluations of their teaching qualities by residents. In the study period residents were invited to evaluate the teaching qualities of their clinical supervisors. Residents concurrently rotating on the service received an invitation via email to fill out an anonymous web-based questionnaire. The information accompanying the invitation stressed the formative purpose of the evaluation and the anonymous and voluntary nature of resident participation. Residents could choose which faculty to evaluate, but they were encouraged to evaluate the faculty that they had had the most contact with. The evaluation period lasted one month. In order to encourage participation, up to three automatic e-mail reminders were sent and the program director also encouraged participation. At the end of the evaluation period, faculty received a feedback report summarizing the residents' anonymized feedback.

Clinical performance data was accessed with the permission of the department head through the hospital's electronic record system for the study period. We accessed de-identified data on patients' demographics, co-morbidities, type of procedure, codes of the responsible anesthesiologist and resident, timing of the anesthesia, as well as the pre-defined clinical measures described below. We included perioperative cases where the evaluated faculty was the sole responsible anesthesiologist as well as cases where a resident was present. During perioperative cases anesthesiologists also work together with anesthesiology assistants. Anesthesiology assistants carry out the orders of the anesthesiologist and follow routine protocols. However, the anesthesiologist is ultimately responsible for the quality of care that is delivered. On the other hand, faculty and residents share the responsibility for clinical tasks whereby residents usually take on the primary responsibility for patient care while faculty supervise from a distance.

Measures

Faculty's teaching performance and role modeling status. Teaching performance was evaluated using the System for Evaluation of Teaching Qualities (SETQ) questionnaire, which has been validated for use in anesthesiology in the Netherlands.[16] SETQ evaluates clinical teaching among supervisors in five domains using 22 items: creating a motivating learning climate (seven items), displaying professional attitude towards residents (three items), communicating learning goals (four items), evaluating residents (four items), and providing feedback (four items). Overall teaching quality was evaluated with the item: "Overall I give this supervisor for his/her teaching qualities the score…" (1-10). All items are evaluated on a 5-point Likert scale, ranging from "Totally Agree" to "Totally disagree" with an additional option "I cannot judge" (Appendix, Table A1). Additionally, residents were asked to rate the clinical supervisor using as a role model with the item: "This supervisor is for me a … role model as a future anesthesiologist" rated on a 5-point scale: 1 = poor, 2 = fair, 3 = average, 4 = good, 5 = excellent. For each clinical supervisor we calculated average teaching score using the mean of all 22 items, as well as the global teaching score (item: "Overall I give this supervisor for his/her teaching qualities the score…") and the role modeling score (item: "This supervisor is for me a role model as a future anesthesiologist") by calculating the average score across all evaluations for each faculty for each evaluation period. Based on previous literature, we included faculty with at least two resident evaluations.[19] Additionally, we checked reliability of scores in our sample using intraclass correlations (ICC), which correspond to the proportion of the total variance in scores that can be explained by differences between faculty.

Clinical performance measures. For each evaluated faculty member we identified all perioperative clinical cases in which they were involved as a sole anesthesiologist or a clinical supervisor of a resident in the period six months after the teaching evaluation. We scored each case on a number of pre-identified quality measures, which were developed based on national norms, previous literature and locally-derived evidence-based protocols.[17,18,20-22] We defined two outcome measures (intra-operative temperature management and pain management) and two process measures (neuromuscular monitoring prior to extubation in the operating room and post-operative nausea and vomiting prophylaxis) (Table 1). For each measure, we identified the patient population and calculated pre-defined indicators. Since faculty members regularly work with residents and anesthesiology assistants during perioperative cases, some of the variation in clinical measures will be explained by residents and anesthesiology assistants. Therefore, we calculated the ICC for every measure in order to identify the proportion of variance in performance that is due to differences between faculty as well as residents and anesthesiology assistants.

Covariates. We included proportion of female respondents and average year of training as they are known respondents' characteristics as covariates for teaching evaluations.[23] We also included faculty, sex, age, years in practice, and whether they participated in a two-day teacher training course. Since clinical measures defined should have been achievable in all patients in the pre-specified patient population regardless of the patients' co-morbidities, we did not include any patient related covariates. However, we included the year as a covariate in our statistical analyses to control for potential differences in patient-mix between the years.

Table 1. Clinical performance measures for the study of associations of faculty teaching performance and perioperative care quality in the period of 6 months after the teaching performance evaluation, 2010-2012

Clinical performance measure	Indicator	Study population
1. Temperature management (outcome measure)	1) At least one temperature was documented intra-operatively (Y/N); 2) Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op) (Y/N)	All surgeries longer than 30 minutes
2. Pain management (outcome measure)	First documented pain score post-operatively: 1) Numeric Rating Scale (NRS) score 0-10 (continuous) 2) NRS score in 3 groups: 0-3 = least pain 4-7 = less pain 8-10 = most pain (reference group)	All surgeries where pain was measured post-operatively
3. Prophylaxis for post-operative nausea and vomiting (process measure)	For each risk factor present, one prophylactic measure should be provided intraoperatively. Also, if there are no risk factors present, no prophylaxis is needed. (Local hospital protocol) Risk factors: Adults (≥18 years):[22] 1. Female gender 2. Previous history of Post-operative nausea and vomiting OR motion sickness 3. Non-smoker (also if stopped) 4. Post-operative use of opioids 5. Operation time >60 min Children (<18 years):[20] 1. Age >3 years 2. Operation time >30 min 3. Strabismus correction surgery Since over-treatment is worse than under-treatment, can divide outcome in 3 groups: 3 = following guideline (number prophylaxis medications given = total risk factors); 2 = overtreatment (more prophylaxis medications given than total risk factors); 1 = undertreatment (fewer prophylaxis medications given than total risk factors)	All surgeries
4. Neuromuscular monitoring (process measure)	1) Train of Four (TOF) ratio or TOF count was performed (Y/N); 2) TOF value >70 (Y/N) 3) TOF value >90 (Y/N)	All surgeries where a patient was intubated, received a neuromuscular block, and was extubated in the OR, excluding intra-operative deaths

Statistical analysis

Of the 768 returned SETQ evaluations, 3 evaluations were excluded because they were empty and 8 evaluations were deleted because faculty received just one evaluation (yielding 757 evaluations for further consideration). In the final dataset, 92 SETQ evaluations had >50% missing values and six items had >20% missing values. Overall, 18% of all values were missing. In order to maximize our sample size while taking into account the uncertainty due to missing values and improve the estimation of teaching performance, we chose to impute missing values using multiple imputation technique using resident sex and post-graduate year, as well as faculty sex, age, previous teacher training and years of experience as predictors.[23] We created 100 imputed datasets and increased maximum number of iterations to 10 to improve convergence.[24] Thereafter, we calculated the overall SETQ scores by calculating the mean of all items across all imputed datasets. Data were checked for outliers.

We conducted hierarchical panel analyses to account for the multilevel nature of the data where patients (level 1) were nested within faculty (level 2), including the year as a covariate in the fixed part of the model. We regressed each clinical performance indicator on faculty's overall teaching score or mean role modeling score, adjusting for relevant characteristics of respondents, faculty and patients. Faculty's global teaching score was used as a sensitivity analysis to check the stability of our results. Using Maximum Likelihood estimation method, we constructed either logistic or linear hierarchical random-intercept regression models for each indicator depending on whether the outcome variable was dichotomous or continuous, respectively. We chose compound symmetry covariance structure since we expected the strength of the association to stay the same over time. We started with a full model, excluding covariates that had estimates close to null, meaning they did not contribute to the model. In order to answer our second research question, we included an interaction term between resident and faculty to indicate if a resident was involved in the case. In all analyses, we applied random intercept models where patients were clustered within faculty. Since anesthesiology assistants work according to protocols we did not include them in the main analysis. However, we checked the sensitivity of our results by controlling for the anesthesiology assistant. We report results for cases with and without a resident present. Missing data were imputed using the mice package (version 2.25) in R statistical software (version 3.2.3). All other analyses were conducted in IBM SPSS Statistics version 24.0.0.1 (IBM Corp., Armonk, NY).

Results

Sample characteristics

Between 2010-2012, fifty-one residents (response rate 65%) filled out 757 SETQ evaluations of 54 faculty. Respondents' characteristics on SETQ are included in Table 2. During the study period, the faculty were evaluated on average 2.3 times (42 in 2010, 41 in 2011, 39 in 2012) and received on average 6.2 SETQ evaluations (median 5, SD 3.9). Faculty were on average 47 years old (SD 9.8), a majority (63.3%) had completed a teaching course and had on average 10 years of clinical experience (SD 8.8). Mean SETQ score was 3.6 (SD 0.6, range 1.9-4.6), mean global teaching score was 3.6 (SD 0.6, range 1.67-5.00) and mean role modeling score was 3.6 (SD 0.7, min = 1.67, max = 4.70). Differences between faculty explained majority of mean SETQ (64%), global teaching (56%) and role modeling scores (55%). Of over 15000 patient encounters, 43% were together with a resident. Differences between faculty explained 1.5%-16.6% of variance in outcome measures in the absence of a resident. When a resident was present, residents accounted for 1.3-15.7% of variance in outcome measures and faculty for 0.6-5.1% (Table 2).

Association between faculties' teaching, role modeling and clinical performance

Tables 3-5 report results of the hierarchical panel analyses for the associations of the mean SETQ score, mean global teaching score and mean role modeling score with clinical performance measures, respectively. Our results show that among patients where no resident was present, teaching faculty with higher mean SETQ scores were significantly more likely to check neuromuscular function (TOF ratio or count) in patients who were extubated in the operating room (OR 1.67, 95% CI 1.17-2.39). Moreover, these patients were more likely to have a good neuromuscular function before being extubated (TOF >70: OR 1.96, 95% CI 1.40 – 2.75; TOF >90: OR 1.94, 95% CI 1.40 – 2.68). Results of our sensitivity analyses using mean global teaching score as a predictor were similar (TOF value >70: OR 1.39, 95% CI 1.04 – 1.85; TOF value >90: OR 1.43, 95% CI 1.08 – 1.90). Additionally, in our sensitivity analyses using global teaching score as a predictor of teaching quality instead of the mean SETQ score, we found a significant association between mean global teaching score and overtreatment for postoperative nausea and vomiting prophylaxis (OR 1.49, 95% CI 1.08 – 2.04). This association was also present but did not reach significance when mean SETQ score was used as the main predictor of teaching quality. All associations disappeared when resident was present. We did not find any significant associations between faculty's mean role modeling score and any of the clinical performance measures. However, after controlling for the anesthesiology assistant in our sensitivity analyses, we found that patients of faculty with higher role modeling scores were also more likely to achieve TOF value >70 (OR 1.33, 95% CI 1.02 – 1.73) or TOF value >90 (OR 1.39, 95% CI 1.08 – 1.79). Again, in the presence of a resident this association was not significant. (Appendix, Tables A2-4) Significant covariates varied based on the type of outcome measure. Notably, residents' involvement was associated with a significantly higher likelihood of achievement of normothermia, measurement of TOF value or count, as well as TOF value >70 or >90 (results not shown).

Table 2. Variance in outcome due to differences between faculty and/or residents in cases where the faculty cared for the patient alone (with resident absent) and when a resident was present

	Resident absent		Resident present		
Clinical performance measure	**ICC faculty**	**ICC assistant**	**ICC faculty**	**ICC assistant**	**ICC resident**
1. Temperature management					
At least one temperature was documented intra-operatively (yes/no)	0.018	0.033	0.013	0.039	0.033
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op) (yes/no)	0.015	0.022	0.007	0.010	0.057
2. Pain management					
1=NRS score 0-3/ 2=NRS score 4-7/ 3=NRS score 8-10 (ordinal)	0.013	0.028	0.006	0.026	0.019
NRS score 0-10 (continuous)	0.015	0.034	0.010	0.034	0.013
3. Neuromuscular monitoring					
TOF ratio or TOF count performed (yes/no)	0.166	0.128	0.051	0.091	0.157
TOF value was >70 (yes/no)	0.138	0.115	0.043	0.058	0.130
TOF value was >90 (yes/no)	0.116	0.109	0.038	0.053	0.111
4. Post-operative nausea and vomiting prophylaxis					
1=Overtreatment/2=undertreatment/3=according to guideline (ordinal)	0.028	0.034	0.007	0.028	0.057

Abbreviations: ICC indicates intraclass correlation; NRS, Numeric Rating Scale; TOF, Train of Four

Table 3. The associations of faculty's teaching performance (mean SETQ score) with perioperative care quality

		Resident absent			Resident present		
Perioperative care measure	**N**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**OR**	**Lower 95% CI**	**Upper 95% CI**
1. Temperature management							
At least one temperature was documented intra-operatively (yes/no)[a]	7688	0.94	0.822	1.07	0.98	0.84	1.13
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op)[b]	7176	1.05	0.91	1.21	0.93	0.80	1.09
2. Pain management[c]	6645						
NRS score 0-3 (least pain)		0.87	0.66	1.14	1.06	0.80	1.41
NRS score 4-7		0.95	0.72	1.26	0.99	0.74	1.32
NRS score 8-10 (most pain)		ref	ref	ref	ref	ref	ref
3. Neuromuscular monitoring							
TOF ratio or TOF count performed[d]	3421	1.67*	1.17	2.39	1.11	0.77	1.61
TOF value >70[e]	3421	1.96***	1.40	2.75	1.10	0.79	1.54
TOF value >90[f]	3571	1.94***	1.40	2.68	1.10	0.80	1.52
4. Post-operative nausea and vomiting prophylaxis[g]	7729						
Overtreatment		1.43	0.99	2.07	1.00	0.69	1.44
Undertreatment		1.05	0.83	1.33	1.04	0.81	1.35
According to guideline		ref	ref	ref	ref	ref	ref
	N	***b***	**Lower 95% CI**	**Upper 95% CI**	***b***	**Lower 95% CI**	**Upper 95% CI**
5. Pain management[h]							
First documented pain score post-operatively (NRS score 0-10)	6645	0.09	-0.11	0.28	-0.12	-0.35	0.11

Abbreviations: SETQ indicates System for Evaluation of Teaching Qualities; OR, odds ratio; CI, confidence intervals; TOF, Train of Four; NRS, Numeric Rating Scale; ref, reference category; *b*, unstandardized linear regression coefficient

*$P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

[a] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[b] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[c] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[d] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, patient age, patient gender

[e] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[f] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[g] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[h] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

Table 4. The associations of faculty's global teaching performance (mean global teaching score) with perioperative care quality

		Resident absent			Resident present		
Perioperative care measure	**N**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**OR**	**Lower 95% CI**	**Upper 95% CI**
1. Temperature management							
At least one temperature was documented intra-operatively (yes/no)[a]	8097	0.94	0.84	1.04	0.99	0.88	1.11
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op)[b]	7176	0.99	0.88	1.13	0.93	0.81	1.06
2. Pain management[c]	6645						
NRS score 0-3 (least pain)		0.98	0.79	1.23	1.11	0.88	1.39
NRS score 4-7		0.99	0.79	1.24	1.01	0.80	1.26
NRS score 8-10 (most pain)		ref	ref	ref	ref	ref	ref
3. Neuromuscular monitoring							
TOF ratio or TOF count performed[d]	3421	1.36	1.00	1.84	1.03	0.76	1.40
TOF value >70[e]	3421	1.39*	1.04	1.85	0.99	0.75	1.30
TOF value >90[f]	3571	1.43*	1.08	1.90	1.03	0.79	1.35
4. Post-operative nausea and vomiting prophylaxis[g]	7729						
Overtreatment		1.49*	1.08	2.04	1.05	0.77	1.42
Undertreatment		1.05	0.85	1.29	1.09	0.88	1.35
According to guideline		ref	ref	ref	ref	ref	ref
	N	***b***	**Lower 95% CI**	**Upper 95% CI**	***b***	**Lower 95% CI**	**Upper 95% CI**
5. Pain management[h]							
First documented pain score post-operatively (NRS score 0-10)	6645	0.01	-0.16	0.18	-0.07	-0.25	0.11

Abbreviations: OR indicates odds ratio; CI, confidence intervals; TOF, Train of Four; NRS, Numeric Rating Scale; ref, reference category; *b*, unstandardized linear regression coefficient

$^{*}P < 0.05$; $^{**}P < 0.001$; $^{***}P < 0.0001$

[a] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[b] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[c] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[d] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[e] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist , gender case-mix of respondents on teaching evaluation, patient age, patient gender

[f] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[g] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[h] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

Table 5. The associations of faculty's role modeling (mean role modeling score) with perioperative care quality

		Resident absent			Resident present		
Perioperative care measure	**N**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**OR**	**Lower 95% CI**	**Upper 95% CI**
1. Temperature management							
At least one temperature was documented intra-operatively (yes/no)[a]	8097	0.99	0.90	1.09	1.03	0.93	1.15
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op)[b]	7176	0.99	0.89	1.10	0.95	0.84	1.07
2. Pain management[c]	6645						
NRS score 0-3 (least pain)		1.02	0.84	1.24	1.08	0.83	1.40
NRS score 4-7		1.00	0.82	1.21	1.02	0.77	1.34
NRS score 8-10 (most pain)		ref	ref	ref	ref	ref	ref
3. Neuromuscular monitoring							
TOF ratio or TOF count performed[d]	3421	1.15	0.88	1.50	0.93	0.71	1.23
TOF value >70[e]	3421	1.18	0.91	1.52	0.93	0.72	1.21
TOF value >90[f]	3571	1.25	0.98	1.59	1.03	0.80	1.31
4. Post-operative nausea and vomiting prophylaxis[g]	7729						
Overtreatment		1.28	0.98	1.68	0.97	0.74	1.28
Undertreatment		1.13	0.95	1.36	1.09	0.90	1.31
According to guideline		ref	ref	ref	ref	ref	ref
	N	***b***	**Lower 95% CI**	**Upper 95% CI**	***b***	**Lower 95% CI**	**Upper 95% CI**
5. Pain management[h]							
First documented pain score post-operatively (NRS score 0-10)	6645	-0.02	-0.17	0.13	-0.07	-0.23	0.10

Abbreviations: OR indicates odds ratio; CI, confidence intervals; TOF, Train of Four; NRS, Numeric Rating Scale; ref, reference category; *b*, unstandardized linear regression coefficient

$^{*}P < 0.05$; $^{**}P < 0.001$; $^{***}P < 0.0001$

[a] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[b] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year , gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[c] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[d] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[e] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

[f] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, patient gender

[g] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[h] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

Discussion

Main findings

In this study we examined the associations between anesthesiology faculty's teaching performance and role modeling status with clinical performance. Residents accounted for a larger proportion of variability in clinical outcomes than the faculty working together or alone in all measures studied. We found that, when working without a resident, higher teaching performance as well as role modeling scores were associated with better clinical performance on perioperative neuromuscular monitoring. While involvement of residents was associated with better performance on normothermia as well as neuromuscular monitoring, there was no association between better teaching performance or role modeling scores and any of the clinical indicators.

Explanation of findings

Our main finding is that anesthesiology faculty's teaching performance is associated with higher likelihood and quality of neuromuscular monitoring of patients prior to extubation, which disappeared when a resident was involved in the case. We also found that the studied neuromuscular monitoring indicators had about five times more of their variance explained by faculty, than any of the other measures in this study. In other words, faculty vary in their practices around neuromuscular monitoring more so than any other measures. Since very little variability exists between faculty on the other clinical measures, it could explain why we did not find any associations of faculty's teaching performance with the other indicators of clinical performance.

Another explanation could be due to differences in the teaching value inherent in the different clinical activities. Neuromuscular monitoring represents a high-stakes decision and a good "teaching moment" for anesthesiologists, the outcome of which can have serious consequences for the patient. This corresponds to the findings of Wenrich et al., who reported that teaching bedside clinical skills to pre-clerkship students had a positive impact on staff's own clinical skills through expanding staff's knowledge and skills, deconstructing the clinical experience and greater self-reflection, including greater awareness of being a role model.[11] It is possible that, in practice, only those faculty's clinical skills improve that receive the most attention during teaching situations, especially those involved in high stakes clinical situations. In our study, the decision to extubate a patient in the OR represents a high-stakes situation. Residents' perceptions of faculty's teaching performance may, therefore, be a better indication of faculty' own knowledge and skills, where higher scores indicate faculty's own deeper and broader clinical practice and knowledge. This positive impact of teaching on the teacher has already been widely demonstrated and used in peer teaching in medical education.[25,26]

Our second main finding is that, when a resident is present, faculty's teaching performance no longer appeared to predict clinical performance. This is not surprising since faculty and residents share the responsibility for clinical tasks whereby residents usually take on the primary responsibility for patient care while faculty supervise from a distance. In our study, residents contributed a larger proportion of variability in clinical outcomes compared to faculty working alone or together with a resident. Indeed, the presence of the resident in itself had a positive impact on several clinical performance measures. Since the variability between faculty was also quite low, it was possible that associations with faculty's teaching performance were too weak to detect. Additionally, our choice of predictor measures could also have impacted the results. In this exploratory single center study, we only studied overall teaching performance as an indicator of teaching quality. However, since SETQ is a multi-dimensional instrument, which covers several aspects of teaching from creating a supportive learning climate and having a professional attitude towards residents to evaluating residents and giving them feedback. Not all of these teaching aspects will equally contribute to clinical performance in a situation where faculty act as a clinical supervisor, rather than a clinical teacher. Previous research has pointed out clinical supervision of residents as an important aspect of ensuring and improving patient safety and quality of care in clinical teaching settings[27,28] where a great variability in supervision styles exists.[29,30] Therefore, we would expect that in a clinical setting, more patient-centered domains of clinical teaching, such as feedback and evaluation[31], would be more likely to be associated with clinical performance measures than those that are more educationally-centered domains of clinical teaching evaluations. Studies involving larger sample sizes and multiple centers are needed to further explore the relation of individual domains of teaching performance to clinical performance indicators.

Similar to our expectations and those expressed in the literature[5,7,14], we found an association between faculty's mean role modeling score and clinical performance (TOF value >70 and >90) only in our sensitivity analyses. Kravet et al. used frequency of being named as a role model by trainees as an approximation of a faculty's role modeling status, and found an a strong association with perceptions of clinical excellence.[14] However, in our study, we used a single 5-point Likert scale question to evaluate faculty's role modeling performance, which may be a less sensitive predictor of clinical performance than frequency of being named a role model. Similar to teaching performance, role modeling is a multi-dimensional concept whereby role models demonstrate a number of characteristics and behaviours, where good clinical skills are only one part.[32,33] We used a single item to evaluate role modeling status, which did not capture the complexity of role modeling or might have been interpreted differently by residents.[2,3] In medical education literature, several typologies of role models have been recognized, namely teacher/supervisor, physician and person.[3,34] Although the SETQ questionnaire used in this study does not distinguish between them, the new redesigned SETQ Smart questionnaire includes all of these typologies as a part of the questionnaire.[15] Additionally, unlike teaching behaviours, faculty are not always conscious of their role modeling behaviours. Effective role modeling requires the faculty to be aware of their role modeling and discuss it with their students, as well as make a conscious effort to articulate what they are role modeling and why.[7] For this reason, the overall teaching performance score measured by the multi-dimensional SETQ questionnaire may be preferable to single-item evaluations of role modeling in predicting faculty's own clinical performance.

Limitations

This study has some limitations. First, all measures of teaching performance, global teaching quality and role modeling status were on a 5-point Likert scale, which could have caused ceiling effects. SETQ Smart, a questionnaire for evaluation in anesthesiology residency teaching, was recently validated for use in six countries. This questionnaire uses 7-point Likert scale, which would have been preferable. However, the department in our study had used the SETQ Smart only once as a pilot and therefore SETQ was chosen for this study. Also, we did not assess the subscales of SETQ, all of which could potentially have unique contributions to faculty's clinical performance. Larger sample sizes with more centers are needed in order to explore the role of the individual domains of teaching performance. Since we used questionnaires to evaluate faculty's teaching performance, there is a possibility that our results of teaching evaluations still suffer from selection bias. We minimized the chance for selection bias by achieving an acceptable response rate of 65% and using respondent characteristics (resident gender and post-graduate year) as predictors during imputation of SETQ data as well as the main analysis. Nonetheless, collecting data on characteristics of non-respondents would have made our analyses more complete. Lastly, due the multi-factorial nature of clinical performance measures, it was impossible to control for all factors that could have confounded our results. In particular, registration of some quality measures is depended on other team members, such as the post-operative pain scores, which are usually registered by Post Anesthesia Care Unit nurses.

Future research

The results of this study should be repeated in other specialties and a larger multi-center sample. The different domains of teaching performance should be evaluated in relation to clinical performance in order to identify specific aspects of teaching that may be amenable to faculty development opportunities. Future research could clarify whether lack of variability in some clinical performance measures are due to faculty themselves or systemic reasons out of their influence. Although there has been a lot of attention for using patient outcomes as indicators of patient safety and quality of care, more research is urgently needed to identify physician-sensitive clinical outcomes, where differences among faculty and/or residents explain a high proportion of variance in outcome.[35,36] The creation of relevant measures of faculty's clinical performance could aid in further studies in this area while improving our understanding of the multifactorial nature of patient outcomes. Furthermore, other measures of clinical performance, such as multisource feedback instruments, should be investigated. Multisource (or 360-degree feedback) instruments are gaining popularity as valid and reliable tools to assess physicians' performance[37], making comparisons between specialties possible.

References:

1. Spencer J. Learning and teaching in the clinical environment. *BMJ*. 2003;326(7389):591-594.
2. Boerebach BC, Lombarts KM, Scherpbier AJ, Arah OA. The teacher, the physician and the person: exploring causal connections between teaching performance and role model types using directed acyclic graphs. *PLoS One*. 2013;8(7):e69449.
3. Boerebach BC, Lombarts KM, Keijzer C, Heineman MJ, Arah OA. The teacher, the physician and the person: how faculty's teaching performance influences their role modelling. *PLoS One*. 2012;7(3):e32089.
4. Balmer D, Serwint JR, Ruzek SB, Ludwig S, Giardino AP. Learning behind the scenes: perceptions and observations of role modeling in pediatric residents' continuity experience. *Ambul Pediatr*. 2007;7(2):176-181.
5. Paice E, Heard S, Moss F. How important are role models in making good doctors? *BMJ*. 2002;325(7366):707-710.
6. Maudsley RF. Role models and the learning environment: essential elements in effective medical education. *Acad Med*. 2001;76(5):432-434.
7. Cruess SR, Cruess RL, Steinert Y. Role modelling--making the most of a powerful teaching strategy. *BMJ*. 2008;336(7646):718-721.
8. Fluit CR, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med*. 2010;25(12):1337-1345.
9. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med*. 2004;19(9):971-977.
10. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med*. 2005;20(12):1159-1164.
11. Wenrich MD, Jackson MB, Ajam KS, Wolfhagen IH, Ramsey PG, Scherpbier AJ. Teachers as learners: the effect of bedside teaching on the clinical skills of clinician-teachers. *Acad Med*. 2011;86(7):846-852.
12. Tortolani AJ, Risucci DA, Rosati RJ. Resident evaluation of surgical faculty. *J Surg Res*. 1991;51(3):186-191.
13. Mourad O, Redelmeier DA. Clinical teaching and clinical outcomes: teaching capability and its association with patient outcomes. *Med Educ*. 2006;40(7):637-644.
14. Kravet SJ, Christmas C, Durso S, Parson G, Burkhart K, Wright S. The intersection between clinical excellence and role modeling in medicine. *J Grad Med Educ*. 2011;3(4):465-468.
15. Lombarts KM, Ferguson A, Hollmann MW, Malling B, Collaborators S, Arah OA. Redesign of the System for Evaluation of Teaching Qualities in Anesthesiology Residency Training (SETQ Smart). *Anesthesiology*. 2016;125(5):1056-1065.
16. Lombarts KM, Bucx MJ, Arah OA. Development of a system for the evaluation of the teaching qualities of anesthesiology faculty. *Anesthesiology*. 2009;111(4):709-716.
17. Ehrenfeld JM, McEvoy MD, Furman WR, Snyder D, Sandberg WS. Automated near-real-time clinical performance feedback for anesthesiology residents: one piece of the milestones puzzle. *Anesthesiology*. 2014;120(1):172-184.
18. Hyder JA, Niconchuk J, Glance LG, et al. What can the national quality forum tell us about performance measurement in anesthesiology? *Anesth Analg*. 2015;120(2):440-448.
19. Boerebach BC, Lombarts KM, Arah OA. Confirmatory factor analysis of the System for Evaluation of Teaching Qualities (SETQ) in graduate medical training. *Eval Health Prof*. 2016;39(1):21-32.
20. Eberhart LH, Geldner G, Kranke P, et al. The development and validation of a risk score to predict the probability of postoperative vomiting in pediatric patients. *Anesth Analg*. 2004;99(6):1630-1637.
21. Anesthesiology Performance Improvement and Reporting Exchange (ASPIRE). https://www.aspirecqi.org/aspire-measures. Accessed April 19, 2017.
22. Apfel CC, Laara E, Koivuranta M, Greim CA, Roewer N. A simplified risk score for predicting postoperative nausea and vomiting: conclusions from cross-validations between two centers. *Anesthesiology*. 1999;91(3):693-700.
23. Arah OA, Heineman MJ, Lombarts KM. Factors influencing residents' evaluations of clinical faculty member teaching qualities and role model status. *Med Educ*. 2012;46(4):381-389.
24. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci*. 2007;8(3):206-213.
25. Peters M, ten Cate O. Bedside teaching in medical education: a literature review. *Perspect Med Educ*. 2014;3(2):76-88.

26. Bene KL, Bergus G. When learners become teachers: a review of peer teaching in medical student education. *Fam Med*. 2014;46(10):783-787.

27. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad Med*. 2012;87(4):428-442.

28. Schumacher DJ, Slovin SR, Riebschleger MP, Englander R, Hicks PJ, Carraccio C. Perspective: beyond counting hours: the importance of supervision, professionalism, transitions of care, and workload in residency training. *Acad Med*. 2012;87(7):883-888.

29. Goldszmidt M, Faden L, Dornan T, van Merrienboer J, Bordage G, Lingard L. Attending physician variability: a model of four supervisory styles. *Acad Med*. 2015;90(11):1541-1546.

30. Farnan JM, Burger A, Boonyasai RT, et al. Survey of overnight academic hospitalist supervision of trainees. *J Hosp Med*. 2012;7(7):521-523.

31. Piquette D, Moulton CA, LeBlanc VR. Model of interactive clinical supervision in acute care environments. Balancing patient care and teaching. *Ann Am Thorac Soc*. 2015;12(4):498-504.

32. Wright SM, Kern DE, Kolodner K, Howard DM, Brancati FL. Attributes of excellent attending-physician role models. *N Engl J Med*. 1998;339(27):1986-1993.

33. Wright S. Examining what residents look for in their role models. *Acad Med*. 1996;71(3):290-292.

34. Ullian JA, Bland CJ, Simpson DE. An alternative approach to defining the role of the clinical teacher. *Acad Med*. 1994;69(10):832-838.

35. Lipner RS, Weng W, Arnold GK, Duffy FD, Lynn LA, Holmboe ES. A three-part model for measuring diabetes care in physician practice. *Acad Med*. 2007;82(10 Suppl):S48-52.

36. Holmboe ES, Weng W, Arnold GK, et al. The comprehensive care project: measuring physician performance in ambulatory practice. *Health Serv Res*. 2010;45(6 Pt 2):1912-1933.

37. Donnon T, Al Ansari A, Al Alawi S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: a systematic review. *Acad Med*. 2014;89(3):511-516.

Appendix

Table A1. Items in System for Evaluation of Teaching Qualities (SETQ) questionnaire (evaluated on a 5-point Likert scale)

Learning climate

1. Encourages residents to participate actively in discussions.
2. Stimulates residents to bring up problems.
3. Teaches residents time management.
4. Keeps to teaching goals; avoids digressions.
5. Motivates residents to study further.
6. Stimulates residents to keep up with the literature.
7. Prepares well for teaching presentations and talks.

Professional attitude towards residents

8. Listens attentively to residents.
9. Is respectful towards residents.
10. Is easily approachable during on-calls.

Communication of goals

11. States learning goals clearly.
12. States relevant goals.
13. Prioritizes learning goals.
14. Repeats stated learning goals periodically.

Evaluation of residents

15. Evaluates residents' specialty knowledge regularly.
16. Evaluates residents' analytical abilities regularly.
17. Evaluates residents' application of knowledge to specific patients regularly.
18. Evaluates residents' medical skills regularly.

Feedback

19. Regularly gives positive feedback to residents.
20. Gives corrective feedback to residents.
21. Explains why residents are incorrect.
22. Offers suggestions for improvement.

Table A2. Results of the sensitivity analysis controlling for the anesthesiology assistant in the association of faculty's teaching performance (mean SETQ score) and perioperative care quality

		Resident absent			Resident present		
Perioperative care measure	**N**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**OR**	**Lower 95% CI**	**Upper 95% CI**
1. Temperature management							
At least one temperature was documented intra-operatively (yes/no)[a]	7665	0.96	0.83	1.10	0.98	0.84	1.14
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op)[b]	7151	1.06	0.92	1.21	0.91	0.79	1.07
2. Pain management[c]	6628						
NRS score 0-3 (least pain)		0.90	0.64	1.25	1.08	0.85	1.36
NRS score 4-7		0.97	0.69	1.36	1.02	0.84	1.24
NRS score 8-10 (most pain)		ref	ref	ref	ref	ref	ref
3. Neuromuscular monitoring							
TOF ratio or TOF count performed[d]	3414	1.91*	1.32	2.77	1.21	0.83	1.77
TOF value >70[e]	3414	2.29***	1.61	3.25	1.22	0.86	1.72
TOF value >90[f]	3564	2.24***	1.59	3.14	1.18	0.85	1.64
4. Post-operative nausea and vomiting prophylaxis[g]	7703						
Overtreatment		1.39	0.95	2.02	1.00	0.68	1.463
Undertreatment		1.00	0.79	1.27	0.99	0.76	1.28
According to guideline		ref	ref	ref	ref	ref	ref
	N	***b***	**Lower 95% CI**	**Upper 95% CI**	***b***	**Lower 95% CI**	**Upper 95% CI**
5. Pain management[h]							
First documented pain score post-operatively (NRS score 0-10)	6628	0.05	-0.15	0.26	-0.02	-0.18	0.14

Abbreviations: SETQ indicates System for Evaluation of Teaching Qualities; OR, odds ratio; CI, confidence intervals; TOF, Train of Four; NRS, Numeric Rating Scale; ref, reference category; *b*, unstandardized linear regression coefficient

*$P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

[a] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[b] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[c] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[d] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[e] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[f] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[g] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender

[h] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender

Table A3. Results of the sensitivity analysis controlling for the anesthesiology assistant in the association of faculty's global teaching performance (mean global teaching score) and perioperative care quality

		Resident absent			Resident present		
Perioperative care measure	**N**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**OR**	**Lower 95% CI**	**Upper 95% CI**
1. Temperature management							
At least one temperature was documented intra-operatively (yes/no)[a]	8072	0.95	0.85	1.07	0.98	0.87	1.11
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op)[b]	7151	1.00	0.89	1.13	0.92	0.81	1.04
2. Pain management[c]	6628						
NRS score 0-3 (least pain)		1.01	0.80	1.27	1.12	0.89	1.41
NRS score 4-7		1.00	0.79	1.26	1.03	0.81	1.30
NRS score 8-10 (most pain)		ref	ref	ref	ref	ref	ref
3. Neuromuscular monitoring							
TOF ratio or TOF count performed[d]	3414	1.54*	1.12	2.12	1.14	0.83	1.57
TOF value >70[e]	3414	1.60*	1.19	2.15	1.11	0.84	1.47
TOF value >90[f]	3564	1.61*	1.20	2.17	1.11	0.84	1.46
4. Post-operative nausea and vomiting prophylaxis[g]	7703						
Overtreatment		1.47*	1.06	2.03	1.05	0.77	1.43
Undertreatment		1.02	0.82	1.26	1.05	0.85	1.31
According to guideline		ref	ref	ref	ref	ref	ref
	N	***b***	**Lower 95% CI**	**Upper 95% CI**	***b***	**Lower 95% CI**	**Upper 95% CI**
5. Pain management[h]							
First documented pain score post-operatively (NRS score 0-10)	6628	-0.01	-0.18	0.16	-0.06	-0.24	0.12

Abbreviations: OR indicates odds ratio; CI, confidence intervals; TOF, Train of Four; NRS, Numeric Rating Scale; ref, reference category; *b*, unstandardized linear regression coefficient

*$P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

[a] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender, anesthesiology assistant

[b] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[c] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[d] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[e] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist , gender case-mix of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[f] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender, anesthesiology assistant

[g] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender, anesthesiology assistant

[h] Variables included in model: mean global teaching score, interaction term (global score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

Table A4. Results of the sensitivity analysis controlling for the anesthesiology assistant in the association of faculty's role modeling (mean global teaching score) and perioperative care quality

		Resident absent			Resident present		
Perioperative care measure	**N**	**OR**	**Lower 95% CI**	**Upper 95% CI**	**OR**	**Lower 95% CI**	**Upper 95% CI**
1. Temperature management							
At least one temperature was documented intra-operatively (yes/no)[a]	8072	1.01	0.91	1.11	1.03	0.92	1.16
Initial temperature at emergence was ≥36 degrees Celsius (30 min before End Anesthesia time, or 15 min after Start Post-Op)[b]	7151	1.00	0.90	1.12	0.95	0.84	1.07
2. Pain managementc	6628						
NRS score 0-3 (least pain)		1.04	0.85	1.27	1.11	0.90	1.38
NRS score 4-7		1.01	0.82	1.24	1.04	0.83	1.30
NRS score 8-10 (most pain)		ref	ref	ref	ref	ref	ref
3. Neuromuscular monitoring							
TOF ratio or TOF count performed[d]	3414	1.27	0.95	1.68	1.00	0.75	1.33
TOF value >70[e]	3414	1.33*	1.02	1.73	1.02	0.78	1.33
TOF value >90[f]	3564	1.39*	1.08	1.79	1.09	0.85	1.40
4. Post-operative nausea and vomiting prophylaxis[g]	7703						
Overtreatment		1.25	0.95	1.65	0.97	0.73	1.29
Undertreatment		1.11	0.92	1.33	1.07	0.89	1.29
According to guideline		ref	ref	ref	ref	ref	ref
	N	***b***	**Lower 95% CI**	**Upper 95% CI**	***b***	**Lower 95% CI**	**Upper 95% CI**
5. Pain management[h]							
First documented pain score post-operatively (NRS score 0-10)	6628	-0.03	-0.19	0.12	-0.05	-0.22	0.12

Abbreviations: OR indicates odds ratio; CI, confidence intervals; TOF, Train of Four; NRS, Numeric Rating Scale; ref, reference category; *b*, unstandardized linear regression coefficient

*$P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

[a] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender, anesthesiology assistant

[b] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year , gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[c] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[d] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[e] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

[f] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, total years of experience as specialist, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, patient gender, anesthesiology assistant

[g] Variables included in model: mean role modeling score, interaction term (role modeling score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient gender, anesthesiology assistant

[h] Variables included in model: mean SETQ score, interaction term (SETQ score*resident), resident present (Y/N), year, teach-the-teacher training (Y/N), gender specialist, gender case-mix of respondents on teaching evaluation, mean post-graduate year of respondents on teaching evaluation, patient age, patient gender, anesthesiology assistant

CHAPTER 7

Discussion

Discussion

Rising healthcare costs and concerns about patient safety have called for greater accountability in post-graduate medical education (PGME). Up till now medical education research has almost exclusively focused on the quality of care delivered by residents, overlooking other potential associations with patient care quality and patient outcomes. The purpose of this thesis was to investigate how quality in PGME is related to the quality of care provided in clinical teaching departments, using the residency learning climate and teaching quality as starting points of our investigations. In order to answer the overall research question of this thesis, "How is quality of PGME related to indicators of care quality in clinical teaching departments?", we first investigated the quality of instruments used to evaluate the residency learning climate and inpatient experiences in clinical teaching departments. Validity and reliability of two key instruments to determine the learning climate and patient care experience in teaching departments was demonstrated in chapters 2 and 4. The association between indicators of PGME and health care quality was explored in chapters 3, 5, 6. In chapter 3, higher departments' learning climate scores were associated with greater odds of adverse perinatal outcomes, but not maternal outcomes, in 16 nontertiary obstetrics and gynecology teaching departments. These associations were largely driven by low Apgar scores. In chapter 5 departments' learning climate scores were positively associated with patients' experiences of feeling of safety and communication with doctors in 86 teaching departments across 15 specialties. The individual subscales of the learning climate showed positive as well as negative associations with patient care experience. Finally, in chapter 6, individual anesthesiology faculty's teaching performance and role modeling scores were associated with better performance on indicators of neuromuscular function in perioperative care. This association was however attenuated when faculty were working with a resident. Figure 1 presents the statistically significant findings from these chapters.

In this section we will discuss the major findings from the studies in light of theory and current literature. First, we will discuss the validity of the instruments used to measure the learning climate and patient care experience before discussing the insights from studies of association between learning climate, teaching quality and care quality. We will weave their lessons into a single story about the relationship between quality of residency training and quality of patient care in clinical teaching departments. We will then address implications for future research and practice, ultimately arguing for a need for greater integration of PGME and health care delivery in clinical teaching departments. Finally, we will conclude by reviewing the strengths and limitations of the approach taken in this dissertation.

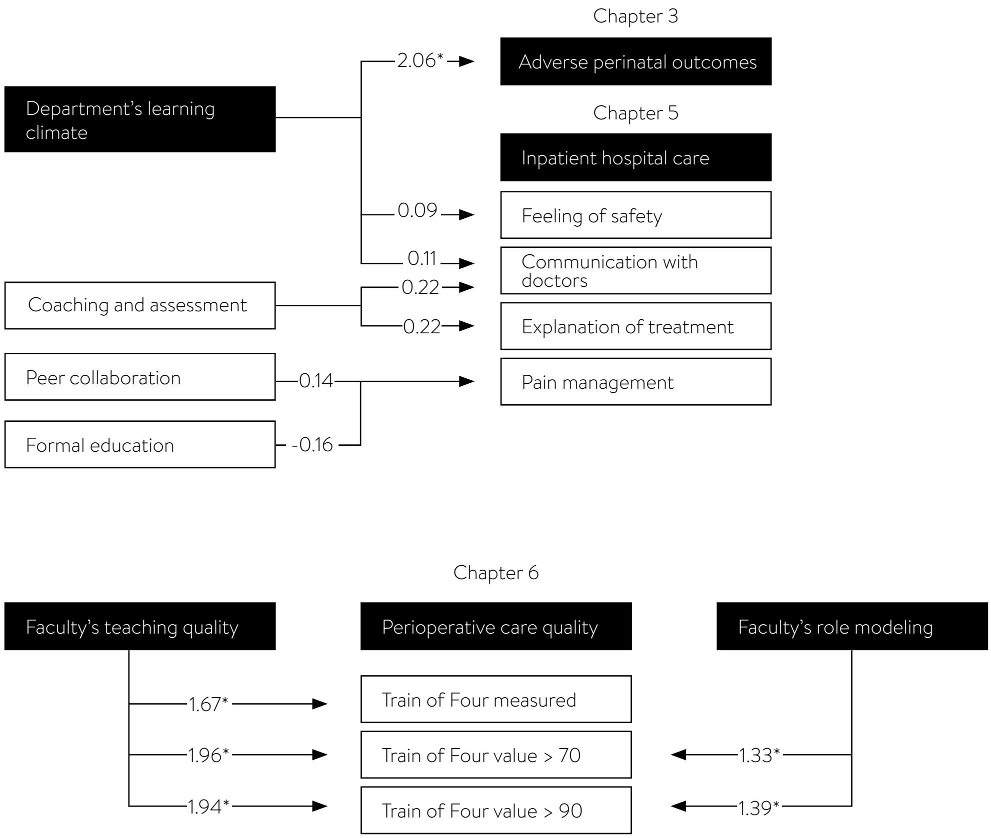


Figure 1. Overview of major findings of the association studies.
*Adjusted odds ratios, the remaining coefficient are adjusted unstandardized linear regression coefficients.

Validity and reliability of learning climate and patient care experience instruments

Valid and reliable instruments are needed to study the relationship between PGME and quality of care. In the first part of this thesis we set out to review the validity of instruments used to evaluate the learning climate and patient experiences of inpatient hospital care. Current conceptions of validity emphasize that the interpretation of scores for a particular use or uses of an instrument or test should be supported by evidence on the scoring of items as well as their generalization, extrapolation and implications (Box 1).[1,2] Possible implications of scores for quality of patient care, therefore, represent only a part of the validity argument, for which the other aspects of validity serve as a foundation. In this section, we review the evidence on the validity and the reliability of the D-RECT and the CQI Inpatient Care questionnaire. These considerations are required for making valid conclusions about their associations with predictors, exposures or outcomes.

Box 1. Examples of types of evidence in the four components of Kane's validity argument.

Scoring: evaluation of items and their wording, appropriateness of raters for assessment, response process including having sufficient opportunities to observe performance being evaluated

Generalization: sampling strategy, required sample size, reliability (Cronbach's α or generalizability theory)

Extrapolation: identification of domains using factor analyses, differential functioning of items, identification of relationship with other variables, effects of interventions

Implications: actions based on assessment results, intended and unintended consequences of assessment, setting a pass/fail standard

In chapter 2 we reviewed the original 50-item version of the D-RECT and the applicability of the items in light of the modernization of postgraduate medical education in the Netherlands. By removing the items of the D-RECT that were no longer applicable in the current context or had low factor loadings in the exploratory factor analysis (EFA), we reduced the construct-irrelevant variance in D-RECT scores improving its validity in scoring.[3] In shortening the D-RECT we also provide evidence of generalization demonstrating good reliability of the D-RECT domains. Next, we contributed evidence for D-RECT's extrapolation by investigating whether similar domains could be measured on the department level (its unit of analysis). The results of the confirmatory factor analysis (CFA) showed that the resulting 9-factor 35-item D-RECT had an acceptable fit and low inter-scale correlations on the department level meaning that aggregated scores can be interpreted similarly for department level evaluations as for individual residents' evaluations. Furthermore, the nine factors identified in our validation procedure could be mapped to affective, cognitive, and instrumental theoretical components typical to all organizational climates. In summary, in chapter 2 we demonstrated that the learning climate of a teaching department can be validly and reliably measured using a shorter questionnaire and fewer respondents, which has an advantage over the original 50-item version supporting its generalization and extrapolation. In other studies in chapters 3 and 5, we showed, using intraclass correlations (ICC), the reliability of D-RECT group means, which is important when studying differences between departments' learning climate scores. In chapter 3, the group level ICC (or ICC(2)) was 0.69 meaning that 69% of the variance in the overall D-RECT scores that can be explained due to differences between departments. In chapter 5, ICC(2) was 0.76 for the overall score and 0.57-0.83 for the subscale scores. ICC(2) of 0.70 is considered to be adequate in organizational climate research.[4] This additional validity evidence supports the use of the D-RECT to study differences between learning climates in clinical teaching departments.

In contrast to the D-RECT which is mainly used for quality improvement in PGME, the CQI Inpatient Hospital Care has been widely used by insurance companies to compare hospitals and make decisions about contracting of healthcare services.[5] However, since its original validation, the questionnaire has been shortened, but was not validated on the level of its common uses (department and hospital). Therefore, our validation study (chapter 4) provides an essential step towards building a sound validity argument, including the internal structure on the individual and department levels (extrapolation) as well as its reliability on individual, department and hospital levels (generalization). We found a less acceptable fit and a significant overlap between domains when scores of the CQI Inpatient Hospital Care questionnaire were aggregated to the department level. This suggests that although individual patients perceive some domains as being separate, these differences in domains are less important at the department level, which has implications for interpretation of scores on different levels of analysis. Furthermore, the evidence of generalization shows that some domains are not as reliable, indicating they require more attention in the future and their scores should be interpreted with caution. In order to support the current use(s) of the CQI Inpatient Hospital Care, additional validity evidence is needed for all four components of the validity argument on both patient and aggregated levels. We considered the use of the questionnaire, however, appropriate for the multilevel analyses in our studies, because only disaggregated patient experience data on individual level was included, which showed good CFA fit parameters in the validation study presented in chapter 4. Furthermore, when we compared the fit parameters for individual specialties, including surgery, cardiology, internal medicine, and obstetrics and gynecology, the good fit further supported its applicability for use in the subsequent association study in chapter 5.

The relationship between PGME quality and quality of care in clinical teaching departments

Having reviewed the validity of the instruments used in the thesis, we will now turn the discussion to our findings regarding the relationship between PGME quality and quality of care in clinical teaching departments. We can draw five main conclusions from the studies in this thesis about the relationship of PGME with quality of care and patient outcomes in clinical teaching departments, which are summarized in Box 2. In the paragraphs below we will elaborate on each of these conclusions in light of the validity of the instruments and organizational climate literature.

Box 2. Perspectives on the relationship between PGME quality and quality of care.

1. Valid and reliable instruments on the appropriate level of use are the prerequisite for making solid conclusions about the relationship and the domains being studied. (chapters 2 & 4)
2. A better residency learning climate in clinical teaching departments may place patients at an increased risk of poor outcomes, especially in urgent care situations that require a timely response. Nonetheless, a reciprocal relationship between learning climate and patient outcomes may (co-)exist since poor patient outcomes may impact residency learning climate perceptions. (chapter 3 & 5)
3. Those domains of the residency learning climate most closely related to direct patient care (coaching and assessment with doctors' communication and explanation of treatment, and peer collaboration with pain management) are favourably associated with inpatient care experience, and vice versa, those domains furthest removed from direct clinical care (formal education and pain management) associate unfavourably with inpatient care experience. (chapter 5)
4. Faculty's teaching quality and role modeling evaluations positively predict their clinical performance in high stakes clinical situations with highest clinical impact on the patient and teaching value. Residents' involvement in clinical cases, in turn, is associated with better performance on clinical measures. (chapter 6)

Berenson et al. have argued that to achieve their full potential, validity and reliability evidence of performance measures should reflect their level of use.[6] In our studies we used variables on the level of a department (learning climate and its domains), physician (teaching quality) or patient (patient experience domains, obstetrical complications, or perioperative care outcomes). The validity and reliability evidence of the variables' level of aggregation can improve the interpretation of the associations. For instance, the relatively high inter-scale correlation between patient experience subscales pain management and communication with nurses (chapter 4) suggest that patients' perceptions of how their pain is managed is closely related to their perceptions of nurses-patient interaction. Therefore, a negative association between the learning climate subscale formal education and pain management (chapter 5) may reflect patients' experiences of pain management related to nurses' performance in teaching departments that prioritize formal education for residents. These associations provide cues for the inner workings of teaching departments by indicating areas of interaction between subscales of the learning climate and aspects of healthcare delivery relevant for patient care outcomes. When aggregated to the level of a department, the patient experience domain communication of doctors overlaps with the domain explanation of treatment. Although in our analyses we only used individual-level domain scores, communication with doctors and explanation of treatment may be similar if interpreted from the level of the department. This potential overlap between domains of CQI Inpatient Hospital Care therefore needs to be kept in mind when interpreting the associations. On the other hand, D-RECT domains exhibited good psychometric properties on both the aggregated (department) as well as the individual (resident) levels, which supports its use on both levels.

Our main finding, however, is the negative association of residency learning climate and its domains with patient outcomes and experiences of care. In chapter 3 we found that teaching departments with highest learning climate scores were more likely to have an adverse perinatal event, most likely a low Apgar score. Furthermore, in chapter 5 we found a negative association between the D-RECT domain formal education and patients' experience of pain management in clinical teaching departments. These findings may reflect a potential unintended negative consequence of a supportive learning climate on patient care quality, whereby patients may experience unnecessary risk in those teaching departments with better learning climates. For example, since Apgar scores reflect the quality of the neonatal resuscitation, it is possible that in a resuscitation situation, prioritizing residents' over patients' needs may be better for learning for the resident but worse clinical outcomes for the patient. Additionally, we found similar tensions between patient- and resident-centeredness in examining the associations with inpatient care experience. Learning climate domains that were more focused on patient care (i.e. coaching and assessment and peer collaboration) were associated with improved patients' experiences, while the domain formal education, which was furthest removed from direct clinical care (at least in the short-term) was associated with worse perceptions of care.

This apparent tension could be explained by the fact that, in contrast to generic climates[4], the residency learning climate aims to improve residents' learning and well-being and reflects the extent to which work and training are integrated and tailored to individual residents' learning needs.[7] Assessment of learning climate is therefore interested in evaluating those elements of the clinical learning environment that serve these goals, which are in themselves not necessarily aimed at ensuring patient safety. Since learning in clinical environments and healthcare delivery systems share similar resources (for example, teaching time and space), tensions between patient- and learner-centeredness may arise between learning and patient care when activities coincide but are directed at different goals.[8] In other words, resident-centeredness exemplified by higher learning climate scores in teaching departments could potentially create tensions with direct patients' needs. In addition, in all cases the negative associations involved clinical situations where there was a sense of urgency or a need for a quick response. For instance, in obstetrics and gynecology, Apgar scores were reflective of resuscitation efforts in neonates (chapter 3), inpatient care pain management was reflective of timely response to patients' pain (chapter 5). What these outcome measures have in common is the need for timely and efficient response to the patients' clinical condition.

On the other hand, a negative association between learning climate and patient outcomes and experiences of care could also signify a reciprocal relationship between outcomes and climate perceptions. Beus et al. found that there was a two-way relationship between safety climate and injuries, whereby the relationship from injuries to climate was stronger than from climate to injuries.[9] In another study, Schneider et al. found a reciprocal effect of customer service perceptions on the service climate, but it was smaller than the effect of service climate on customer service perceptions.[10] As Beus explained, a particular outcome, such as a workplace injury, a customer's display of appreciation or an adverse event, can affect a workplace climate through influencing employees' reflections of their experiences with related workplace policies, procedures and practices.[9] While our initial hypothesis was that a supportive learning climate would be associated with better patient outcomes, it is equally possible that occurrence of adverse perinatal outcomes could potentially be a good "teaching moment" or constitute a trigger for residents' learning, thereby influencing their perceptions of their learning climate. The work of Teunissen in the area of workplace based learning supports the role of clinical activities as a starting point for residents' learning.[11] Along with Dornan and Watling, Teunissen argues that learning and clinical work are inseparable in the clinical workplace because they are a part of the same process of engaging in practice.[11-13] Although it may be impossible to separate the causal effects of learning/teaching and clinical care in practice, it may be possible to examine the strengths of these bi-directional relationships from each side. We would expect then that in certain situations one may predominate over the other.

Not only did we find that in certain situations the learning climate can be detrimental to patient care, we also found positive associations between PGME and care quality. These associations seem to be found in clinical situations when teaching, learning and direct care overlap. This is in line with organizational climate literature, in which the associations are stronger when customer contact was high and outcome less tangible.[4] The strong predictive value of the residency learning climate of patients' experiences of communication with doctors can be at least partly explained by the fact that residents are often the first point of contact for hospitalized patients, while faculty take on a supervising role.[14] Furthermore, when faculty's involvement and communication about direct patient care intensifies through coaching and assessment of residents, we see that departments with higher scores on coaching and assessment also had higher scores on patients' perceptions of communication of doctors. Similarly, the overlap between teaching, learning and practicing was also evident in the study of anesthesiology faculty's teaching quality, role modeling and clinical performance (chapter 6). We found that faculty's teaching quality and role modeling positively predict their clinical performance in the clinical situation with highest clinical impact on the patient and teaching value, namely monitoring of neuromuscular function prior to extubation, a clinical situation that also represents an important "teaching moment" for faculty.

Ultimately, learning climate is only a part of the climate of a clinical teaching department, which in turn is a part of a larger hospital organization. Therefore, the ultimate interpretation of the findings rests on our holistic understanding of the complexity of the clinical learning environment in which training takes place. Although their approaches may differ, the goals of PGME and high quality care are indistinguishable. Therefore, in order to fully understand the relationships between educational quality and clinical care in a clinical teaching department, we must conceptualize clinical teaching departments as a holistic system.[15] In a such a holistic system a (dis)balance may exist between patient care and training, whereby demands of patient care and goals of training sometimes compete for scarce resources.[16] It is then not surprising that efforts to improve PGME quality can have unintended negative consequences for patient care in clinical teaching departments. Perhaps, the most striking discussion in recent years has centered around the Duty Hour reform in the United States, which attempted to reduce fatigue and improve patient safety by limiting residents' duty hours. The resulting extensive research of the effect of duty hours on patient care, however, provided mixed results[17] leading to the realization that duty hours have multiple complex relationships through which they effect patient care in the clinical learning environment.[18] Therefore, optimizing both patient care and education cannot be achieved by focusing on a single variable, but through identifying and optimizing many relevant variables at the same time.[19] What is needed, however, is the integration of PGME and healthcare delivery in practice. The following section addresses this issue by discussing where we stand now and what can potentially be done.

The way forward: integrating PGME and health care delivery

Leaders in medical education have already recognized the need to integrate PGME and healthcare delivery systems in order to improve care quality and, ultimately, the health of patients and populations. In a recent perspective published in *Academic Medicine*, Sklar and colleagues plea for "the importance of aligning the purpose and goals of medical education with those of the health care delivery system."[20] Similarly, Wong and Holmboe have called for a reorganization of PGME to better align educational and clinical contexts to benefit both the learner and the patient.[21] In both instances, the authors propose a framework centered around the patient and care quality, where learning is not an end in itself but is subservient to this greater goal. The path to achieving this still remains unclear and is perhaps a topic for another dissertation (or more). Nonetheless, a number of potential areas of integration have been identified in the literature, which can form the stepping stones for the future. Below, we will briefly explain potential areas for integration that have already been identified in the literature and discuss the necessary conditions for achieving this in practice.

The possible areas of integration include consolidating quality assurance systems, aligning undergraduate and postgraduate curriculum, as well as re-imagining teaching and learning in the clinical learning environment. Quality assurance and quality improvement in PGME is mostly driven by entities such as educational institutions or educational regulatory bodies, which require reports on specific educational activities for purposes of financial reimbursement and accreditation of training programs. This is in contrast to quality assurance in healthcare, which is driven by public accountability and demands of the public.[22] Both quality assurance methods rely on the same principles, however, a common framework agreed upon by key stakeholders from both systems will narrow down the focus and streamline quality improvement efforts and lead to a natural integration of both systems.[22] Second, medical training should be better aligned with health care practice. For example, medical students would benefit from early exposures to clinical practice, quality management in healthcare, quality improvement and patient safety, health systems design and health economics.[20] Also, simulation has the potential to fill in the gaps on all levels of training that clinical environments cannot provide by honing the necessary skills trainees need for full participation in the healthcare delivery system.[23] Third, health care practice can also be transformed by stimulating resident-faculty co-learning, a process that involves educating residents alongside faculty, which has a number of advantages by bringing faculty and residents closer together and disseminate latest developments in healthcare quality and safety much quicker.[21,24] In their review, Sholl and colleagues identified a number of interventions for balancing education and practice, as well as their outcomes and mechanisms.[25] According to the authors, the most common interventions were ward round teaching, protected learning time and continuous professional development.[25] Finally, "service" – often regarded by residents as not contributing to their learning – should be recognized as the essential component of learning and quality improvement in residency training.[26,27]

In order to achieve such transformation, a number of existing challenges that prevent meaningful change must be tackled. These include addressing the silos of leadership[28], lack of a common vision for medical education and health care delivery[29] and lack of organizational support to facilitate change.[30] Currently, the linkages between medical education and health care delivery are weak because existing silos of PGME and health care quality do not allow leaders in medical education and leaders in health care quality in the same institutions to communicate well and learn from each other.[28] Thus, creating and supporting leadership positions that focus on bridging education and care quality in clinical learning environments can facilitate collaboration between these inherently connected systems.[25] Furthermore, creating a common organizational vision on both hospital and departmental level as well as in the interaction between the medical school and the hospital, and an environment characterized by clinical efficiency, responsiveness to problems, and collegiality[29] while tackling old discourses not aimed at meaningful change[30] would be necessary for a successful transition towards more integration of the medical education and health care delivery systems. Lastly, consolidating quality assurance systems, aligning undergraduate and postgraduate curriculum, and re-imagining teaching and learning in the clinical learning environment requires a holistic understanding of how PGME interacts with patient care and patient outcomes. However, this understanding is in turn limited by current lack of sufficient validity and reliability evidence on the instruments' level of use, lack of meaningful performance measures to evaluate individuals' and teams' clinical performance, as well as lack of insight into the mechanisms behind the associations. The following section aims to set out an agenda to address these research gaps. Broadly the research agenda is three-pronged: extending the validity argument for existing performance measures, developing a set of meaningful performance measures to evaluate both quality in healthcare delivery and medical education, as well as uncovering mechanisms behind the associations.

The way forward: a research agenda

Extending the validity argument

The key claims, assumptions, and inferences of the existing instruments used to measure quality in both PGME and healthcare should be continuously tested.[1] As Kane argued, the validity argument is only as strong as its weakest link. The proposed use of the instrument and the weakest area of evidence will guide the choice and prioritization of the goals and methods for subsequent validations.[1,2] Both the D-RECT and CQI would benefit from additional evidence on the performance of different scoring rubrics (scoring), effects of paper and web-based evaluation (scoring), effectiveness of actions based on assessment results (implications), intended and unintended consequences of testing (implications), and setting pass/fail standards (ROC curve, implications).

In order to understand the D-RECT results better, the response process should be better understood as well as the relevant department and resident characteristics that determine perceptions of the learning climate. Also, the effect of variation of responses should be investigated. Depending on the purpose of the D-RECT (summative or formative), a different (higher) cutoff for reliability may be required. The associations of the D-RECT with patient outcomes and patient care experience presented in this thesis may require additional validity evidence to be collected for the D-RECT, if it is to be used in high-stakes decisions (for example, accreditation of residency training programs). In particular, scoring (to better understand how the items on the D-RECT are scored by residents) and generalization (to further reduce variance in D-RECT scores and improve precision in order to allow departments to be ranked) may need to be reviewed. High-stakes decisions will require additional evidence of relationships with other important variables in the clinical learning environment, such as identifying the features of residency training programs associated with high and low D-RECT scores (extrapolation) and, most importantly, the implications of these scores for patient care.

On the other hand, CQI Inpatient Hospital Care questionnaire can benefit from improvements in the domains that demonstrate significant overlap, improvements in the response rates by collecting more information on the non-responders and understanding the reasons for non-response. The results of the CQI Inpatient Hospital Care questionnaire are also used by key stakeholders and the public to choose healthcare services. To this extent, validity evidence should be extended to support their use in practice. Finally, replication of our findings is needed in other clinical contexts and with improved instruments in order to exclude any spurious associations related to the questionnaire's validity issues.

Developing a set of meaningful performance measures

Aligning the goals of PGME and health care delivery requires that meaningful clinical performance measures are defined on both organizational (hospital and department) as well as individual levels. Measuring clinical performance of institutions, departments and individuals has become an unmissable part of quality control and improvement in health care and medical education.[31-33] All teaching hospitals engage in teaching, health care delivery and research, yet the balance and priority given to each activity differs between institutions. Therefore, decisions about overall performance of institutions or departments and individual healthcare providers (be it faculty or residents) should be made based on results from a range of performance indicators.[22] Performance measures should not only demonstrate their impact on health of the populations, but also be meaningful to the providers, departments and institutions using them to facilitate meaningful changes.[6] Electronic Health Record (EHR) data collected in registries and quality improvement databases have a great potential for creating meaningful patient-centered composite performance measures that reflect health care systems' and individual providers' contributions to patient and population outcomes, which can be used by healthcare systems, providers and patients to continuously monitor and improve performance.[32,34,35] The identified measures should be valid, reliable as well as meaningful to their end-users, in order to increase acceptance and avoid performance measurement from becoming an administrative burden devoid of meaning.[36-38] Similar to health care delivery, there is a growing need for public accountability and transparency in PGME. There is a need for a consensus on PGME outcomes on individual, program, as well as national levels. In an editorial in *New England Journal of Medicine*, Weinstein argues that a centralized system that systematically collects data on residents, graduates, as well as programs' performance is needed in order to be able to assess the impact of graduates, programs, and the PGME system.[33]

Uncovering mechanisms behind associations

Both quantitative and qualitative methods are needed to uncover the mechanisms behind the associations between learning climate, teaching quality and patient care outcomes. One approach is to study differences related to healthcare and educational processes in teaching departments that score consistently high on the D-RECT and have lower adverse perinatal outcomes and those scoring lower on the D-RECT and have higher adverse perinatal outcomes. An alternative is to study qualitative differences between programs that score consistently high or low on the D-RECT and link them to processes of care. Understanding how the D-RECT is filled in and the potential meaning of various answer patterns (i.e. residents' discordance/concordance on the D-RECT, clustering of high/low scoring domains together, response rate, effect of post-graduate year and other characteristics) may provide additional clues.

Moderators of the learning climate-outcome relationship should also be identified. These can include the types of outcome with less tangible outcomes having stronger associations, or other factors like the complexity of the patient's condition.[39] Finally, climate strength should be researched as a moderator of the relationship between climate and outcomes. Climate strength is defined as the extent of consensus in residents' perceptions of the practices, policies and procedures of a department.[4] In previous studies, climate strength has been found to moderate the relationship between climate and outcome either through more consistent behaviour of residents or through ensuring a more reliable mean resulting in greater validity of associations.

The antecedents of learning climate should also be researched. In service climate research, for example, leadership and leaders' personal characteristics play an important role in determining climate.[4] Important antecedents of the learning climate may include faculty's teaching[40] and personal characteristics[41], as well as program director's leadership style and department's resources.[42] Our research contributes another piece of the puzzle by establishing an association of faculty's teaching performance and their own clinical performance (chapter 6). What is needed, however, is a better understanding of how faculty's teaching performance contributes to the departments' learning climate and clinical performance.

Ultimately, the study of the complex associations between PGME and care quality requires that researchers also adopt a "complexity lens". In 2010, Regehr called for medical education researchers to move away from the search for simple and generalizable conclusions towards a better understanding of the unique contexts in which education and clinical care take place.[43] This means that a fundamental shift in medical education research towards complexity is required. While there are many ways to study complexity, several conceptual frameworks are particularly suited to the medical (educational) context: activity theory, systems thinking and complexity theory. We will shortly explain these theories and how they can be applied in this context.

Activity theory. Activity theory is a sociocultural theory developed originally by Vygotsky and Leontiev and further elaborated upon by Engeström.[44] In medical education research, activity theory has already been applied to studying complexity of collaborative practice on a distributed transplant team[45] and cultural complexity in globalized medical education.[46] Activity theory, developed by Engeström, is particularly adept at exploring the tensions between systems, which could be particularly useful in studying the tensions between patient care and education in clinical learning environments.[47] For instance, de Feijter and colleagues used activity theory to investigate how workplace-based learning in undergraduate medical education creates tensions with safe patient care.[8] Furthermore, activity theory can bring about change in the practices and structures of organizations.[48] Skipper used the Change Laboratory approach in order to integrate activities of an outpatient clinic, patient care and resident training by identifying and tackling the tensions and contradictions arising in day-to-day activities of the outpatient clinic with the double aim of treating patients and training residents.[49] As a result, activity theory is particularly well suited for exploring tensions between the medical educational system and health care delivery system. For example, activity theory would be helpful in studying how organizational processes in both systems have contributed to higher odds of adverse perinatal outcomes in clinical teaching departments found in chapter 3. Additionally, since the outcome is usually clearly defined, once such tensions have been identified and alleviated, changes in both systems, as well as the outcome, can also be measured.

Systems thinking. Having its roots in engineering, systems thinking has recently found its application in studying complexity in medical education research. Systems thinking attempts to understand how the environment constrains, affords or modifies goal-directed behaviour.[50] In contrast to activity theory where an outcome is usually clearly defined, systems thinking analyzes the problem from different perspectives first, which then can help design specific interventions. This is particularly useful in situations where multiple perspectives and conflicting interests exist.[51] The advantage of the systems thinking approach is that it defines the problem not on the level of the individual, but "the level of the situation at which the problem is manifested".[52] Another application of systems thinking is in program evaluation, where systems thinking approach can be applied to study the intended and unintended consequences of PGME. For instance, Rojas and colleagues applied systems thinking using a mixed method approach to study intended and unintended consequences of a surgical boot camp in a surgical residency program.[53] In our example, systems thinking approach can be used to identify the intended and unintended consequences of departments' efforts to improve the residency learning climate.

Complexity theory. Both healthcare and educational systems are considered to be complex adaptive systems.[54-59] A complex adaptive system can be defined as a "collection of individual agents with freedom to act in ways that are not always totally predictable, and whose actions are interconnected so that one agent's actions changes the context for other agents".[58,59] Complex adaptive systems are ubiquitous and have been found in computer modeling[60], the human body[61], organizations[62] and social life.[63] Complexity theory recognizes that tensions are inherent in complex adaptive systems, whereby interaction of complex systems (or their components) can lead to unexpected, non-linear, or even paradoxical, effects.[59,64] Complexity theory considers the global pattern of all inter-dependent parts and how their interactions produce certain outcomes.[65] Complexity theory can, therefore, be applied to study how micro-level behaviours and phenomena, such as individual agency and workplace affordances, may interact with macro-level phenomena, such as structural policies to produce outcomes that are not fully predictable or controllable. As a result, complexity theory can be particularly useful in exploring how processes of a clinical teaching department align around the department's shared vision of patient-centeredness or learner-centeredness, and how they ultimately affect individual interactions and decisions within those systems. In conclusion, it is evident from this discussion that we have the theoretical perspectives necessary to be able to carry out the proposed research to address the complexity of the clinical learning environments.

Strengths and limitations

The strengths of this thesis lie in its timeliness and relevance, its originality and methodological rigour. This thesis comes at a time when competency-based medical education continues its march around the world, in the process changing the face of postgraduate medical education with new developments such as Entrustable Professional Activities (EPA's) and individualization of training duration. While focusing on the long-term goal of educating physicians to deliver high quality care, this thesis makes us think about the value that these advances have for patient care in the short-term and how they may interact with patient care processes in clinical teaching departments. The results are, therefore, highly relevant for both patients and patient care delivery systems as competency-based medical education contuse to be widely implemented across clinical practices. Previous research on this topic has been limited by smaller sample sizes, which limited its generalizability. This thesis uses 'big data' in coming to its conclusions, a trend that will only continue to develop in the future as researchers will use large datasets to search for links between medical education and patient care. Therefore, the lessons from the studies in this thesis can be used for advancing future research on this topic.

Another strength of this thesis lies in the originality of its ideas and methods. Until now, patient outcome research in medical education has focused on evaluating outcomes of individual physicians, graduates or residents.[66-68] Notwithstanding its importance for patient safety and ultimately for quality of care, this perspective disregards the multi-factorial nature of patient outcomes and the team nature of healthcare. By taking the perspective of the patient in this thesis, we add another dimension to the outcome discussion in medical education by bringing our attention to the healthcare context in which medical education takes place. Another strength of this thesis lies in its methodological approach. We applied advanced techniques, such as multiple imputation of missing data, multilevel analyses, and generalizability theory, to come to our conclusions. Moreover, we were also the first to apply bias analyses in medical education research. This statistical technique was developed by Arah and Thompson[69-71] but up to now has only been used in clinical epidemiology research.[72] We believe that this technique will be indispensable to research of complex learning environments in the future.

The methodological rigour in our studies was ensured by the use of high quality national databases as well as reliable instruments. The conclusions of this thesis are based on large samples including surgical and non-surgical specialties, academic and nonacademic teaching hospitals. We also used instruments that have been validated in the same context supporting their validity and reliability (D-RECT, SETQ, CQI Inpatient Hospital Care). In order to protect anonymity and encourage truthfulness of residents' responses, we used a secure and anonymous online platform to collect data about teaching quality (SETQ) and learning climate of teaching departments (D-RECT). We believe this also contributed to a sufficiently high enough response rate of above 60 percent for all of our studies. In our study, we used appropriate level of aggregation depending on the unit of analysis. For example, we aggregated D-RECT scores to the level of the department to indicate the departments' learning climate and SETQ scores to the level of the faculty to indicate teaching performance of a faculty. The use of multilevel regression analyses, on the other hand, allowed us to use disaggregated patient data, which avoids the potential for ecological fallacy when interpreting the results. We adjusted for relevant covariates, multiple associations and accounted for bias sources from selection and uncontrolled confounding. All of these contributed to the methodological rigour of the thesis.

Nonetheless, there are several limitations associated with the work presented here that relate to our choice of measures, the use of registry data, the research design, and cultural transferability of our results. Although we chose inherently patient-centered clinical measures that are amenable to healthcare delivery, the multi-factorial nature of patient outcomes presents a particular challenge in attribution of these outcomes to a physician and/or a department. This challenge is not new: In their 2013 "White paper" on "Achieving the Potential of Health Care Performance Measures" Berenson, Pronovost and Krumholz state that in order to be useful, measures need to be appropriately attributable to the right provider or group of providers, they also need to be measured reliably.[6] In our studies we used intraclass correlation coefficients to represent the proportion of variance in a particular outcome that can be explained by differences in healthcare providers (e.g. faculty, resident) or an organizational unit (e.g. hospital, department). Nonetheless, even with high intraclass correlations, these measures only reflect a very specific aspect of performance that can reliably be measured in practice, leaving out other potentially important (but unmeasurable) aspects of clinical performance.[36] A similar argument can be made for the chosen postgraduate medical education measures. By limiting our focus on residents' perspectives in our studies, we did not take into account other potentially more "objective" measures of postgraduate medical education quality, such as staff to resident ratio, which could have a more direct connection with patient outcomes.

The use of existing registries and data allowed for large-scale studies to be conducted. However, the use of existing registries for research presents several inherent limitations.[73] First, the data in existing registries were not specifically collected for the purposes of our research. As a result, we could not include potentially important variables in our analyses. Furthermore, since the data was collected in the past, it is extremely difficult to go back and collect additional data for the same time period. Another limitation of using the existing databases is that the researchers had only indirect control over the response process (for example, sending reminders) and, as a consequence, the response rate, which was left to the individual departments conducting the questionnaires. A prospective study design can ensure that all relevant data is collected and to maximize the response rate. However, this is a more intensive process requiring more resources, but has the disadvantage of fallout over time and fewer departments being involved than with large-scale cross-sectional studies.

Cross-sectional design of the studies, although easier to implement, limits statements we can make about causality, explanatory mechanism, or direction of the associations. In our studies, we always used postgraduate medical education indicators as "exposure" variables to predict care quality and patient outcomes in teaching departments. However, as discussed earlier, the opposite direction could also be true (i.e. that patient outcomes and quality of patient care may predict perceptions of the learning climate and teaching quality). The opposite hypothesis, thus, merits further empirical exploration. One potential method for improving statements about causality in cross-sectional studies is to separate the "exposure" and "outcome" variables in time as we did in chapter 6. However, in some circumstances it may not be possible, such as in chapter 5, where patients were invited to evaluate their experiences who had been hospitalized in the past year.

Finally, perception-based tools can be influenced by the cultural context in which they are developed and used.[74] Therefore, the results of our studies could not be directly transferrable to other cultural contexts. Particularly, it is possible that the relationships reported in this thesis cannot be found in settings, where there is less pressure or accountability placed on improving the learning climate in postgraduate medical education. Part of the reason for the increased focus on improving the learning climate in residency could be that teaching departments in the Netherlands receive a substantial financial compensation for teaching residents. Another reason could be that Dutch residents experience lower levels of hierarchy in the workplace and have relatively more say in their working conditions and quality of their training.

A final note

Like climate science research that aims to study such complex phenomena as climate change, researching the relationship between PGME and care quality and, especially patient outcomes, requires a sustained multidisciplinary collaboration of a broad range of disciplines.[75] Climate change and quality and safety in health care are not unlike each other. As Vincent and colleagues have put it: "Both [climate change and the perilous state of health care quality and safety] constitute a profoundly serious man-made threat to the public good which have until recently been both ignored and denied but are increasingly being recognized, taken seriously and acted on."[75] As this thesis has shown, PGME can act both as a threat and as a resource for the delivery of high quality patient care, and its contributions to patient care are only now being fully recognized. We hope that, by putting together experts in PGME and quality and safety and aiming for the common goal of high quality and safe patient care, we can minimize potential harm and maximize benefit of PGME for patients and populations now and in the future.

References:

1. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560-575.
2. Kane MT. Current concerns in validity theory. *J Educ Meas*. 2001;38(4):319-342.
3. Schuwirth LW, van derVleuten CP. Programmatic assessment and Kane's validity perspective. *Med Educ*. 2012;46(1):38-48.
4. Schneider B, Ehrhart MG, Macey WH. Organizational climate and culture. *Annu Rev Psychol*. 2013;64:361-388.
5. Delnoij DM, Rademakers JJ, Groenewegen PP. The Dutch consumer quality index: an example of stakeholder involvement in indicator development. *BMC Health Serv Res*. 2010;10:88.
6. Berenson RA, Pronovost PJ, Krumholz HM. Achieving the potential of health care performance measures. Princeton, NJ: Robert Wood Johnson Foundation; 2013.
7. Boor K. The clinical learning climate (dissertation). Amsterdam: Vrije Universiteit Amsterdam, Faculteit Geneeskunde; 2009.
8. de Feijter JM, de Grave WS, Dornan T, Koopmans RP, Scherpbier AJ. Students' perceptions of patient safety during the transition from undergraduate to postgraduate training: an activity theory analysis. *Adv Health Sci Educ Theory Pract*. 2011;16(3):347-358.
9. Beus JM, Payne SC, Bergman ME, Arthur W. Safety Climate and Injuries: An Examination of Theoretical and Empirical Relationships. *J Appl Psychol*. 2010;95(4):713-727.
10. Schneider B, White SS, Paul MC. Linking service climate and customer perceptions of service quality: test of a causal model. *J Appl Psychol*. 1998;83(2):150-163.
11. Teunissen PW. Experience, trajectories, and reifications: an emerging framework of practice-based learning in healthcare workplaces. *Adv Health Sci Educ Theory Pract*. 2014.
12. Dornan T, Boshuizen H, King N, Scherpbier A. Experience-based learning: a model linking the processes and outcomes of medical students' workplace learning. *Med Educ*. 2007;41.
13. Watling C, Driessen E, van der Vleuten CP, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgements. *Med Educ*. 2012;46(2):192-200.
14. Dalia S, Schiffman FJ. Who's my doctor? First-year residents and patient care: hospitalized patients' perception of their "main physician". *J Grad Med Educ*. 2010;2(2):201-205.
15. Schein EH. Organizational psychology. Englewood Cliffs, NJ: Prentice Hall; 1965.
16. Jippes E, Van Luijk SJ, Pols J, Achterkamp MC, Brand PL, Van Engelen JM. Facilitators and barriers to a nationwide implementation of competency-based postgraduate medical curricula: a qualitative study. *Med Teach*. 2012;34(8):e589-602.
17. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ*. 2015;7(3):349-363.
18. Philibert I, Nasca T, Brigham T, Shapiro J. Duty-hour limits and patient care and resident outcomes: can high-quality studies offer insight into complex relationships? *Annu Rev Med*. 2013;64:467-483.
19. Arrighi JA, Hebert JC. Duty hour requirements: time for a new approach? *JAMA*. 2014;312(22):2342-2344.
20. Sklar DP, Hemmer PA, Durning SJ. Medical education and health care delivery: a call to better align goals and purposes. *Acad Med*. 2017.
21. Wong BM, Holmboe ES. Transforming the academic faculty perspective in graduate medical education to better align educational and clinical outcomes. *Acad Med*. 2016;91(4):473-479.
22. Busari JO. Comparative analysis of quality assurance in health care delivery and higher medical education. *Adv Med Educ Pract*. 2012;3:121-127.
23. Norman G. Medical education: past, present and future. *Perspect Med Educ*. 2012;1(1):6-14.
24. Holmboe ES, Foster TC, Ogrinc G. Co-creating quality in health care through learning and dissemination. *J Contin Educ Health Prof*. 2016;36 Suppl 1:S16-18.
25. Sholl S, Ajjawi R, Allbutt H, et al. Balancing health care education and patient care in the UK workplace: a realist synthesis. *Med Educ*. 2017.

26. Kesselheim JC, Cassel CK. Service: an essential component of graduate medical education. *N Engl J Med*. 2013;368(6):500-501.
27. Sanfey H, Cofer J, Hiatt JR, et al. Service or education: in the eye of the beholder. *Arch Surg*. 2011;146(12):1389-1395.
28. Gupta R, Arora VM. Merging the health system and education silos to better educate future physicians. *JAMA*. 2015;314(22):2349-2350.
29. Philibert I. Resident perspectives on duty hour limits and attributes of their learning environment. *BMC Med Educ*. 2014;14 Suppl 1:S7.
30. Whitehead CR, Hodges BD, Austin Z. Captive on a carousel: discourses of 'new' in medical education 1910-2010. *Adv Health Sci Educ Theory Pract*. 2013;18(4):755-768.
31. Lindenauer PK, Shojania KG. The era of big performance measurement: here at last? *Jt Comm J Qual Patient Saf*. 2008;34(6):307-308.
32. Atkins D. The next generation of clinical performance measures. *J Gen Intern Med*. 2016;31 Suppl 1:3-5.
33. Weinstein DF. Optimizing GME by measuring its outcomes. *N Engl J Med*. 2017;0(0):null.
34. Nelson EC, Dixon-Woods M, Batalden PB, et al. Patient focused registries can improve health, care, and science. *BMJ*. 2016;354:i3319.
35. Horstman MJ, Naik AD. A 'Just Culture' for performance measures. *BMJ Qual Saf*. 2015;24(8):486-489.
36. Berenson RA. If you can't measure performance, can you improve it? *JAMA*. 2016;315(7):645-646.
37. Berenson RA. Improving performance, not just what's measured: does the inpatient prospective payment system provide useful lessons? *J Ambul Care Manage*. 2016;39(2):111-114.
38. Berenson RA, Kaye DR. Grading a physician's value--the misapplication of performance measurement. *N Engl J Med*. 2013;369(22):2079-2081.
39. Hofmann DA, Mark B. An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology*. 2006;59(4):847-869.
40. Lombarts KM, Heineman MJ, Scherpbier AJ, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One*. 2014;9(1):e86512.
41. Scheepers RA, Lombarts KM, van Aken MA, Heineman MJ, Arah OA. Personality traits affect teaching performance of attending physicians: results of a multi-center observational study. *PLoS One.* 2014;9(5):e98107.
42. Schneider B, Ehrhart MG, Macey WH. Perspectives on organizational climate and culture. APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization. Washington, DC: American Psychological Association; US; 2011:373-414.
43. Regehr G. It's NOT rocket science: rethinking our metaphors for research in health professions education. *Med Educ*. 2010;44(1):31-39.
44. Cleland J, Durning S, (eds.). Researching medical education. UK: The Association for the Study of Medical Education; 2015.
45. Lingard L, McDougall A, Levstik M, Chandok N, Spafford MM, Schryer C. Representing complexity well: a story about teamwork, with implications for how we teach collaboration. *Med Educ.* 2012;46(9):869-877.
46. Frambach JM, Driessen EW, van der Vleuten CP. Using activity theory to study cultural complexity in medical education. *Perspect Med Educ*. 2014;3(3):190-203.
47. Engestrom Y. Activity theory as a framework for analyzing and redesigning work. *Ergonomics*. 2000;43(7):960-974.
48. Engestrom Y, Sannino A. Studies of expansive learning: Foundations, findings and future challenges. *Educ Res Rev-Neth*. 2010;5(1):1-24.
49. Skipper M, Musaeus P, Nohr SB. The paediatric change laboratory: optimising postgraduate learning in the outpatient clinic. *BMC Med Educ*. 2016;16:42.
50. Cianciolo AT. Deciding 'what to teach' health professionals: a human-centred systems engineering perspective. *Med Educ*. 2014;48(12):1150-1156.
51. Cristancho S. What can we learn from a soft sister? A complementary lens to the systems engineering approach in medical education research. *Med Educ*. 2014;48(12):1139-1141.
52. Cristancho S, Lingard L, Forbes T, Ott M, Novick R. Putting the puzzle together: the role of 'problem definition' in complex clinical judgement. *Med Educ*. 2017;51(2):207-214.

53. Rojas D, Grierson L, Mylopoulos M, Trbovich P, Bagli D, Brydges R. How can systems engineering inform the methods of programme evaluation in health professions education? Med Educ. 2017.
54. Cooper H, Geyer R. Using 'complexity' for improving educational research in health care. *Soc Sci Med*. 2008;67(1):177-182.
55. Bleakley A. Blunting Occam's razor: aligning medical education with studies of complexity. *J Eval Clin Pract*. 2010;16(4):849-855.
56. Thompson DS, Fazio X, Kustra E, Patrick L, Stanley D. Scoping review of complexity theory in health services research. *BMC Health Serv Res*. 2016;16:87.
57. Tan J, Wen HJ, Awad N. Health care and services delivery systems as complex adaptive systems. *Commun Acm*. 2005;48(5):36-44.
58. Greenhalgh T, Plsek P, Wilson T, Fraser S, Holt T. Response to 'The appropriation of complexity theory in health care'. *J Health Serv Res Policy*. 2010;15(2):115-117.
59. Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ*. 2001;323(7313):625-628.
60. Holland JH. Complex adaptive systems. *Daedalus*. 1992;121(1):17-30.
61. Schwab ED, Pienta KJ. Cancer as a complex adaptive system. *Med Hypotheses*. 1996;47(3):235-241.
62. Boisot M, Child J. Organizations as adaptive systems in complex environments: The case of China. *Organ Sci*. 1999;10(3):237-252.
63. Ball P. Complex adaptive systems: An introduction to computational models of social life. *Nature*. 2007;448(7154):647-648.
64. van Rossum TR, Scheele F, Scherpbier AJ, Sluiter HE, Heyligers IC. Dealing with the complex dynamics of teaching hospitals. *BMC Med Educ*. 2016;16:104.
65. Paley J. The appropriation of complexity theory in health care. *J Health Serv Res Policy*. 2010;15(1):59-61.
66. van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med*. 2012;10:65.
67. Asch DA, Epstein A, Nicholson S. Evaluating medical training programs by the quality of care delivered by their alumni. *JAMA*. 2007;298(9):1049-1051.
68. Tamblyn R. Outcomes in medical education: what is the standard and outcome of care delivered by our graduates? *Adv Health Sci Educ Theory Pract*. 1999;4(1):9-25.
69. Arah OA. Bias analysis for uncontrolled confounding in the health sciences. *Annu Rev Public Health*. 2017;38:23-38.
70. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol*. 2008;18(8):637-646.
71. Thompson CA, Arah OA. Selection bias modeling using observed data augmented with imputed record-level probabilities. *Ann Epidemiol*. 2014;24(10):747-753.
72. Arah OA, Sudan M, Olsen J, Kheifets L. Marginal structural models, doubly robust estimation, and bias analysis in perinatal and paediatric epidemiology. *Paediatr Perinat Epidemiol*. 2013;27(3):263-265.
73. Haut ER, Pronovost PJ, Schneider EB. Limitations of administrative databases. *JAMA*. 2012;307(24):2589; author reply 2589-2590.
74. Kikukawa M, Stalmeijer RE, Emura S, Roff S, Scherpbier AJ. An instrument for evaluating clinical teaching in Japan: content validity and cultural sensitivity. *BMC Med Educ*. 2014;14:179.
75. Vincent C, Batalden P, Davidoff F. Multidisciplinary centres for safety and quality improvement: learning from climate change science. *BMJ Qual Saf*. 2011;20 Suppl 1:i73-78.

Summary

Chapter 1: Introduction

The context of this thesis is the increasing need in postgraduate medical education (PGME or residency training) to be more accountable to tax payers, patients and the society that it serves. Towards the end of the 20th century the standards and expectations of quality in patient care have undergone a transition away from being centered around physicians as solo-practicing medical experts towards patient-centered team-based care, where physicians are competent to address population's needs. Medical education has responded to these advances in two ways. The first response was the defining, refining and evaluating of educational outcomes for residents, mostly expressed as competencies, which residents must possess at the end of their training. The second response was the establishment of a robust quality improvement system in PGME aimed at continuously evaluating and improving postgraduate training for residents to optimize their educational outcomes. However, the growing concerns around patient safety in relation to quality of training bolstered by the ballooning costs of PGME have urgently called for the greater use of patient outcomes in medical education research.

Patient care quality indicators and patient outcomes have been used in medical education primarily to study the quality of care delivered by residents and related aspects of residency training programs. However, the focus on the quality of care delivered by residents only does not take into account the complexity of the clinical context in which residency training takes place. As a result, it may overlook other potential ways in which PGME can be related to patient care quality and patient outcomes. In order to better understand how the goals of PGME and high quality care can be better aligned, this thesis aims to explore the relationships between PGME and quality of care by investigating other ways in which PGME can be related to patient outcomes and quality of care, namely how residency training and care are related in clinical teaching departments. Taking a pragmatic approach, this thesis first investigates the validity and reliability of widely used instruments and indicators of quality in both PGME and clinical care to support their use in practice. Next, it aims to establish conceivable relationships of various aspects that define high quality PGME with indicators of patient care quality as well as patient outcomes by exploring the associations between these variables using validated instruments and patient data from existing quality databases, guided by the overarching research question: How are indicators of PGME quality related to indicators of patient care quality and patient outcomes in clinical teaching departments?

Chapter 2

Chapter 2 reports on the validity and reliability of the Dutch Residency Educational Climate Test (D-RECT) in measuring the learning climate in clinical teaching departments. In the Netherlands, establishing and improving the residency learning climate is an important area for continuous quality improvement in PGME. While the D-RECT has been increasingly used to evaluate the learning climate in the PGME setting, it has not yet been tested in its final form and on the actual level of use – the clinical teaching department. The aim of this study was to investigate the validity and reliability of the D-RECT on both the resident and department levels. In 2012 – 2013, Dutch postgraduate medical trainees including residents, interns and fellows filled out 2306 learning climate evaluations of 291 clinical departments in 48 teaching hospitals. The results of the exploratory factor analysis show a 9-factor structure containing 35 items: teamwork, role of specialty tutor, coaching and assessment, formal education, resident peer collaboration, work is adapted to residents' competence, patient sign-out, educational atmosphere, and accessibility of supervisors. The results of the subsequent confirmatory factor analysis indicate an acceptable to good fit of both resident and department levels. The 9-factor structure at the resident and department levels is further supported by the item-total correlations >0.30 indicating that each item contributes to the measurement of the concept learning climate, and inter-scale correlations <0.70 indicating that the D-RECT sub-constructs overlapped by <50%. The results of generalizability analyses show that a minimum of three respondents are needed to assess the overall learning climate reliably and eight to assess the all of the individual subscales, which represents an improvement from the previous version of the D-RECT. The identified subscales can be mapped to Ostroff's taxonomy of organizational climate perceptions consisting of three higher order facets: affective, cognitive, and instrumental.

Chapter 3

Chapter 3 reports on a cross-sectional observational study investigating the association between residency learning climate and adverse perinatal and maternal outcomes in non-academic obstetrics and gynecology teaching departments in the Netherlands. In 2013, a total of 103 trainees including residents, interns and fellows in 16 obstetrics and gynecology teaching departments evaluated their learning climate using the D-RECT. For the same year, anonymized data on adverse maternal and perinatal outcomes for the same departments was retrieved with permission from the Netherlands Perinatal Registry (PRN), excluding deliveries meeting the criteria for transfer to a tertiary care center as well as any births with congenital abnormalities. Adverse perinatal outcomes included fetal or early neonatal mortality, five-minute Apgar score <7, or neonatal intensive care unit admission for ≥24 hours. Adverse maternal outcomes included postpartum hemorrhage and/or blood transfusion, death, uterine rupture, or third- or fourth-degree perineal laceration. The results of the multilevel logistic regressions indicate that higher department's residency learning climate scores are associated with a significantly greater odds of an adverse perinatal outcome even after adjustment for maternal and department characteristics. When the learning climate scores are divided in tertiles, compared to those in the lowest tertile, departments in the middle tertile have a 46% greater odds of an adverse perinatal outcome, while departments in the highest tertile have 69% greater odds of an adverse perinatal outcome. The increase in odds of adverse perinatal outcomes seems to be driven mainly by low five-minute Apgar scores (<7) as opposed to other adverse perinatal outcomes, which are very rare. There was no significant association between department learning climate and adverse maternal outcomes. The protective effect of lower learning climate scores remains in sensitivity analyses using quartiles instead of tertiles as well as subgroup analyses of deliveries performed by residents only and excluding multiple births and stillbirths. Bias analyses for uncontrolled confounding, including selection bias, show that the relationship would be weakened if the unmeasured confounder set simultaneously increased (or decreased) the learning climate score (a desirable result) and the odds of an adverse outcome (an undesirable result).

Chapter 4

Chapter 4 re-focuses on a different type of outcome by investigating quality of care from the perspective of the patient. The study reported in this chapter investigates the psychometric properties of a widely used Consumer Quality Index (CQI) Inpatient Hospital Care. The different purposes for which the questionnaire has been used, including external accountability and internal quality assurance, resulted in different versions of the questionnaire to be produced. The purpose of the study was to investigate the internal validity and reliability of a shortened version of the questionnaire on patient and department levels as well as the number of patient evaluations needed per department and per hospital to generate reliable assessments. A total of 22924 CQI Inpatient Hospital Care questionnaires were completed by adult patients ≥16 years old who had been hospitalized for ≥24 hours between 1 January 2013 and 31 December 2014 in 23 Dutch academic and nonacademic hospitals. Confirmatory factor analysis demonstrated a good fit of the 35 item and 9-factor structure of the questionnaire on the individual level. Individual subscales show an acceptable reliability and low inter-scale correlations (<0.70). However, the results of the confirmatory factor analysis for scores aggregated to the level of the department demonstrate a less desirable fit. In particular, inter-scale correlations show a significant overlap between two subscales (communication of doctors and explanation of treatment). In total, 4-8 departments and 50 respondents per department are needed to reliably evaluate subscales rated on a 4-point scale (communication with nurses, communication with doctors, own contribution, explanation of treatment, pain management, communication about medication, feeling of safety), and 10 departments with 100-150 respondents per department for binary subscales (admission, information at discharge). Different numbers of respondents may, however, be required whether the questionnaire is used for summative (external reporting and high stakes decision-making) or formative (internal quality improvement) purposes. Compared to other questionnaires reported in the literature, the shortened CQI Inpatient Hospital Care questionnaire needs fewer respondents to evaluate patient experiences of a single department, while roughly similar sized samples are needed to reliably evaluate all subscales on the hospital level.

Chapter 5

Going deeper into the associations between learning climate and patient care quality, the cross-sectional observational study in chapter 5 investigates the association of residency learning climate as well as its domains with inpatient care experience in 86 teaching departments across 18 academic and nonacademic hospitals and 15 specialties in 2013 and 2014. This time the outcome of interest is the overall patient care experience as well as the individual domains of patient experience at clinical teaching departments. Patients ≥16 years old who were hospitalized for ≥24 hours between 1 January 2013 and 31 December 2014 were invited to fill out the CQI Inpatient Hospital Care questionnaire. Postgraduate medical trainees (residents, fellows and interns) currently rotating or who had recently completed their rotation at the department evaluated the department's learning climate in the same year using the D-RECT. In total, 1201 D-RECT evaluations and 6718 CQI Inpatient Hospital Care evaluations were analyzed. After controlling for respondent and department characteristics as well as multiple testing, the results of multilevel linear regressions show that a department's higher scores on residency learning climate are associated with a significant but small increase in inpatients' experiences of communication with doctors and feeling of safety. Among learning climate subscales, higher scores on coaching and assessment and peer collaboration are positively associated with higher patients' experiences of communication with doctors and explanation of treatment, respectively. On the other hand, higher scores on the learning climate subscale formal education are associated with lower pain management scores. These associations remain the same when Bartlett factor scores are used instead of simple means for patient experience subscales. The findings support the role of teamwork, supervision and communication inherent in the patient-centered subscales coaching and assessment and peer collaboration as being important for quality of care. In contrast, the subscale formal education is more resident-centered, and as a result, can strain already limited clinical resources (such as availability of faculty and residents, space and time) away from direct patient care.

Chapter 6

Chapter 6 shifts the focus from the learning climate to clinical teaching quality by investigating the hypothesis that good clinical teachers are also good clinicians in a retrospective observational study of anesthesiology faculty in an academic teaching hospital in the Netherlands. Faculty's teaching performance and role modeling were evaluated in 2010-2012 using the extensively validated System for Evaluation of Teaching Qualities (SETQ). The associations of faculty's mean SETQ, global teaching and role modeling scores with several pre-defined clinical performance measures in the period of 6 months after the faculty's clinical teaching evaluation were studied, namely: (1) intra-operative temperature monitoring and achievement of normothermia, (2) post-operative pain scores, (3) neuromuscular monitoring using Train of Four (TOF) count/value prior to patient extubation in the operating room and subsequent achievement of TOF value >70 or >90, (4) postoperative nausea and vomiting prophylaxis. In total, 54 individual faculty were evaluated by 51 residents yielding a total of 757 SETQ evaluations. Of over 15000 patient encounters analyzed, 43% are when faculty supervised a resident. After adjustment for patient, faculty and respondent characteristics on SETQ questionnaire, hierarchical panel analyses demonstrate that higher SETQ teaching and role modeling scores are associated with faculty's better performance on neuromuscular monitoring, including neuromuscular monitoring using Train of Four (TOF) count/value prior to patient extubation in the operating room and subsequent achievement of TOF value >70 or >90, but not with other outcome measures. While involvement of residents in perioperative care is associated with better performance on indicators of normothermia as well as neuromuscular monitoring, there is no association between faculty's teaching performance or role modeling scores and any of the clinical indicators of perioperative care quality in perioperative cases with resident involvement.

Chapter 7: Discussion

Chapter 7 synthesizes the research from previous chapters and weighs up the evidence in light of the existing literature, provides recommendations for practice and sets out a plan for future research. The strengths and weaknesses of the thesis are also explored. The summary of results provides the answer to the main research question "how are indicators of PGME quality related to indicators of care quality and patient outcomes in clinical teaching departments?" The suggestion of our research is that high quality training and high quality care do not always go hand in hand, as it might have initially been expected. On the one hand, the better a teaching department scores on residency learning climate, the higher its rate of complications. On the other hand, hospitalized patients in departments with higher scoring learning climates report feeling safer during hospitalization and report better communication with their doctors. Clinical supervisors who are excellent clinical teachers and role models are also better clinicians, however, that does not necessarily translate to improvements in clinical performance when supervising residents. The discussion of the results underscores the need for valid and reliable instruments, on which solid conclusions about these associations can be made. The studies reported in this thesis provide clues to the potential mechanisms behind these associations. First, if present, the associations were most evident in high stakes clinical situations as represented by Apgar scores in obstetrical care which reflect the degree of neonatal resuscitation, timeliness and degree of pain management of hospitalized patients, as well as neuromuscular function monitoring prior to extubation, all of which are highly time-sensitive and have a high teachable value. Second, the direction of the associations of learning climate subscales seemed to depend on whether the subscale was more patient-centered (coaching and assessment and peer collaboration) or resident-centered (formal education), where better performance on patient-centered subscales is associated with improved patient outcomes and resident-centered subscales were associated with worse patient-related outcomes. Here, organizational climate literature provides additional potential explanations for the negative associations. Negative associations can represent a reciprocal relationship between residents' learning climate perceptions and patient outcomes, where outcomes may drive perceptions. Additionally, negative associations can be explained by the tensions that arise when certain aspects of PGME training and demands of patient care share limited resources. In this chapter, it is argued that the ultimate interpretation of the findings rests on a holistic understanding of the complexity of the clinical learning environment in which training takes place as well as the need for better integration of PGME and healthcare delivery systems in practice. Future research directions should continue to expand the validity evidence for existing tools based on the goals and the level of their use, develop new meaningful clinical performance measures, and study the complexity of the clinical learning environments to better understand the mechanisms behind the associations by adopting a "complexity lens". To this end, several conceptual frameworks exist that are particularly suited to the medical (educational) context: activity theory, systems thinking, and complexity theory.

Samenvatting

Hoofdstuk 1: Inleiding

Dit proefschrift is geschreven in het licht van de toenemende vraag aan medisch specialistische vervolgopleidingen om meer rekenschap af te leggen over de kwaliteit van de door hen opgeleide specialisten aan de belastingbetaler, de patiënt en de maatschappij. Aan het eind van de 20^{e} eeuw heeft een kentering plaatsgevonden in de standaarden voor en verwachtingen van de kwaliteit van de patiëntenzorg: om tegemoet te komen aan veranderende zorgbehoeften is de aandacht verschoven van individuele, medische experts naar teams van artsen en andere zorgverleners die zorgdragen voor de patiënt. Het medisch onderwijs heeft op twee manieren gereageerd op deze ontwikkelingen: de eerste reactie was het definiëren, verfijnen en evalueren van onderwijseinddoelen voor artsen-in-opleiding-tot- specialist (AIOS), vaak uitgedrukt in competenties waarover jonge artsen dienen te beschikken aan het einde van hun opleiding. De tweede reactie was het definiëren van een kwaliteitsbeleid en het opzetten van een robuust kwaliteitssysteem voor medische vervolgopleidingen gericht op het continu evalueren en verbeteren van deze vervolgopleidingen. Het bewaken van de opleidingskwaliteit kan echter op gespannen voet staan met het waarborgen van de patiëntveiligheid. In combinatie met de hoge kosten van de medische vervolgopleidingen is het van groot belang ook gebruik te maken van patiënten uitkomsten in het onderzoek naar onderwijs.

Het primaire doel van het gebruik van klinische prestatie-indicatoren en patiëntenuitkomsten in medisch onderwijs was tot dusver om de kwaliteit van zorg die AIOS in de praktijk leveren te bestuderen, inclusief de factoren die daarop van invloed kunnen zijn. Binnen deze definiëring van de geleverde zorg door AIOS wordt echter geen rekening gehouden met de complexe, klinische omgeving waarin het opleiden plaatsvindt. Teneinde beter te begrijpen hoe de doelen van de medische vervolgopleiding enerzijds en de kwaliteit van geleverde zorg anderzijds beter op elkaar kunnen worden afgestemd, beoogt dit proefschrift de relatie tussen de medische vervolgopleidingen en kwaliteit van zorg te exploreren. Rekening houdend met de complexe, klinische context waarbinnen zowel opleiding als patiëntenzorg plaatsvinden, wordt de relatie tussen de medische vervolgopleiding en kwaliteit van zorg onderzocht binnen de context van klinische (opleidings)afdelingen. De overkoepelende onderzoeksvraag luidt: Hoe verhouden de kwaliteitsindicatoren voor de medische vervolgopleiding zich tot de indicatoren voor kwaliteit van de patiëntenzorg en patiëntenuitkomsten op afdelingen waar AIOS worden opgeleid? Vanuit een pragmatische benadering wordt als eerste de validiteit en betrouwbaarheid van veelgebruikte instrumenten en kwaliteitsindicatoren in zowel de vervolgopleidingen als patiëntenzorg onderzocht, om zo bij te dragen aan de optimalisatie van het gebruik van deze instrumenten in de praktijk. Ten tweede wordt nader ingezoomd op de verschillende aspecten van kwaliteit van de medische vervolgopleiding. Er wordt getracht verbanden te vinden tussen kwaliteitsaspecten van de medische vervolgopleiding enerzijds en kwaliteitsindicatoren voor patiëntenzorg en daadwerkelijke patiënten uitkomsten anderzijds. Deze verbanden worden onderzocht met behulp van gevalideerde instrumenten en patiëntgegevens uit bestaande kwaliteitsdatabases.

Hoofdstuk 2

Hoofdstuk 2 rapporteert de validiteit en betrouwbaarheid van de *Dutch Residency Educational Climate Test* (D-RECT) voor het meten van het opleidingsklimaat van klinische afdelingen. In Nederland speelt het ondersteunen en verbeteren van het opleidingsklimaat een belangrijke rol in continue kwaliteitsverbetering van de medische vervolgopleidingen. Hoewel de D-RECT steeds vaker wordt gebruikt om het opleidingsklimaat te evalueren, is de vragenlijst nog niet getest in zijn definitieve vorm noch op het feitelijke niveau van gebruik - de afdeling. Het doel van deze studie was dan ook om de validiteit en betrouwbaarheid van de D-RECT te onderzoeken op zowel het individuele als het afdelingsniveau. In 2012 en 2013 vulden Nederlandse A(N)IOS en fellows in totaal 2306 D-RECT vragenlijsten in om het opleidingsklimaat van 291 afdelingen in 48 ziekenhuizen te evalueren. Een exploratieve factoranalyse resulteerde in de clustering van 35 vragen in 9 factoren: opleidingssfeer, werken in een team, rol formele opleider, begeleiden en toetsen, gepland onderwijs, samenwerking peers, aansluiting werk bij AIOS, toegankelijkheid supervisoren, en overdracht. De resultaten van de confirmatieve factoranalyse duidden op een acceptabele tot goede fit van deze 9 factoren op zowel individueel als afdelingsniveau. De gevonden structuur werd verder ondersteund door de item-totaalcorrelaties (>0,30), die aantoonden dat elk item bijdraagt aan de meting van het fenomeen opleidingsklimaat. Ook de correlaties tussen de factoren (<0,70) bevestigden de gevonden structuur door aan te tonen dat de 9 factoren voor minder dan 50% overlapten. De resultaten van de generaliseerbaarheidsanalyses lieten zien dat minimaal drie respondenten nodig zijn om het algehele opleidingsklimaat betrouwbaar te beoordelen en acht respondenten om alle afzonderlijke subschalen betrouwbaar te beoordelen. Deze resultaten zijn een verbetering ten opzichte van de prestaties van de vorige versie van de D-RECT. De geïdentificeerde subschalen vinden een theoretische inbedding in de beschreven taxonomie van Ostroff, die schrijft dat de percepties van organisatieklimaten uit drie overkoepelende facetten bestaan: affectief, cognitief en instrumenteel.

Hoofdstuk 3

Hoofdstuk 3 rapporteert de bevindingen van een cross-sectioneel observationeel onderzoek naar de associatie tussen het opleidingsklimaat en ongunstige perinatale en maternale uitkomsten op niet-academische afdelingen verloskunde en gynaecologie in Nederland. In 2013 evalueerden 103 A(N)IOS en fellows van 16 afdelingen verloskunde en gynaecologie, hun opleidingsklimaat met behulp van de D-RECT. Met toestemming van de Perinatale Registratie Nederland (PRN) werden voor dezelfde afdelingen en hetzelfde jaar geanonimiseerde gegevens over ongunstige maternale en perinatale uitkomsten verkregen. Bevallingen die voldeden aan de criteria om plaats te mogen vinden in een derdelijns instelling en geboorten met aangeboren afwijkingen werden geëxcludeerd. Perinatale complicaties werden gedefinieerd als foetale of vroege neonatale mortaliteit, een 5-minuten Apgar-score <7, of opname op een neonatale intensive care afdeling ≥24 uur. Maternale complicaties werden gedefinieerd als fluxus postpartum en/of het ondergaan van een bloedtransfusie, maternale dood, uterusruptuur, of perineale laceratie van de derde of vierde graad. De resultaten van de multilevel logistische regressieanalyse toonden dat hogere D-RECT scores gepaard gaan met een significant grotere kans op een perinatale complicatie. Dit effect bleef zelfs zichtbaar na correctie voor maternale en afdelingskenmerken. Wanneer de scores van het opleidingsklimaat verdeeld werden in drie groepen op basis van tertielen, hadden afdelingen in de middengroep een 46% grotere kans op een perinatale complicatie vergeleken met afdelingen in het laag-scorende groep; afdelingen in het hoog-scorende groep hadden zelfs een 69% groter kans op een ongunstige perinatale uitkomst. Deze toename in kans bleek voornamelijk te worden veroorzaakt door lage 5-minuten Apgar-scores (<7) in plaats van de andere ongunstige perinatale uitkomsten die zeer zeldzaam zijn. Er werd geen significant verband gevonden tussen het opleidingsklimaat van de afdeling en adverse maternale uitkomsten. Het beschermende effect van lagere D-RECT scores bleef stabiel in sensitiviteitsanalyses, waarin de D-RECT scores in kwartielen (in plaats van tertielen) werden verdeeld. Ook in subgroepanalyses waarin (i) meerlingen en doodgeboorten werden geëxcludeerd, en (ii) alleen geboorten onder leiding van de AIOS werden geïncludeerd, bleven de resultaten stabiel. Biasanalyses waarin werd gecorrigeerd voor ongecontroleerde confounding, inclusief selectiebias, toonden aan dat de relatie zou worden verzwakt als de niet-gemeten confounder-set tegelijkertijd de D-RECT score (een gewenst resultaat) en de kans op een ongunstige uitkomst (een ongewenst resultaat) zou verhogen (of verminderen).

Hoofdstuk 4

Hoofdstuk 4 focust op de kwaliteit van de zorg vanuit het perspectief van de patiënt. De studie die in dit hoofdstuk wordt beschreven, onderzoekt de psychometrische eigenschappen van de veel gebruikte *Consumer Quality Index* (CQI) Ziekenhuisopname vragenlijst. Het gebruik van de vragenlijst voor uiteenlopende doelen, inclusief interne kwaliteitszorg en het afleggen van rekenschap aan externe partijen, heeft ertoe geleid dat verschillende versies van de vragenlijst zijn ontwikkeld. Daarom was het doel van deze studie om de interne validiteit en betrouwbaarheid van een verkorte versie van de vragenlijst op zowel het patiënt- en het afdelingsniveau te onderzoeken, en te berekenen hoeveel patiëntenevaluaties per afdeling en per ziekenhuis nodig zijn voor een betrouwbare beoordeling. In totaal werden 22924 CQI ingevuld door volwassen patiënten (16 jaar en ouder) die tussen 1 januari 2013 en 31 december 2014 in 23 Nederlandse academische en niet-academische ziekenhuizen waren opgenomen (>24 uur). Een confirmatieve factoranalyse toonde aan dat de 9 factor structuur waarin de 35 vragen clusterden een goede fit had op individueel niveau. Individuele subschalen toonden een acceptabele betrouwbaarheid en waren in geringe mate gecorreleerd (<0,70). De confirmatieve factoranalyse op afdelingsniveau toonde echter een minder goede fit. Met name de correlaties tussen de twee subschalen 'communicatie van artsen' en 'uitleg bij behandeling' toonden aanzienlijke overlap. Voor betrouwbare evaluatie bleken in totaal 4-8 afdelingen en 50 respondenten per afdeling nodig te zijn voor de subschalen met een 4 punts antwoordschaal (communicatie met verpleegkundigen, communicatie met artsen, eigen inbreng, uitleg bij behandeling, pijnbeleid, communicatie rond medicatie, gevoel van veiligheid). Voor binaire subschalen bleken 10 afdelingen met 100-150 respondenten per afdeling nodig (inhoud opnamegesprek, informatie bij ontslag). Het aantal benodigde respondenten kan echter variëren afhankelijk van het summatieve (rekenschap afleggen naar externe partijen) of formatieve (interne kwaliteitszorg) gebruik van de vragenlijst. In vergelijking met andere vragenlijsten die in de literatuur worden gerapporteerd, heeft de verkorte CQI Ziekenhuisopname minder respondenten nodig om de ervaringen van patiënten van één afdeling te evalueren. Een vergelijkbaar aantal respondenten is nodig om alle subschalen op ziekenhuisniveau betrouwbaar te kunnen evalueren.

Hoofdstuk 5

In hoofdstuk 5 wordt meer inzicht gegeven in het verband tussen het opleidingsklimaat en de kwaliteit van de patiëntenzorg. Deze relatie is middels cross-sectionele observationele data onderzocht op 86 afdelingen in 18 academische en niet-academische ziekenhuizen voor 15 specialismen. Dit keer wordt als uitkomstmaat de totale patiëntenervaring als ook de individuele domeinen van patiëntervaringen gebruikt. Patiënten ≥16 jaar die tussen 1 januari 2013 en 31 december 2014 gedurende ≥24 uur in het ziekenhuis werden opgenomen, werden uitgenodigd de vragenlijst voor de CQI Ziekenhuisopname in te vullen. AIOS, ANIOS en fellows die in deze periode op de afdeling werkzaam waren of die onlangs hun stage op de afdeling hadden afgerond, evalueerden het opleidingsklimaat van de afdeling met behulp van de D-RECT. In totaal werden 1201 D-RECT-evaluaties en 6718 CQI-evaluaties geanalyseerd. Na controle voor respondent- en afdelingskenmerken en het doen van meerdere evaluaties, tonen de resultaten van multilevel lineaire regressieanalyses aan dat hogere opleidingsklimaatscores samenhangen met een significant kleine toename in de gerapporteerde ervaringen van patiënten met communicatie met artsen en met een ervaren veiligheidsgevoel. Op de opleidingsklimaat-subschalen hangen hogere scores op begeleiden en toetsen en samenwerken met collega's positief samen met de ervaring van patiënten omtrent communicatie met artsen en uitleg van artsen over de behandeling. Aan de andere kan hangen hogere scores op de opleidingsklimaat-subschaal formeel onderwijs samen met lagere pijnbestrijdingsscores. Deze verbanden blijven ongewijzigd wanneer Bartlett-factorscores worden gebruikt in plaats van eenvoudige gemiddelden voor de CQI-subschalen. De bevindingen ondersteunen de rol van teamwork, supervisie en communicatie, inherent in de patiëntgerichte subschalen begeleiden en toetsen en samenwerking peers, als belangrijk voor de kwaliteit van patiëntenzorg. Daarentegen is de subschaal gepland onderwijs meer gericht op de arts-assistent, wat voor spanning kan zorgen tussen de al beperkte middelen (zoals beschikbaarheid van docenten en AIOS, ruimte en tijd) en daarom afleiden van directe patiëntenzorg.

Hoofdstuk 6

In hoofdstuk 6 verschuift de focus van het opleidingsklimaat op een afdeling naar de opleiderskwaliteiten van supervisoren. In een retrospectieve observationele studie op de anesthesiologie afdeling in een academisch ziekenhuis in Nederland is onderzocht of goede opleiders (alle leden van een opleidersgroep) ook goede clinici zijn. De opleiderskwaliteiten van de supervisoren en hun status als rolmodel werden in 2010-2012 geëvalueerd met behulp van de uitgebreid gevalideerde *System for Evaluation of Teaching Qualities* (SETQ). De verbanden tussen de gemiddelde SETQ-scores, de algemene opleidingskwaliteit en de rolmodel scores van supervisoren en enkele klinische performance maten van de opleiders die werden verzameld in de periode tot 6 maanden na de SETQ-evaluatie, werden bestudeerd. De volgende klinische prestatie-indicatoren werden daarbij onderzocht: (1) intra-operatieve temperatuurmonitoring en het bereiken van normothermie, (2) post-operatieve pijnscores, (3) neuromusculaire monitoring met behulp van *Train of Four* (TOF) waarde voorafgaand aan extubatie van de patiënt in de operatiekamer en de daaropvolgende TOF-waarde >70 of >90, (4) profylaxe voor postoperatieve misselijkheid en braken. In totaal werden 54 individuele supervisoren beoordeeld door 51 arts-assistenten die in totaal 757 SETQ-evaluaties invulden. In 43% van de meer dan 15000 geanalyseerde patiënten trad de anesthesioloog op als supervisor van een arts-assistent. Na correctie van de SETQ-score voor patiënten-, opleiders- en respondentenkenmerken, toonden hiërarchische panelanalyses aan dat hogere SETQ- en rolmodelscores samenhangen met betere prestaties van de opleiders op neuromusculaire monitoring, inclusief neuromusculaire monitoring met behulp van TOF waarde en TOF-waarde >70 of >90, maar niet met andere uitkomstmaten. Hoewel de betrokkenheid van arts-assistenten bij perioperatieve zorg wordt geassocieerd met betere prestaties op indicatoren van normothermie en neuromusculaire monitoring, is in deze gevallen geen verband gevonden tussen de opleiderskwaliteit- en rolmodelscores van supervisoren en de perioperatieve indicatoren.

Hoofdstuk 7: Discussie

In hoofdstuk 7 wordt het onderzoek uit de voorgaande hoofdstukken samengevat, afgezet tegen bestaande literatuur en worden er aanbevelingen voor praktijk en toekomstig onderzoek gedaan. De sterke en zwakke punten van het proefschrift worden hierbij ook besproken. Deze samenvatting biedt het antwoord op de overkoepelende onderzoeksvraag "Hoe relateren de indicatoren van de kwaliteit van medische vervolgopleidingen aan indicatoren van kwaliteit van zorg en patiëntenuitkomsten op klinische afdelingen waar arts-assistenten worden opgeleid?". Ons onderzoek suggereert dat een kwalitatief hoogwaardige opleiding en hoogwaardige zorgverlening niet altijd hand in hand gaan, zoals in eerste instantie verwacht was. Enerzijds blijkt dat afdelingen die beter scoren op (evaluaties van) het opleidingsklimaat, slechtere patiëntenuitkomsten realiseren, uitgedrukt in meer perinatale complicaties. Anderzijds toonden we aan dat klinische patiënten op afdelingen met hoger opleidingsklimaatscores zich veiliger voelen en een betere communicatie met hun artsen rapporteren. Opleiders die uitstekende opleiderskwaliteiten hebben en een rolmodel zijn voor arts-assistenten, zijn ook betere clinici; dat vertaalt zich echter niet noodzakelijkerwijs in betere klinische prestaties tijdens supervisie van arts-assistenten. De discussie van de bevindingen van dit proefschrift benadrukt wederom de noodzaak van valide en betrouwbare instrumenten, met behulp waarvan solide conclusies over de gevonden associaties kunnen worden getrokken. De gerapporteerde studies in dit proefschrift bieden aanwijzingen voor de mogelijke mechanismen achter deze associaties. Ten eerste waren de associaties, indien aanwezig, het meest evident in die klinische situaties waarin er veel voor de patiënten op het spel stond (*high stakes*), zoals de Apgar-score - die de mate van neonatale reanimatie weerspiegelt -, de tijdigheid en mate van pijnmanagement van klinische patiënten en de neuromusculaire functiemonitoring voorafgaand aan extubatie. Al deze maten zijn zeer tijdsgevoelig en vormen tegelijkertijd belangrijke opleidingsmomenten. Ten tweede leek de richting van de associaties van de verschillende subschalen van het opleidingsklimaat af te hangen van de gerichtheid van de schaal op de patiënt (begeleiden en toetsen en samenwerking peers) of de AIOS (gepland onderwijs). Zo bleek dat betere prestaties op patiëntgerichte subschalen geassocieerd zijn met verbeterde patiëntenervaringen en AIOS-gerichte subschalen juist met slechtere patiëntenervaringen. Een mogelijke verklaring voor de negatieve associatie kan gevonden worden in de literatuur over organisatieklimaat. Negatieve associaties kunnen een aanwijzing zijn voor een wederkerige relatie, waarbij negatieve patiëntenuitkomsten de percepties van arts-assistenten over hun opleidingsklimaat positief kunnen beïnvloeden. Bovendien kunnen negatieve associaties worden verklaard door de spanningen die (kunnen) ontstaan wanneer de beperkt beschikbare middelen, zoals tijd en capaciteit, moeten verdeeld tussen de medische vervolgopleiding en de patiëntenzorg. In dit hoofdstuk wordt beargumenteerd dat de uiteindelijke interpretatie van de bevindingen berust op een holistisch perspectief op de complexiteit van de klinische leeromgeving waarin opleiding (leren) plaatsvindt, als ook op de behoefte aan betere integratie van medische vervolgopleidingen en zorgverlening in de praktijk. Teneinde de mechanismen achter de in dit proefschrift gerapporteerde associaties beter te kunnen begrijpen, moet toekomstig onderzoek doorgaan met het verzamelen van het validiteitsbewijs voor bestaande tools, het ontwikkelen van nieuwe, betekenisvolle klinische prestatie-indicatoren en het bestuderen van klinische leeromgevingen vanuit een complexiteitsperspectief Voor deze laatste opgave bestaan verschillende conceptuele kaders die bijzonder geschikt zijn voor de medische (onderwijskundige) context: *activity theory*, systeemdenken en complexiteitstheorie.

Valorization

Social relevance

Social accountability of post-graduate medical education (PGME) to the public already underpins its efforts to educate future physicians who can meet the needs of the public. Part of this accountability is to ensure that not only the outcomes but also the processes of educating physicians are socially responsible. Thus the fundamental principle of healthcare "first, do no harm" applies not only to individual practitioners but also to their education. This thesis is, thus, socially relevant because it centers the dialogue of PGME on the patient. By doing so, this thesis brings two defining aspects of teaching hospitals together: quality of training and quality of patient care. Taking a national sample of both academic and non-academic teaching hospitals in a variety of specialties across the country, the thesis demonstrates that while in some cases what's good for the resident is also good for patients (chapters 5 and 6), this is not always the case, as with adverse perinatal outcomes in chapter 3 and patients' experiences of pain management in chapter 5. The findings are directly relevant to the public because they confirm that there is a connection between residency training and patient outcomes. Understanding how these connections are formed will aid our understanding of the potential pitfalls of PGME in workplace learning environments, and ways to remedy them.

Economic relevance

As mentioned at the beginning of this thesis, training residents costs the Dutch government over a billion euros a year and the United States government over 15 billion dollars, which provides an additional reason for focusing on clinical (outcome) measures in PGME. The studies reported in the thesis demonstrate how existing electronic health records (chapter 6) and existing clinical quality data (chapter 3 and 5) can be adapted for use in PGME research. In particular we employed the intra-class correlation coefficient as a way to determine the variance attributable to department or individual supervisor differences. These techniques can be used to identify PGME-sensitive clinical performance measures in the future. PGME-sensitive outcome measures have the potential to be used by policy makers as an additional marker for residency training program evaluation, and as a way to quantify the impact of PGME on patient care quality in order to support evidence-based innovations in PGME. The additional insight into clinical performance in teaching departments can stimulate residency programs to better align their quality improvement activities with clinical outcomes.

Target groups

Ultimately the purpose of this research is to benefit both residents in their learning to become competent physicians as well as patients who are being cared for in teaching hospitals. The results of this thesis are particularly relevant to clinical supervisors, who oversee residents' clinical work and are responsible for ensuring patient safety, providing feedback to residents about their performance, evaluating their competencies and creating a supportive learning climate. The study results underscore that maintaining this balance in practice is an important part of ensuring both high quality residency training and patient care. Based on the research results in chapter 5, clinical supervisors are advised to increase their presence and involvement in observation and feedback to residents as it is associated with better patients' experiences of hospital care. This is in line with previous studies' results that more direct clinical supervision has been associated with improved patient safety and outcomes.[1,2] Thus, efforts to support and enhance supervisors' involvement in residents' work not only promotes learning for residents, it also benefits patient care. Study results reported in chapter 6 additionally provide support for the continuation of efforts to improve clinical teaching qualities in all faculty insofar as better clinical teaching qualities are associated with improvements in faculty's own clinical performance.

In terms of organization of residency training, the study results in chapter 5 support teamwork between residents as being important for patient care, as better teamwork among residents was associated with better patients' perceptions of pain management. As front line healthcare providers, residents should be supported in the practical aspects of delivering care. Teamwork has been already recognized as an important ingredient in patient care outcomes and teamwork training as an important part of medical education.[3,4] On the other hand, since those activities that take residents away from their regular clinical duties, for example, for formal educational sessions, it is recommended that extra supports are in place to ensure continuity of care is maintained.

In terms of program evaluation and auditing, we recommend to include clinical performance measures as part of the evaluation process of training quality in clinical workplaces. For the healthcare inspectorate, it is recommended to be aware of the potential impact that education can have on healthcare processes and to consider residency training as a part of the healthcare delivery system. Greater collaboration between leadership in both quality of healthcare and residency training is required especially when large innovations are being proposed. In the Netherlands, PGME has recently undergone a complete overhaul towards competency-based training, which has affected all training programs and teaching hospitals. Since PGME is so interwoven with patient care delivery, the integration of similar innovations in PGME should be carefully facilitated as to prevent adversely affecting patient care initially.

Activities and products

Aside from the resulting academic publications and presentations at various international and national conferences, this thesis has produced two updated instruments for practical use: the Dutch Residency Educational Climate Test (D-RECT) for the evaluation of the residency learning climate in clinical teaching departments (chapter 2) as well as the Consumer Quality Index (CQI) Inpatient Hospital Care for the evaluation of patient care experiences during hospitalization (chapter 4). The updated version of the D-RECT questionnaire has been made available through www.professionalperformanceonline.nl, an online platform widely used by residency programs in the Netherlands to evaluate the residency learning climate. Using this online platform, programs can invite residents to fill out the D-RECT questionnaire. Anonymous responses are then collected in a feedback report at the end of the evaluation period provided the minimum number of residents have filled out the questionnaire. This feedback report summarizes the results in such a way that it can be used by residency programs to identify areas for improvement. Prior to the release of the new version of the D-RECT, all programs using the old questionnaire on the online platform were notified. Additionally, workshops have been organized to introduce and discuss the use of D-RECT in practice.

The updated CQI Inpatient Hospital Care questionnaire (chapter 4) has been made widely available for use through open-access publication of results. The results of the validation study were presented to interested parties in the Netherlands. As this questionnaire is now only used for hospitals' own internal quality improvement activities, rather than national benchmarking or for purchasing care by insurance companies, there is insufficient interest in the uptake of the new results. This is understandable since coordinated effort and considerable financial investment is required to achieve sufficient response rates for this questionnaire. Given the necessary infrastructure and funding to facilitate this process, the use of CQI Inpatient Hospital Care has the potential to be fully implemented in the Netherlands.

The results of the association studies reported in chapters 3, 5, and 6 have been presented at workshops to residents, clinical supervisors, and policy advisors on quality improvement in both healthcare and PGME. In these workshops the participants were asked to think about their own clinical practice in relation to study results and come up with examples of where tension between quality of residency training and quality of care might arise and how it can be resolved. In addition to this, on September 6, 2015, the author presented a pre-conference workshop together with prof. Cees van der Vleuten on researching patient care outcomes in residency training to the alumni of Maastricht University's School for Health Professions Education in Glasgow, United Kingdom. The author has prepared a number of workshops on evaluating and improving the learning climate in clinical practice for regional medical educational symposia and conferences (in Utrecht and Egmond aan Zee) as well as internationally (for the Swedish anesthesiology delegation as well as post-graduate medical training in Thailand).

Innovation

The validation study of the D-RECT, reported in chapter 2, resulted in a shorter version of the questionnaire (35 instead of 50 items and 9 instead of 11 domains), which can be reliably evaluated using smaller number of respondents (8 instead of 11 respondents). This is an improvement because it takes residents a shorter period of time to complete the questionnaire while showing improvements in its validity and reliability. The validation of the CQI Inpatient Hospital Care, reported in chapter 4, has resulted in a valid and reliable shorter questionnaire, which can be used for multiple purposes such as internal quality assurance as well as external reporting and benchmarking, provided that the minimum number of respondents has been met. Because of this improvement, a single questionnaire can be used.

This thesis also utilizes innovative approaches, such as bias analyses techniques, to studying the complex relationship between quality of residency training and quality of care. These methods can be applied to studying the effects of unknown confounding in other relationships when confounders have not or could not be measured. This has the potential to improve the quality of similar studies in the future.

In addition, this thesis demonstrated the application of existing registries and the electronic health record data to study clinical performance. Tying PGME to clinical performance also directly addresses the demand of public accountability and enables PGME to optimize its contribution to the improvement of quality of care and health of populations. The study results from chapter 6 have identified several clinical performance measures that could be used to create a real-time feedback on clinical performance for anesthesiology faculty and/or residents. Using real-time feedback for faculty and residents can help facilitate individual performance monitoring and improvement among all stages of medical education continuum. This is particularly relevant for faculty after 10 years of practice as improvement in patient outcomes tends to taper off in the second and third decades after graduation.[5]

Schedule and implementation

Results of the conducted studies were presented at international and national conferences as well as through workshops at medical education symposia in the Netherlands. The research findings will be incorporated into future workshops on quality improvement in PGME given by the Professional Performance research group. To help disseminate the research findings, printed copies of the PhD thesis will be sent to members of key stakeholders in Dutch post-graduate medical specialty training and care quality, including the Dutch Ministry of Health, quality improvement departments, hospital-wide education committees, department heads, program directors and clinical supervisors, as well as international researchers in the same field.

The author is currently working together with several researchers from The Netherlands, Canada and the United States on a collaborative paper and a grant proposal. The goal of the collaborative is: to better understand the relationship between quality of residency training and quality of care, to define evidence-based PGME-sensitive clinical performance measures using electronic health record data with the aim of facilitating real-time feedback to residents and physicians. Both the paper and the grant proposal build directly on this work and combine researchers' experience, knowledge and resources from these countries to create a greater impact and streamline research efforts. An abstract of the paper has been submitted for an oral presentation at the Second World Summit on Competency-Based Medical Education in Basel, Switzerland, on 24-25 August, 2018. The group also plans to apply for international grants to support further work and research in this area, where creating usable and meaningful products is part of the group's ultimate aim.

References:

1. Farnan JM, Burger A, Boonyasai RT, et al. Survey of overnight academic hospitalist supervision of trainees. *J Hosp Med.* 2012;7(7):521-523.
2. Van Der Leeuw RM, Boerebach BC, Lombarts KM, Heineman MJ, Arah OA. Clinical teaching performance improvement of faculty in residency training: A prospective cohort study. *Med Teach.* 2015:1-7.
3. Davenport DL, Henderson WG, Mosca CL, Khuri SF, Mentzer RM, Jr. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg.* 2007;205(6):778-784.
4. Kerry MJ, Schmutz JB, Eppich WJ. Speaking up in defence of teamwork training towards patient safety: a response. *Med Educ.* 2017;51(5):561-562.
5. Epstein AJ, Srinivas SK, Nicholson S, Herrin J, Asch DA. Association between physicians' experience after training and maternal obstetrical outcomes: cohort study. *BMJ.* 2013;346:f1596.

Acknowledgements

Dear Kiki, Cees and Renée, over the past four years you have become my pillars, akin to family, guiding me through the labyrinth of the PhD and some big life decisions on the way. :) Kiki, from the moment of our first meeting at the interview in the summer of 2013 you and Maas Jan made me feel welcomed in the AMC as the only non-Dutch PhD student in the Professional Performance research group. You really know how to strike a balance between having a full understanding of what's going on in the personal lives of your *promovendi* and giving the "nudge" when it's needed – in my opinion an optimal learning climate! So many memories, big and small, to recount here, but I will always remember us dancing, shoeless, on the rooftop of the W hotel as the sunrise was breaking at 3am at my wedding in St. Petersburg, long after most of the wedding guests had gone home. Thank you!

Cees, thank you for placing your trust in me and my abilities even when you initially hadn't interviewed me for this position. I always looked forward to coming down to the *zoning zuiden* for our *promo-overleggen*. You always knew the right thing to say, and your critical and practical input came at just the right moments, which often came very shortly after I had sent the original email. Thank you for being so accessible and supportive from the very beginning. And thank you for giving me the opportunity to stay half a year longer as a PhD student at SHE to continue my research with outstanding researchers from three different countries. I cannot imagine a more stimulating and interesting environment to work in!

Renée, without you the whole PhD experience would have been so different! You've taken me under your wing, taught me the ins and outs of managing the process behind the scenes, while keeping my head sane! And all of this was frequently done with a wine at hand sitting comfortably on your couch, or with a cocktail somewhere in a café! We celebrated the successes and learned from the rejections together, so I didn't feel like I was alone at any point during this journey. I *proost* to our journey continuing in the future as we both travel all corners of the world!

Dear Onyi, starting as a co-author, you came on to the team later on in the PhD. You were always an indispensable part of planning, carrying out and writing up of every study. Our meetings during *Heusdenweek* were very productive regardless of how limited the time was. I remember always leaving with more ideas and learning something new after our every conversation, no matter if it was about causality, bias analyses, investing & cryptocurrency or being a parent. Thank you for your insights and enthusiasm making even the most complicated concepts seem so easy to understand.

Dear Maas Jan, a special thanks to you too. Although you were not part of my supervising team, it surely did not feel like it during our meetings. Perhaps, since you and Kiki interviewed me, the first impression your warmth and friendly attitude has stayed with me throughout the years. During our *promo-overleggen* with Kiki, the discussion went often beyond the paper as you were always interested in what was happening with my PhD as a whole. Even outside of the AMC, you were just as friendly and warm, such as that time we saw each other during the *pauze* in the Concertgebouw.

Dear Professional Performance research group, you are an unmissable part of what makes it worthwhile getting up on those rainy and cold Dutch mornings! Having great people to share the highs and the lows with made all the difference. Ben, we started as roommates and ended up spending last weeks of my PhD again together in the same room – the circle is complete. Thank you for your readiness to help explain important concepts and to step in whenever needed, I really appreciate that. Renée, I admire your organizational and personal skills! You always know how to ask the right questions and are always interested in what's going on below the surface in any conversation. I enjoyed being Milou's paranimf together with you, and I hope our friendship will continue growing! Irene, how I miss your laugh and sunny energy around the AMC!! We had amazing times together bubbling about relationships and I could always share the deepest thoughts with you. Thank you for always listening and putting everything into the larger perspective of life! We need to *proost* to that with a *wijntje!* Guusje, Lenny, Joost, Joris, Renée vd L, Nelleke and Myra, you make the PPO-AMC family complete. I have had a blast during our dinners, *uitjes*, celebrations of *promoties* and weddings, cooking sessions in Heusden (sometimes complete with swimming) and especially for making it *gezellig* in the AMC. Maarten, we found a common passion in cryptocurrencies together and I hope Ripple goes up soon! Iris, just a few weeks into your PhD as our newest colleague, you already fit in so well in the team. Good luck on your journey and I am excited to see you grow!

Milou, you have a special place in my heart and my *dankwoord*. Starting together as PhD'ers on the same day, our paths would be forever intertwined, not just in publishing studies together, but also as good girlfriends. Through life's turns, we shared everything together from the frustrations of *inlezen* and rejections from journals to the highs of weddings, vacations and our PhD defenses and the lows of hospital stays. I love how we keep each other going forward and motivating each other to achieve more. I think another trip to Greece is in order!

Mirja and Elisa, *mijn paranimfen!* Thank you for taking on this task with such enthusiasm and positive energy. I am in awe of your abilities as researchers and as organizers, and I am happy both of you will stand next to me as I defend my thesis.

Dear *Maastrichtsche collega's*, Lorette, Carolin, Andrea, Koos, Ellen, Sanne & Sanne, Tiuri, Emmaline, Danielle, Anique, Janneke, Jimmy, Lilian, Nicky and others I might have forgotten. Your welcoming spirit in Maastricht made the long train trips something I looked forward to! Part of me wishes I lived in Maastricht just to be part of the buzz of the *promovendikamer*. I enjoyed our dinners at conferences, the HAM-week, and numerous celebrations. Merci!

Dear JC'ers, Anouk, Nesibe, Marianne, Tiuri, Emma, Chantal, Jessica, Karsten, thank you for the we had fun together during the conferences, the *proefpromoties*, countless journal clubs, and, of course, the infamous JC dinners. I've learned a lot from all of you!

Dear friends, knowingly or unknowingly, you are also all a part of this PhD journey. Nora, moving to the houseboat on Da Costakade was probably the best decision in my life. Not only because I met the love of my life there but also because I really enjoyed going to concerts at the Concertgebouw together, celebrating Thanksgiving and Sinterklaas, playing piano, and sharing most wonderful moments soaking up the sun on the deck watching boats and geese pass by. And who can forget St. Petersburg too?! Zina, from our first Dutch class at Babel, learning a new language and getting used to a new culture, we navigated every curve-ball this immigrant's life hurled at us, and we did it while enjoying all the culture and beauty of this country! Nienke, I'm so happy I met you before even moving to the Netherlands during our *semi-arts stage* at *spoedeisendehulp*. Since then we became good friends sharing in each other's career successes, relationship stories, weddings and expanding families. I'm lucky to have a friend like you! Sarah, Egon, Sander, Laura, Heather, Whit and Vic, thank you for all the *gezelligheid* and keeping the good wine and good food flowing!

Мама и папа, it's true that distance makes the heart grow fonder. The past few years have only brought us closer together and I am happy with how we have become as a family, which has grown from just 3 of us to almost 5! I love you and am grateful for the foundation you have given me to be successful in everything I do. Hats off to both of you!

My dear Matt, little did I know that we would meet on a houseboat in Amsterdam – me from Canada and you from Australia. I guess in Dutch this would be called *een goede bijvangst*. Others would say it was like *de kat op het spek binden!* When you know, you just know. Amsterdam will always have a special place in my heart! Thank you for being so supportive of all my endeavours, you are the best partner one can ask for. What a pleasure it is to be loved by you and to love you back knowing we have our whole lives to do that. Now with our little one on the way, I can't wait to step into the next chapter in our lives вместе.

Curriculum Vitae

Alina was born on October 9, 1985 in a small town Chelyabinsk-65 (later renamed Ozersk), Chelyabinskaya Oblast, Russia to two nuclear engineers. In 2000, when she was 14, they moved to Richmond Hill, Canada where Alina attended Langstaff Secondary School between 2000 and 2004 graduating cum laude. In 2004 she entered the Health Studies and Physiology undergraduate programs at the University of Toronto in Canada, for which she received Honours Bachelor of Science with High Distinction graduating in 2008. Alina stayed at the University of Toronto to study medicine from 2008 to 2012. After receiving her Doctor of Medicine diploma in June 2012, Alina moved to the Netherlands, where she entered a full-time PhD program in October 2013 at the School for Health Professions Education in Maastricht University in collaboration with Professional Performance Research Group at the Academic Medical Centre in Amsterdam. During her PhD she mentored University of Amsterdam medical students in their pre-clerkship and clerkship years and participated as an organizing committee member for the SHE Academy conference in 2015. In 2017, Alina was a Visiting Scholar at the Wilson Center in Toronto, Canada. Her research received the Best Research Paper Award at the AMEE conference in Helsinki, Finland.

Currently, she continues to work as a researcher at School for Health Professions Education and the Professional Performance Research Group to set up an international collaboration in research into quality of medical education and quality of care. After finishing her PhD studies, Alina hopes to continue her medical career as a resident in family medicine in Canada. Her ambition is to combine clinical practice and research in the future. She currently lives in Amsterdam together with her husband Matthew Kirkwood.

SHE dissertations series

The SHE Dissertation Series publishes dissertations of PhD candidates from the School of Health Professions Education (SHE) who defended their PhD theses at Maastricht University. The most recent ones are listed below. For more information go to: https://she.mumc.maastrichtuniversity.nl

Hikspoor, J. (5-12-2017) Development of the heart and vessels in the caudal part of the human body

Boymans, T. (06-10-2017) Hip arthroplasty in the very elderly: anatomical and clinical considerations

Zaidi, Z. (04-10-2017) Cultural hegemony in medical education: exploring the visibility of culture in health professions

Harrison, C. (20-09-2017) Feedback in the context of high-stakes assessment: can summative be formative?

Mekonen, H. (30-06-2017) Development of the axial musculo-skeletal system in humans

Taylor, T. (29-03-2017) Exploring Fatigue as a Social Construct: Implications for Work Hour Reform in Postgraduate Medical Education

McLellan, L. (29-03-2017) Prescribing the right medicine: Perspectives on education and practice

Ignacio, J. (09-02-2017) Stress Management in Crisis Event Simulations for Enhancing Performance

Bolink, S. (19-01-2017) Functional outcome assessment following total hip and knee arthroplasty; Implementing wearable motion sensors

Beckers, J. (09-12-2016) With a little help from my e-portfolio. Supporting students' self-directed learning in senior vocational education

Giroldi, E. (07-12-2016) Towards skilled doctor-patient communication. Putting goal-directed and context-specific communication into (educational) practice

Huwendiek, S. (25-11-2016) Virtual patients for learning of clinical reasoning

Bohle-Carbonell, K. (28-09-2016) May I ask you…? The influence of individual, dyadic & network factors on the emergence of information exchange in teams

Ginsburg, S. (01-09-2016) Hidden in plain sight, the untapped potential of written assessment comments

Koops, W. (08-06-2016) Computer-supported collaborative learning in clinical clerkships

Schlegel, C. (08-06-2016) Simulated and standardized patients in health profession education: the impact of quality improvement

Sorensen, J. 01-06-2016 Obstetric simulation: designing simulation-based medical education and the role of physical fidelity

Kok, E. (01-04-2016) Developing visual expertise: from shades of grey to diagnostic reasoning in radiology

Van den Eertwegh, V. (11-11-2015) Unravelling postgraduate communication learning; from transfer to transformative learning

Gingerich, A. (03-09-2015) Questioning the rater idiosyncrasy explanation for error variance, by searching for multiple signals within the noise

Goldszmidt, M. (02-09-2015) Communication and reasoning on clinical teaching teams, the genres that shape care and education

Slootweg, I. (19-06-2015) Teamwork of Clinical Teachers in Postgraduate Medical Training

Al-Eraky, M. (21.05.15) Faculty development for medical professionalism in an Arabian context

Wearne, S. (08-04-2015) Is it remotely possible? Remote supervision of general practice registrars

Embo, M. (13-03-2015) Integrating workplace learning, assessment and supervision in health care education

Zwanikken, P. (23-01-2015) Public health and international health educational programmes for low- and middle-income countries: questioning their outcomes and impact

Kazimir Malevich
Black Square, 1913
© State Tretyakov Gallery, Moscow

The cover of the thesis is inspired by the famous painting 'The Black Square' by Kazimir Malevich. When the Black Square was first shown at The Last Futurist Exhibition in 1915, it was exhibited in the top corner beneath the ceiling, the part of the room usually reserved for Russian Orthodox icons. This fundamentally questioned long held assumptions and relationships in prerevolutionary Russia. In this thesis the author presents new findings inviting the reader to create a new understanding of the relationship between quality of residency training and quality of health care delivery.