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 taxpayers, 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 peri-
natal 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.