

Resident-sensitive quality measures

Citation for published version (APA):

Schumacher, D. J. (2020). *Resident-sensitive quality measures: defining the future of patient-focused assessment*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. <https://doi.org/10.26481/dis.20200319ds>

Document status and date:

Published: 01/01/2020

DOI:

[10.26481/dis.20200319ds](https://doi.org/10.26481/dis.20200319ds)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Resident-Sensitive Quality Measures:
Defining the Future of
Patient-Focused Assessment

Daniel John Schumacher

The research reported here was carried out at



in the School of Health Professions Education



and at Cincinnati Children's Hospital Medical Center



This work was supported by funding from the Division of Emergency Medicine and a Place Outcomes Research Award at Cincinnati Children's Hospital Medical Center.

ISBN

978-94-6380-727-2

Design/lay-out

Bregje Jaspers, ProefschriftOntwerp.nl, Nijmegen

Print

ProefschriftMaken | www.proefschriftmaken.nl

© Daniel John Schumacher, 2019

All rights are reserved. No part of this book may be reproduced, distributed, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission of the author.

Resident-Sensitive Quality Measures: Defining the Future of Patient-Focused Assessment

DISSERTATION

to obtain the degree of Doctor at 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
Thursday, 19th of March 2020,
at 16:00 hours.

by

Daniel J. Schumacher

Supervisors:

Prof. dr. C.P.M. van der Vleuten

Prof. dr. C. Carraccio (American Board of Pediatrics)

Co-supervisors:

Dr. J.O. Busari (Zuyderland Medisch Centrum Heerlen)

Dr. E. Holmboe (Accreditation Council for Graduate Medical Education)

Assessment Committee:

Prof. dr. A.B.H. de Bruin (Chair)

Prof. dr. P.L.P. Brand, Professor of Clinical Education, Universitair Medisch Centrum Groningen

Prof. dr. T.J. ten Cate, Professor of Medical Education, Universitair Medisch Centrum Utrecht

Prof. dr. I.C. Heyligers

Prof. dr. P.W. Teunissen

Table of Contents

Chapter 1	Introduction	9
Chapter 2	The Future of High-Quality Care Depends on Better Assessment of Physician Performance	25
Chapter 3	The Power of Contribution and Attribution in Assessing Educational Outcomes for Individuals, Teams, and Programs	33
Chapter 4	Developing Resident-Sensitive Quality Measures: A Model From Pediatric Emergency Medicine	47
Chapter 5	Developing Resident-Sensitive Quality Measures: Engaging Stakeholders to Inform Next Steps	67
Chapter 6	Initial Implementation of Resident-Sensitive Quality Measures in the Pediatric Emergency Department: A Wide Range of Performance	91
Chapter 7	Resident-Sensitive Quality Measures in the Pediatric Emergency Department: Exploring Relationships With Supervisor Entrustment and Patient Complexity and Acuity	107
Chapter 8	Contribution of Resident-Sensitive Quality Measure Data to Making an Entrustment Decision in a Pediatric Residency	129
Chapter 9	A Feasibility Study to Attribute Patients to Primary Interns on Inpatient Ward Teams Using Electronic Health Record Data	153
Chapter 10	Discussion	185
	Summary	207
	Samenvatting	215
	Valorization	223
	Acknowledgements	229
	Curriculum Vitae	233
	SHE Dissertation Series	235



Chapter 1

Introduction



The foundation of competency-based medical education (CBME) is ensuring that the outcomes of training prepare graduates to meet the needs of populations of patients.¹ Hall and Lloyd have taken this objective one step further to propose that the purpose of CBME is to provide *better* care.² These goals make aligning the outcomes of learners with the outcomes of patients a fundamental interest in CBME. However, understanding of the relationships among medical education, quality of care, and patient outcomes has advanced minimally over the past 40 years. A 1978 article from Magraw and colleagues about the Work Group on the Education of the Health Professions and the Nation's Health called for studies focused on the relationship between training and patient outcomes.³ Nearly 40 years later, Weinstein's editorial in the *New England Journal of Medicine* bemoaned the exact same gap in the medical education research agenda.⁴

While evidence is limited, early work in this area has emerged in the past decade. In their 2012 systematic review of the effects of residency training on patient outcomes, van der Leeuw and colleagues found limited evidence of the effect of residency training on later practice.⁵ Since that time, several large studies have emerged that demonstrate that the residency institution has an impact on future practice, with lower performing sites producing lower performing physicians after training.⁶⁻¹² Not surprisingly, then, other studies have described a lack of preparedness of residents to care for patients with common conditions or to enter fellowship training.^{13,14} Furthermore, during training, increased resident autonomy has been shown to lead to increased errors, raising concerns as to whether appropriate supervision is ensured.¹⁵ These findings have led the Institute of Medicine (now National Academy of Medicine) in the United States to call for innovation funding to address these problems and have led the Medicare Payment Advisory Commission to call into question care reimbursement if these challenges cannot be addressed.¹⁶⁻¹⁸

While these are concerning trends in the literature, other studies have demonstrated the role that residents play in ensuring patient care quality during training. For example, resident transitions have been shown to be associated with higher mortality for hospitalized patients.^{19,20} Furthermore, departments' patient safety climates have been shown to be associated with residents' self-reported patient safety behavior.²¹

Calls to Link Educational and Patient Care Outcomes

With conflicting, concerning, and limited evidence of the link between educational outcomes and patient care outcomes to date, further work in this area is important. This need is underscored by recent increases in calls to improve graduate medical education (GME) by measuring its outcomes,^{4,22} including looking to "big data" to elucidate the relationship between education

and health outcomes.^{23,24} However, we still need reliable metrics for GME outcomes to harness as big data.²⁵ The work of Gillespie and colleagues to develop an educational data registry to determine educational outcomes longitudinally is an example of an effort that could help move the medical education community toward tracking meaningful GME outcome metrics,²⁶ allowing graduates' practice data to be used to drive medical education reform.²⁷ Caverzagie and colleagues have also detailed potential performance-based metrics for GME that could be linked to GME funding; however, they also note the inherent challenges of a performance-based GME funding model.²⁸ Most of this discussion has focused on evaluating the outcomes of medical education at the program and institutional level. However, Smirnova and colleagues advocate four levels of clinical performance indicators useful to consider in GME: individual, program, institutional, and national.²⁹ *This thesis focuses on individual residents, whose data can be aggregated to inform program, institutional, and national outcomes.*

Calls for taming big data and developing outcome metrics for GME have a practical focus. However, a more fundamental problem may be inhibiting a focus on determining the relationship between education outcomes and health outcomes. Namely, GME is not focused on the patient, and it should be in order to align educational and clinical outcomes.³⁰ Rather, medical education and health care systems function as separate silos. This reality is likely due, at least in part, to how these entities have grown in ways that overlap more by chance than by intentional design in response to societal needs over the past century.³¹ Moving forward, health care delivery systems must own their role in preparing GME trainees for future care that meets societal needs,³² and the purpose and goals of medical education must align with those of the health care delivery system.³³ Furthermore, medical education and quality improvement (QI) function as silos, and we must also bridge these silos to link educational outcomes in GME to relevant clinical outcomes.^{34,35} Underscoring these silos, there has been a vociferous movement in the past decade in the United States of board-certified physicians demanding that continuous QI efforts not be part of their ongoing maintenance of certification.^{36,37}

Gaps in Resident Quality Improvement Experiences

In the United States, the Accreditation Council for Graduate Medical Education (ACGME) has instituted a Clinical Learning Environment Review (CLER) program that includes a focus on ensuring that clinical training institutions both embrace and execute patient safety and QI at their institutions as well as meaningfully involve residents in these efforts. The architects of CLER posit that "graduate medical education must include training and active participation in quality and safety initiatives by every resident physician."³⁸ In the lessons learned from CLER visits at institutions so far, it is evident that substantial progress in these areas is sorely needed.³⁹⁻⁴¹ Institutions often do not engage residents in patient safety and QI even at the level of competence

for the institution. Not surprisingly, then, Butler and colleagues recently found that residents may not fully understand the vision of quality improvement (QI), are confused about QI basics, and feel that they do not play a valuable role in QI efforts.⁴² Cooke and colleagues summarize the current state of QI training well:

The current methods of teaching quality and safety to health professionals are inadequate for the task. To the extent that quality and safety are addressed at all, traditional pedagogies are used, involving a narrow focus on content transmission from experts to neophytes, didactic sessions spatially and temporally distant from clinical work, and a linear approach to applying content in the form of quality and safety projects segregated from the provision of actual patient care.⁴³

Gaps in Assessment

Complicating the gaps that have been created by the silos of medical education and health care delivery systems as well as the silos of medical education and QI, there are also gaps within medical education that inhibit efforts to establish the association between outcomes in medical education and outcomes for patients.

Unwanted Variability in Assessment

The first gap that traditional assessment efforts present is variability in performance judgments between assessors.⁴⁴⁻⁴⁶ For example, Kogan and colleagues found that assessors use different frames of reference (e.g., self versus other residents at same stage of training), high levels of inference (e.g., resident crossing arms means he is comfortable versus uncomfortable in patient encounter), variable ways of translating resident performance into a numerical rating (e.g., from a lack of understanding of what the numbers mean), and factors external to the resident that influence their ratings (e.g., feared response of the resident to feedback).⁴⁶ The issue of numeric ratings is also raised by Gingerich and colleagues, who found that assessors form categorical judgments about trainees and may introduce errors when converting these to performance ratings.⁴⁷ Gingerich et al have also described the fallibility of assessors; they can be limited by human cognition and immediate context.⁴⁸ Finally, expertise can influence an assessor's decisions,⁴⁵ and performance variability across contexts can also serve as sources of unwanted variability in assessment.⁴⁹

Alignment of Assessment to Patient Care

A second gap that traditional performance assessments pose is that they have not always been aligned with determining requisite skills to provide the care that patients need, often not including patients in the assessment equation at all.¹ However, recent advancements in

the field of assessment in medical education are taking clear steps to consider the patient in performance assessment. As an example, in 2016, Norcini advocated that assessment should support the movement toward patient-centered care by enabling change (through delineating requisite competencies, ensuring they will be taught), enhancing and creating learning (through formative efforts), and ensuring competence (through summative efforts).⁵⁰ Furthermore, the 2018 Consensus Framework for Good Assessment states that the interaction of assessments with both educational as well as health care systems should be a focus of future work.⁵¹ Finally, Kogan and colleagues have proposed the Accountable Assessment for Quality Care and Supervision (AAQCS) equation, which posits that safe, effective, patient-centered care is the product of trainee performance and appropriate supervision.⁵²

Patient-Centered Performance Assessment: Promise in Entrustable Professional Activities and Quality Measures

Entrustable Professional Activities (EPAs)

The AAQCS equation explicitly brings the patient into the conceptual framework for assessment. However, there are practical challenges with executing it. For example, supervisors may not, perhaps even likely do not, provide appropriate supervision.⁵³ This lacuna has led Carraccio and Englander to partner with Holmboe and Kogan to propose a unifying framework of milestones, competencies, and EPAs to practically apply the AAQCS equation in the clinical learning environment.⁵⁴ EPAs are foundational activities of a profession that an individual can be trusted to do with varying levels of supervision, culminating in being entrusted to perform the activity without supervision and then being trusted to supervise others.^{55,56} EPAs represent a synthetic approach to assessment that integrates competencies and their related milestones. The latter are brief behavioral narratives that describe stages in development from novice to expert.⁵⁷⁻⁵⁹ Carraccio and colleagues note that EPAs can provide a shared mental model of performance that can decrease unwanted variability in assessment, addressing known challenges with traditional assessment efforts.⁵⁴ Furthermore, EPAs provide a lens of supervision and the opportunity to assess learners using a construct-aligned scale,⁶⁰ reflecting the priorities and expertise of assessors who supervise trainees in the clinical learning environment. These types of scales have been shown to be more reliable than other global scales in work-based assessment efforts.^{61,62}

Quality Measures

With known gaps created by the silos of medical education and QI, quality measures that are sensitive to the work that residents complete can not only enhance meaningful involvement of residents in QI work but can also serve as a patient-centered assessment approach, indicating the quality of care provided by residents.²⁹ Using such metrics for assessment purposes also

addresses the gaps created by the silos of medical education and health care delivery, focusing on the intersection of educational and patient care outcomes.

Warm and Mathis have suggested that ACGME requirements for residency training should include measuring and improving patient care outcomes.⁶³ Furthermore, Jaffe and colleagues have advocated partnering with institutional quality improvement experts to design resident quality and safety reports.⁶⁴ However, traditional quality metrics do not align well with the work that residents perform.⁶⁵ Paradoxically, quality measures sometimes attributed to supervising physicians may actually be attributable to residents.⁶⁶

Once quality measures relevant to the work that residents perform are clearly delineated, scholars have advocated the use of electronic health records (EHRs) to produce such measures.²⁹ The EHR has already been used to determine resident experience (e.g., conditions seen),⁶⁷ making its use for obtaining quality metrics for residents a logical next step. However, the EHR was not designed for this purpose, so using it in this manner requires careful thought.⁶⁸ To this end, Fiske and colleagues have advocated for health information counselors to help harness large amounts of data into information that is meaningful to clinical practice.⁶⁹ There are also an increasing number of studies investigating the use of machine learning to automate physician assessment,⁷⁰ and learning analytics is advocated for organizing and presenting large amounts of assessment data in a useful format.⁷¹

Challenges and Limitations to Patient-Centered Assessment Using EPAs and Quality Measures

EPAs

As noted previously, EPAs offer an assessment approach that aligns with the tasks that supervisors are already performing in the workplace: determining how much supervision is needed for residents to ensure safe and effective care for patients with conditions they have intimate familiarity with. However, most EPA efforts to date have defined EPAs without placing focus on empirical evidence, including the practical implementation of them.⁷²⁻⁷⁵ Thus, whether or not they present an assessment approach with higher utility⁷⁶ or better sensitivity to care that patients receive when practically implemented is not yet known. Furthermore, faculty may favor making informal entrustment decisions that can be reconsidered based on circumstances rather than summative ones, which could limit a focus on outcomes.⁷⁷ Finally, Hatala and colleagues have questioned whether entrustment ratings are just another rating, with the same assessor variability challenges discussed previously, or if they are capturing meaningful supervision decisions.⁷⁸

Quality Measures

Quality measures relevant to the work that residents complete face a few fundamental challenges. First, attributing care to individual residents presents challenges, with Sebok-Syer and colleagues arguing:

Some statistical approaches, such as variance component analysis, can quantitatively ascribe a portion of variance in patient care and/or outcomes attributable to differences between residents as opposed to other members of the team. However, although useful, such an approach does not explain how specific residents' behavior actually contributes to patient care or outcomes. Furthermore, attribution, or the focus on an individual resident, may fail to consider the various interactions that exist between residents and other members of a health care team that produce care processes or outcomes.⁷⁹

Indeed, individual competence is necessary but insufficient, with collective, or team, competence having fundamental importance given that care is provided by teams.⁸⁰ Even looking at a smaller unit than the entire team, performance is the result of a dyad, composed of, for example, a trainee and a supervisor.⁷⁹

Another potential criticism of quality measures for individual residents is that they will likely, by definition, need to be process measures, although outcome measures are preferable. However, most existing quality measures are process measures and not outcome measures, so perhaps this potential concern can be assuaged.⁸¹ Adding support to this view, Kaplan and colleagues demonstrated the ability to identify individual physician effects among diabetes process and intermediate outcome measures in the practice environment outside training.⁸² However, the potential differences in attributing measures to residents compared to physicians in practice following training, which Kaplan et al considered, may be another fundamental challenge to quality measures for residents. For example, it is possible to use billing data to attribute care to attending physicians for the purpose of determining patient-level metrics in the inpatient setting, but residents do not bill, making a similar approach infeasible in GME.⁸³

A final challenge to using quality measures among residents is the lack of alignment of many existing measures with the work that residents perform.⁶⁵ Therefore, there is a need to both identify existing measures and also create new ones.²⁹

Main Argument and Research Questions

Over the past several decades, there have been calls to better evaluate the outcomes of medical education and determine if they meet the needs of patients. In recent years, these calls have

increased and the QI movement has become a mainstream focus in medicine. However, the association between outcomes in medical education and patient care outcomes is still not well understood. Furthermore, most efforts in this area seek to understand relationships at the program and institutional level. However, we believe that focusing on the individual resident is also paramount. We graduate, certify, and credential individuals, and this is not likely to change in the foreseeable future. Therefore, considering the association between individual learner outcomes and the care these individuals provide is important, and we believe that quality measures can provide a path forward in this area. As noted above, there are insufficient quality measures for the work that residents complete; this lack has led us to our main research question:

What are the characteristics of developing and implementing quality measures, which we term resident-sensitive quality measures (RSQMs), that are largely attributable to residents and important to patient care?

As detailed above, we believe that EPAs, in addition to RSQMs, can provide a patient-focused approach to assessment. Therefore, we defined a secondary research question to further explore the association between individual learner outcomes and the care these individuals provide:

What is the association between entrustment decisions made about residents and their performance on RSQMs?

Our studies took place in the U.S. pediatric residency context. Specifically, we focused on the pediatric emergency department for our initial efforts in this novel research direction.

Research Paradigms Underlying the Research Studies

We used post-positivist paradigm for most of the studies in this program of research.⁸⁴ We assume that there is an association between entrustment decisions and RSQMs, but reality makes this association too complex to be fully understood. Similarly, objective knowledge of reality is our goal, but that is not achievable. Rather, we must focus on testing and rejecting hypotheses about the relationship between entrustment decisions and RSQMs.

As detailed below, one of the studies employs a constructivist research paradigm.⁸⁴ This study focuses on how clinical competency committee members (individuals who review resident assessment data for the purpose of making summative decisions) view, use, and prioritize RSQMs, a novel type of assessment data, alongside assessment data they traditionally review. In this study, we believe multiple truths can exist and that those truths are created by and between individuals. In other words, the reality of how clinical competency committee members

view, use, and prioritize RSQMs is socially based and can change. Therefore, knowledge must be co-constructed by research participants and researchers through dialogue.

Overview of Studies

In *chapter 2*, we provide additional detail for why we believe EPAs and quality measures can provide a future for resident performance assessment that ensures high quality care. Then, in *chapter 3*, we detail our justification and approach for focusing on individuals when many scholars are focusing on programs and institutions in seeking to determine the association between educational outcomes and patient care outcomes. We then turn to our approach for developing and prioritizing RSQMs in *chapters 4 and 5*, detailing the RSQMs we developed for three illnesses (asthma, bronchiolitis, and closed head injury) in the pediatric emergency department setting. *Chapter 6* details the performance of these RSQMs when implemented in the pediatric emergency department. In *chapter 7* focus turns to the association between RSQMs and entrustment decisions as well as patient encounter characteristics. *Chapter 8* seeks to determine how RSQMs may be used as assessment data for making summative assessment decisions at the level of the clinical competency committee. *Chapter 9* looks to the future by delineating a model for attributing patients to residents in the EHR, which can in turn lead to attributing RSQMs to patients using EHR data. Finally, *chapter 10* makes a call for medical education to take the lead in achieving high value health care through placing a focus on the patient, citing RSQMs as a patient-focused assessment approach to realize this goal.

Since this thesis is based on published journal articles, some overlap is inevitable.

References

1. 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.
2. Hall JN, Lloyd T. Defining the target: Competency-based medical education for better patient care. *Med Teach*. 2018;40(6):644.
3. Magraw RM, Fox DM, Weston JL. Health professions education and public policy: A research agenda. *J Med Educ*. 1978;53(7):539-546.
4. Weinstein DF. Optimizing GME by measuring its outcomes. *N Engl J Med*. 2017;377(21):2007-2009.
5. 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(1):1.
6. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA*. 2009;302(12):1277-1283.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Phillips RL, Petterson SM, Bazemore AW, Wingrove P, Puffer JC. The effects of training institution practice costs, quality, and other characteristics on future practice. *Ann Fam Med*. 2017;15(2):140-148.
13. Mattar SG, Alseidi AA, Jones DB, et al. General surgery residency inadequately prepares trainees for fellowship. *Ann Surg*. 2013;258(3):440-449.
14. Wiest FC, Ferris TG, Gokhale M, Campbell EG, Weissman JS, D B. Preparedness of internal medicine and family practice residents for treating common conditions. *JAMA*. 2002;288(20):2609-2614.
15. 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.
16. Eden J, Berwick D, Wilensky G. *Graduate Medical Education That Meets the Nation's Health Needs*. Washington, DC: Institute of Medicine of the National Academies. 2014.
17. Medicare Payment Advisory Commission. Report to the Congress: Improving incentives in the Medicare program, June 2009.

18. Medicare Payment Advisory Commission. Graduate medical education financing: Focusing on educational priorities. In: *Report to Congress: Aligning Incentives in Medicare*. Washington, DC 2010:103-126.
19. Denson JL, McCarty M, Fang Y, Uppal A, Evans L. Increased mortality rates during resident handoff periods and the effect of ACGME duty hour regulations. *Am J Med*. 2015;128(9):994-1000.
20. Denson JL, Jensen A, Saag HS, et al. Association between end-of-rotation resident transition in care and mortality among hospitalized patients. *JAMA*. 2016;316(21):2204-2213.
21. Silkens MEWM, Arah OA, Wagner C, Scherpier AJJA, Heineman MJ, Lombarts KMJM. The relationship between the learning and patient safety climates of clinical departments and residents' patient safety behaviors. *Acad Med*. 2018;93(9):1374-1380.
22. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. *Acad Med*. 2018;93(7):975-978.
23. Chahine S, Kulasegaram K, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Acad Med*. 2018;93(6):829-832.
24. Arora VM. Harnessing the power of big data to improve graduate medical education: Big idea or bust? *Acad Med*. 2018;93(6):833-834.
25. Weinstein DF. The elusive goal of accountability in graduate medical education. *Acad Med*. 2015;90(9):1188-90.
26. Gillespie C, Zabar S, Altshuler L, et al. The research on medical education outcomes (ROMEO) registry: Addressing ethical and practical challenges of using "bigger," longitudinal educational data. *Acad Med*. 2016;91(5):690-695.
27. Triola MM, Hawkins RE, Skochelak SE. The time is now: Using graduates' practice data to drive medical education reform. *Acad Med*. 2018;93(6):826-828.
28. Caverzagie KJ, Lane SW, Sharma N, et al. Proposed performance-based metrics for the future funding of graduate medical education: Starting the conversation. *Acad Med*. 2018;93(7):1002-1013.
29. Smirnova A, Sebok-Syer SS, Chahine S, et al. Defining and adopting clinical performance measures in graduate medical education: Where are we now and where are we going? *Acad Med*. 2019;94(5):671-677.
30. 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.
31. Ludmerer K. *Time to Heal: American Medical Education from the Turn of the Century to the Era of Managed Care*. New York, NY: Oxford University Press, Inc.; 1999.
32. Roemer BM, Azevedo T, Blumberg B. Looking at graduate medical education through a different lens: A health care system's perspective. *Acad Med*. 2015;90(9):1231-5.
33. Sklar DP, Hemmer PA, Durning SJ. Medical education and health care delivery: A call to better align goals and purposes. *Acad Med*. 2017;93(3):384-390(387).
34. 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;92(12):1740-1748.

35. Sargeant J, Wong BM, Campbell CM. CPD of the future: A partnership between quality improvement and competency-based education. *Med Educ.* 2018;2(1):125-135.
36. Ackerman KG, Lee B, Kushner JA. Dissatisfaction with maintenance of certification in academic pediatrics. *Pediatr Res.* 2015;79:240.
37. Teirstein PS. Boarded to death—Why maintenance of certification is bad for doctors and patients. *N Engl J Med.* 2015;372(2):106-108.
38. Weiss KB, Bagian JP, Nasca TJ. The clinical learning environment: The foundation of graduate medical education. *JAMA.* 2013;309(16):1687-1688.
39. Nasca TJ, Wiess KB, Bagian JP. Improving clinical learning environments for tomorrow's physicians. *N Engl J Med.* 2014;370(11):991-992.
40. Wagner R, Weiss KB, Passiment ML, Nasca TJ. Pursuing excellence in clinical learning environments. *J Grad Med Educ.* 2016;8:124-127.
41. Weiss KB, Co JPT, Bagian JP, Committee CE. Challenges and opportunities in the 6 focus areas: CLER National Report of Findings 2018. *J Grad Med Educ.* 2018;10(4 Suppl):25-48.
42. Butler JM, Anderson KA, Supiano MA, Weir CR. "It feels like a lot of extra work": Resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med.* 2017;92(7):984-990.
43. Cooke M, Ironside PM, Ogrinc GS. Mainstreaming quality and safety: A reformulation of quality and safety education for health professions students. *BMJ Qual Saf.* 2011;20(Suppl 1):i79-i82.
44. Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtens AM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract.* 2007;12(2):239-260.
45. Govaerts MJ, Schuwirth LW, van der Vleuten CP, Muijtens AM. Workplace-based assessment: Effects of rater expertise. *Adv Health Sci Educ Theory Pract.* 2011;16:151-165.
46. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: A conceptual model. *Med Educ.* 2011;45(10):1048-1060.
47. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: Rethinking the etiology of rater errors. *Acad Med.* 2011;86(10):S1-S7.
48. Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: assessor cognition from three research perspectives. *Med Educ.* 2014;48(11):1055-1068.
49. Eva KW. On the generality of specificity. *Med Educ.* 2003;37(7):587-588.
50. Norcini JJ. The role of assessment in supporting the movement toward patient-centred care. *Perspect Med Educ.* 2017;6(1):5-6.
51. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach.* 2018:1-8.
52. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014;89(5):721-727.
53. Kennedy TJ, Regehr G, Baker GR, Lingard LA. 'It's a cultural expectation...' The pressure on medical trainees to work independently in clinical practice. *Med Educ.* 2009;43(7):645-653.

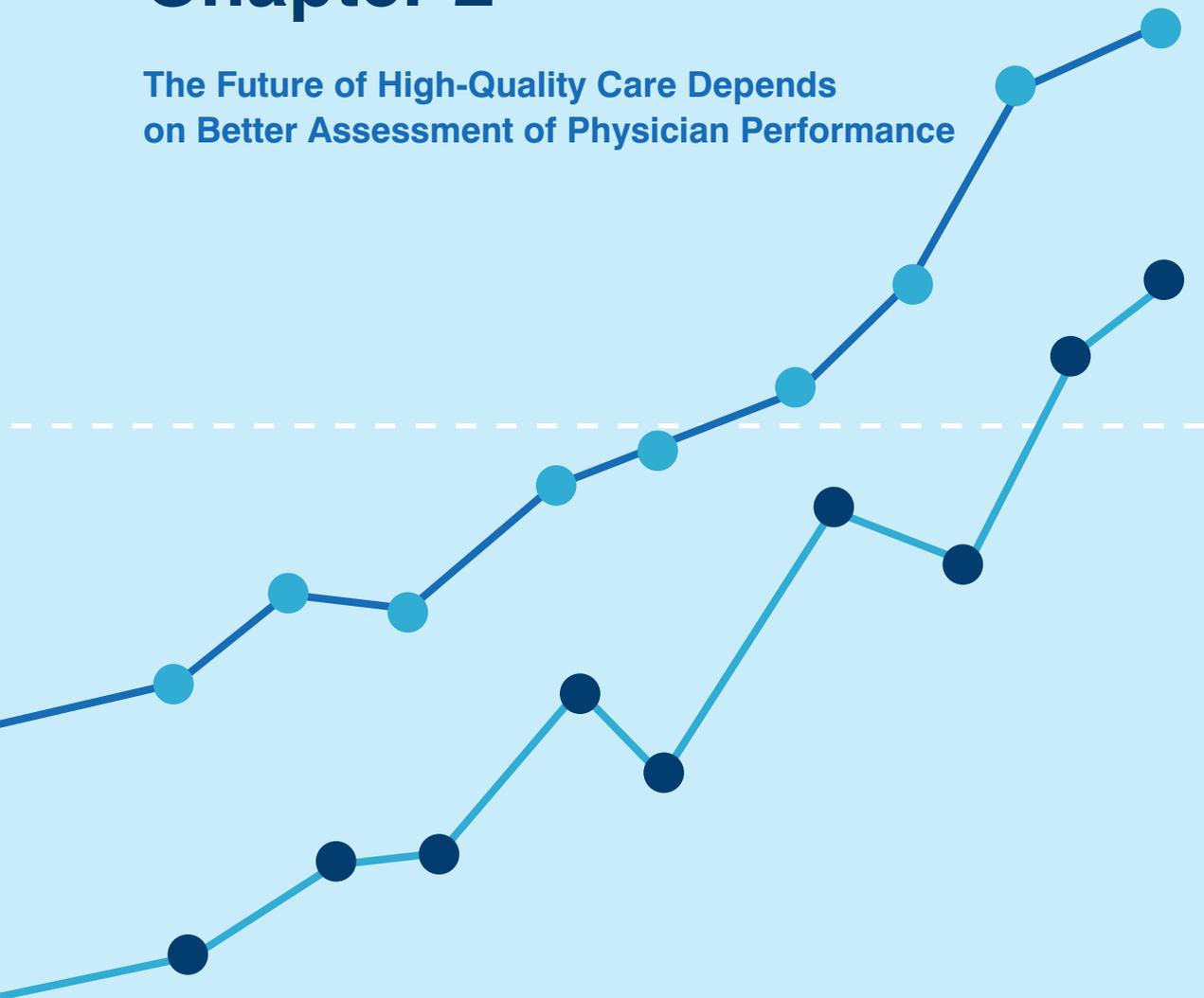
54. Carraccio C, Englander R, Holmboe ES, Kogan JR. Driving care quality: Aligning trainee assessment and supervision through practical application of entrustable professional activities, competencies, and milestones. *Acad Med.* 2016;91(2):199-203.
55. ten Cate O SF. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542-547.
56. ten Cate O, Snell L, Carraccio C. Medical competence: The interplay between individual ability and the health care environment. *Med Teach.* 2010;32:669-675.
57. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—Rationale and benefits. *N Engl J Med.* 2012;366:2324-2325.
58. Hicks PJ, Schumacher DJ, Benson BJ, et al. The pediatrics milestones: Conceptual framework, guiding principles, and approach to development. *J Grad Med Educ.* 2010;2(3):410-418.
59. Hicks PJ, Englander R, Schumacher DJ, et al. Pediatrics milestone project: Next steps toward meaningful outcomes assessment. *J Grad Med Educ.* 2010;2(4):577-584.
60. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: Outlining their usefulness for competency-based clinical assessment. *Acad Med.* 2016;91(2):186-190.
61. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: Construct alignment improves the performance of workplace-based assessment scales. *Med Educ.* 2011;45(6):560-569.
62. Crossley J, Jolly B. Making sense of work-based assessment: Ask the right questions, in the right way, about the right things, of the right people. *Med Educ.* 2012;46(1):28-37.
63. Warm EJ, Mathis BR. Ambulatory education: Time to move from process to outcome. *J Grad Med Educ.* 2019;11(2):143-145.
64. Jaffe R, Diemer G, Caruso J, Metzinger M. Creating provider-level quality reports for residents to improve the clinical learning environment. *J Grad Med Educ.* 2017;9(3):381-382.
65. 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.
66. Lau BD, Streiff MB, Pronovost PJ, Haider AH, Efron DT, Haut ER. Attending physician performance measure scores and resident physicians' ordering practices. *JAMA Surg.* 2015;150(8):813-814.
67. Levin JC, Hron J. Automated reporting of trainee metrics using electronic clinical systems. *J Grad Med Educ.* 2017;9(3):361-365.
68. Sebok-Syer SS, Goldszmidt M, Watling CJ, Chahine S, Venance SL, Lingard L. Using electronic health record data to assess residents' clinical performance in the workplace: The good, the bad, and the unthinkable. *Acad Med.* 2019;94(6):853-860.
69. Fiske A, Buyx A, Prainsack B. Health information counselors: A new profession for the age of big data. *Acad Med.* 2019; 94(1):37-41.
70. Dias RD, Gupta A, Yule SJ. Using machine learning to assess physician competence: A systematic review. *Acad Med.* 2019; 94(3):427-439.
71. Chan T, Sebok-Syer S, Thoma B, Wise A, Sherbino J, Pusic M. Learning analytics in medical education assessment: The past, the present, and the future. *AEM Educ Train.* 2018; 2(2):178-187.

72. Hauer KE. Seeking trust in entrustment: Shifting from the planning of entrustable professional activities to implementation. *Med Educ.* 2019;53(8):752-754.
73. Meyer EG, Chen HC, Uijtdehaage S, Durning SJ, Maggio LA. Scoping review of entrustable professional activities in undergraduate medical education. *Acad Med.* 2019.
74. Shorey S, Lau TC, Lau ST, Ang E. Entrustable professional activities in health care education: a scoping review. *Med Educ.* 2019;53(8):766-777.
75. Mink RB, Schwartz A, Herman BE, et al. Validity of level of supervision scales for assessing pediatric fellows on the common pediatric subspecialty entrustable professional activities. *Acad Med.* 2018;93(2):283-291.
76. van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Adv in Health Sci Educ.* 1996;1(1):41-67.
77. van Loon KA, Teunissen PW, Driessen EW, Scheele F. Formal versus informal judgments: Faculty experiences with entrustment in graduate medical education. *J Grad Med Educ.* 2018;10(5):537-542.
78. Hatala R, Ginsburg S, Hauer KE, Gingerich A. Entrustment ratings in internal medicine training: Capturing meaningful supervision decisions or just another rating? *J Gen Intern Med.* 2019;34(5):740-743.
79. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: Implications for assessment and entrustment. *Med Educ.* 2018;52(Suppl. 1):23-24.
80. Lingard L. Paradoxical truths and persistent myths: Reframing the team competence conversation. *J Contin Educ Health Prof.* 2016;36 Suppl 1:S19-21.
81. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med.* 2016;374(6):504-506.
82. Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Improving the reliability of physician performance assessment: Identifying the "physician effect" on quality and creating composite measures. *Med Care.* 2009;47(4):378-387.
83. Herzke CA, Michtalik HJ, Durkin N, et al. A method for attributing patient-level metrics to rotating providers in an inpatient setting. *J Hosp Med.* 2018;13(7):470-475.
84. Bergman e dFJ, Frambach J, Godefrooij M, Slootweg I, Stalmeijer R, van der Zwet J. A guide to research paradigms relevant to medical education. *Acad Med.* 2012;87(4):545.



Chapter 2

The Future of High-Quality Care Depends on Better Assessment of Physician Performance



Schumacher DJ, van der Vleuten CPM, Carraccio CL. "The Future of High-Quality Care Depends on Better Assessment of Physician Performance."

JAMA Pediatrics 2016;170(12):1131-1132

Ensuring that the performance of physicians meets the needs of patients is central to safe care. However, assessment of physician performance across the continuum of training and practice is inadequate. Often isolated as only an education problem, this assessment gap is also a clinical care problem.¹

Clinical Training

The goal of competency-based education, the purported focus of graduate medical education training during the past 2 decades, is to design and implement educational programs that achieve outcomes necessary to optimize patient and population outcomes.² Unfortunately, performance of residency graduates suggests this goal is often not achieved.^{3,4} Indeed, the Medicare Payment Advisory Committee and Institute of Medicine have focused on shortcomings in care delivery, calling for changes in reimbursements and transformation funding, respectively, to address them.^{3,4}

A partial explanation for these shortcomings is the failure to align training curricula with the requisite skills for practice. However, inability to achieve optimal assessment of trainee performance in the workplace also presents a significant barrier to adequate judgments of if, and when, trainees are prepared to meet the needs of patients. Addressing these deficiencies is imperative because ensuring adequate preparation (through training) and ability (through assessment) to care for patients is central to safe care.

The Transition to Practice

Concerns around clinical performance assessment continue after training. Initial board certification and credentialing in the United States focus on a combination of residency program attestation regarding Accreditation Council for Graduate Medical Education competencies, knowledge assessment through certifying examinations, professional references, and procedure logs when applicable. The most robust component of clinical performance assessment in this list, the first item, depends on valid and reliable decisions made by residency and fellowship program leadership. While this is appropriate, it places significant weight on these decisions, which are often made without valid and reliable assessment data.⁵

Subsequent maintenance of certification through medical boards focuses on an unrestricted license to practice, continued knowledge acquisition and assessment, and quality improvement. However, work-based performance assessment is left to individual institutions employing physicians. Ongoing activities to maintain privileges at these institutions typically focus on

meeting Joint Commission requirements for ongoing professional practice evaluation and focused professional practice evaluation, the scope of which are determined by the institution's choice and interpretation of measures. Unfortunately, assessment of actual performance through these activities is often based on broad performance measures designed to be one size fits all (e.g., readmission within 48 hours that is impacted by several factors beyond the individual practitioner), easy to document and/or achieve but not necessarily meaningful (e.g., 10 sedations completed within a year to maintain this privilege), or some combination of these characteristics.

Better Performance Assessment: A Framework

Recent advances in work-based assessment in medical education show promise for advancing assessment of physician performance to ensure high-quality care. Providing an exemplary framework, the concept of entrustable professional activities (EPAs), first introduced about a decade ago and gaining exponential international momentum over the past few years, conceptualizes physician assessment in a patient-centered manner.⁶ Beginning with the end in mind, EPAs delineate the major “professional activities” an individual in a given specialty must be “entrusted” to perform to provide quality care for patients. Examples include caring for a well newborn (pediatrician), performing a laparoscopic procedure (general surgeon), and leading a health care team (all specialties). Level of supervision needed to safely perform each professional activity serves as the basis for assessment. When an individual no longer requires supervision, he or she is entrusted with unsupervised practice in that EPA. This framework, consistent with the intent of competency-based education, originates from the question, “What essential activities does a given physician need to perform to safely and effectively care for patients?” Importantly, EPAs focus on synthetic activities of physicians that span several competencies and their associated milestones. This is contrary to the traditional approach to medical education, which typically focuses on more granular aspects of physician competence (e.g., gathering a history, communicating with members of the health care team, and acting professionally) with the hope that the sum of the parts will produce a good physician in the end.

From a clinical performance assessment standpoint, EPAs offer distinct advantages over traditional approaches. First, they situate assessment in tasks that assessors understand deeply. For example, most pediatricians have a robust conceptualization of what caring for a well newborn looks like. This stands in stark contrast to the typical assessment task in medicine that focuses on more isolated granular elements of performance such as “clinical reasoning.” Here, we generally call on an untrained assessor to rate performance on a more cognitively, rather than behaviorally, based task without a shared mental model of what constitutes desired performance. Indeed, recent work by Crossley and Jolly⁷ has found that assessment methods

that are aligned with what the assessor actually does in the clinical learning environment (e.g., supervise learners) have greater reliability and utility. The EPA framework also places focus on what matters for the patient: Can the individual safely perform this activity without supervision, and, if not, how much supervision do they need to ensure safe and effective care delivery? Finally, EPAs provide the potential to link the educational continuum, using core EPAs for entering residency as building blocks for specialty EPAs that can serve as a framework for performance assessment through training and into at least the first cycle of maintenance of certification.

Better Performance Assessment: A Call to Action and Research Agenda

Two things seem clear: (1) the future of high-quality care depends on better assessment of physician performance, and (2) contemporary approaches to assessment, including EPAs, provide promise for better work-based assessment. To maximize the potential in contemporary assessment approaches to ensure physicians can be entrusted to provide high-quality care, a well-aligned innovation and research agenda is critical. This agenda should be simple yet bold, reaching yet practical, and future focused yet grounded in what can begin today.

We recommend the following items form the foundation of this agenda:

1. Entrustable professional activities should be implemented into programs of assessment across the continuum of training and practice. This task should focus on identifying core EPAs that compose each specialty and asking assessors in the authentic clinical environment to choose the amount of supervision needed for performing each EPA. The number and sources of assessments to make valid and reliable entrustment decisions (i.e., deciding the level of supervision a physician needs) for each given EPA is unknown. Making meaningful entrustment decisions using EPAs as an assessment framework will require a combination of quantitative data, qualitative data, a variety of data sources, and expert consensus.
2. Research should target the association between entrustment decisions and indicators of quality care delivered by physicians. This will require identifying and accessing existing databases of relevant measures as well as developing new measures in performance areas where they are currently lacking. Achieving these goals and then determining the association between quality measures and entrustment decisions will produce much needed validity and reliability evidence to link educational and patient care outcomes.
3. With only 2% of global health expenditure dedicated to health education for all professions combined,² these efforts should receive funding and resource prioritization from public and private sectors.

References

1. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med*. 2014;89(5):721-727.
2. 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.
3. Medicare Payment Advisory Committee. Graduate Medical Education financing: focusing on educational priorities. In: *Report to Congress*. Washington, DC: Aligning Incentives in Medicare; 2010:103-126.
4. Committee on the Governance and Financing of Graduate Medical Education. *Graduate Medical Education That Meets the Nation's Health Needs*. Washington, DC: National Academies Press; 2014.
5. Driessen E, Scheele F. What is wrong with assessment in postgraduate training? lessons from clinical practice and educational research. *Med Teach*. 2013;35(7):569-574.
6. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542-547.
7. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Med Educ*. 2012;46(1):28-37.



Chapter 3

The Power of Contribution and Attribution in Assessing Educational Outcomes for Individuals, Teams, and Programs



Schumacher DJ, Dornoff E, Carraccio C, Busari J, van der Vleuten C, Kinnear B, Kelleher M, Sall D, Warm E, Martini A, Holmboe E. "The Power of Contribution and Attribution in Assessing Educational Outcomes for Individuals, Teams, and Programs." [published online ahead of print December 10, 2019]. *Academic Medicine*. doi: 10.1097/ACM.0000000000003121

Abstract

Recent discussions have brought attention to the utility of contribution analysis for evaluating the effectiveness and outcomes of medical education programs, especially for complex initiatives such as competency-based medical education. Contribution analysis focuses on the extent to which different entities *contribute* to an outcome. Given that health care is provided by teams, contribution analysis is well suited to evaluating the outcomes of care delivery. Furthermore, contribution analysis plays an important role in analyzing program- and system-level outcomes that inform program evaluation and program-level improvements for the future. Equally important in health care, however, is the role of the individual. In the overall contribution of a team to an outcome, some aspects of this outcome can be *attributed* to individual team members. For example, a recently discharged patient with an unplanned return to the emergency department to seek care may not have understood the discharge instructions given by the nurse or may not have received any discharge guidance from the resident physician. In this example, if it is the nurse's responsibility to provide discharge instructions, that activity is attributed to him or her. This and other activities attributed to different individuals (e.g., nurse, resident) combine to contribute to the outcome for the patient. Determining how to tease out such attributions is important for several reasons. First, it is physicians, not teams, that graduate and are granted certification and credentials for medical practice. Second, incentive-based payment models focus on the quality of care provided by an individual. Third, an individual can use data about his or her performance on the team to help drive personal improvement. In this article, the authors explored how attribution and contribution analyses can be used in a complimentary fashion to discern which outcomes can and should be attributed to individuals, which to teams, and which to programs.

We split paradoxes so reflexively that we do not understand the price we pay for our habit. The poles of a paradox are like the poles of a battery: hold them together, and they generate the energy of life; pull them apart, and the current stops flowing.

Parker Palmer¹

Discussions over the past several years have focused on the importance of evaluating the outcomes of medical education programs and efforts.²⁻⁴ The call for this focus is not new.⁵ In fact, competency-based medical education (CBME) has become the prevailing approach to medical education over the past 2 decades, and it puts primary focus on ensuring that outcomes of medical education programs prepare individuals to meet the needs of patients.⁶⁻⁹ However, despite the popularity of CBME, meaningful transition to this approach, which focuses on the outcome of meeting the needs of populations that we serve, has been challenging. Stated simply, we do not reliably know if graduates are prepared for clinical practice after training because assessment of medical trainees typically lacks defensible validity and reliability, and information on education and patient care outcomes is limited at best. This problem is pervasive across the education, training, and practice continuum, and the Institute of Medicine (now the National Academy of Medicine) and the Medicare Payment Advisory Commission have bemoaned the shortcomings of residency training in meeting the needs of patients.^{10,11} Greater understanding and improvement in the area of medical education program outcomes as well as practice outcomes will require determining how some aspects of care and processes can be *attributed* to individuals, whereas other aspects of care and processes are the result of several individuals *contributing* to the outcome. However, only the latter has received focus in recent conceptual discussions in the literature.

Scholars have recently advocated for contribution analysis as an approach to meaningfully evaluate the effectiveness and outcomes of medical education programs,^{12,13} especially for complex initiatives for which it is important to determine how several components must come together to produce an outcome, such as those of CBME.¹⁴ Contribution focuses on the extent to which different entities *contribute* to an outcome. However, the full contribution story for how an outcome is produced contains components that can be *attributed* to individuals that have varying degrees of influence on producing that outcome. For example, a recently discharged patient with an unplanned return to the emergency department to seek care may not have understood the discharge instructions given by the nurse or may not have received any discharge guidance from the resident physician. In this example, activities that are *attributable* to individuals (e.g., nurse, resident) combine to *contribute* to the outcome for the patient.

As this example illustrates, health care is the result of teamwork, which is likely why contribution analysis has rightly received attention. However, focusing on aspects of care *attributed* to individuals and providing these individuals with data related to their performance on the team

also has value. Recent work by some authors of this paper has resulted in the creation of resident-sensitive quality measures. These quality measures are based on their importance to a specific illness as well as the likelihood of the action that is measured being performed by a resident and not by another member of the team or the team collectively. Thus, resident-sensitive quality measures are an example of an effort that can address outcomes of care attributable to individuals.^{15,16} While these measures are still being studied, they have potential utility given current trends in managed care payment models, such as the Merit-based Incentive Payment System, a program of the Centers for Medicare and Medicaid Services (CMS), that seek to reimburse for care provided partly on the basis of the demonstrated quality of that care provided by individuals. Furthermore, focusing on individuals may be the most effective way to drive their personal improvement.^{17,18} Additionally, without individual-level feedback, it is easy to fall prey to fundamental attribution error, a bias to view oneself as responsible for positive outcomes but not for undesirable ones.¹⁹ While program- and system-level outcomes are important for informing program evaluation and program-level improvements for the future, we can and should graduate, certify, and credential individuals, and we may be required to use managed care payment models—such as the CMS Merit-based Incentive Payment System—that are based on demonstrated quality of care provided by individuals. It is not practical to carry out processes such as certifying or credentialing at the team level; such efforts would require new decisions several times a day, each time a team member changed. Thus, it is important to be able to determine an individual's performance. In summary, while a focus on contribution is important, we also need to focus on what aspects of care can be largely *attributed* to individuals.

If we apply Parker Palmer's analogy,¹ quoted in the epigraph, to the topic at hand, we can see attribution and contribution as the two opposite poles of a battery. By considering both, we can generate sufficient voltage to power the understanding of outcomes in medical education for individuals as well as for teams and programs. However, if we follow the recent conversation in the medical education literature, which has focused solely on contribution analysis and advocated "moving away from attribution analysis in favor of contribution analysis,"¹² there's a risk that the "current stops flowing" to consideration of both individual- and program-level outcomes in medical education. Heeding the words of Palmer relies on the "power of 'and,'" that is, seeing the benefit that both contribution and attribution bring to analysis, making the whole greater than the sum of its parts. It is possible, perhaps even likely, that those who have written about contribution analysis also value a focus on attribution for different purposes. However, we believe that explicitly defining the role of attribution for individuals, as well as its role in the attribution-to-contribution continuum, is important for 2 reasons: realizing the full potential of attribution and making attribution transparent. Therefore, in this article, we will explore how attribution can be used to assess individuals in medical education. We will first briefly review contribution analysis to detail how this approach can be harnessed to create an argument for attribution analysis.

Then we will delineate how attribution analysis can be used to discern what outcomes can and should be attributed to individuals on teams, in programs, and within systems.

Focusing on Programs and Teams: Contribution Analysis

Contribution analysis emerged in 2001 as a means of qualitatively evaluating civic programs and the outcomes they are seeking to achieve. John Mayne, who developed contribution analysis, felt that the changing culture of public administration at the time required greater accountability for the outcomes of civic programs.²⁰ He thought that contribution analysis could help “managers, researchers, and policymakers to arrive at conclusions about the contribution their program has made to particular outcomes.”²⁰ Recent years, and the continued momentum of CBME, have seen similar calls for accountability in the outcomes of medical education programs.^{2-4,21} These circumstances make contribution analysis an excellent fit for this purpose, as medical education scholars have advocated.¹²

The complex nature of the real world means there are many influences on an outcome. The goal of contribution analysis is to determine the extent to which individual entities affect an overall outcome.²² In this analysis, it is assumed there will be a given level of uncertainty present in the determination of the drivers of the outcomes of interest. As shown in Table 1, Mayne describes 6 steps in contribution analysis^{12,20}: (1) set out the causal problem to be addressed, (2) develop the postulated theory of contribution and identify risks to the theory, (3) gather the existing evidence to support the theory of contribution, (4) assess the contribution story and the contestations to it, (5) seek out additional evidence, and (6) revise and strengthen the contribution story. As detailed in the next section, Table 1 further delineates how these steps can be applied to determining the contribution of program outcomes, the attribution of individual outcomes, and the interactions between contribution and attribution in CBME.

Table 1 Mayne's 6 Steps of Contribution Analysis Applied to Contribution of Program Outcomes, Attribution of Individual Outcomes, and Interactions Between Contribution and Attribution in Competency-Based Medical Education

Step	Contribution analysis (Mayne²⁰)	Contribution analysis to determine outcomes for programs	Attribution analysis to determine outcomes for individuals	Contribution-attribution interaction considerations
1	Set out the causal problem to be addressed.	Recognize how influencing factors may shape the implementation of the program.	Set out the attributable factor to the individual trainee on the team.	Competent individuals (attribution) can come together to form an incompetent team (contribution), but teams can be competent even when 1 team member is incompetent. ²⁷
2	Develop the postulated theory of contribution and identify risks to the theory.	Identify the factors that are most likely to influence the outcomes, as well as the assumptions and risks anticipated for the outcomes.	Explain and justify why these tasks are attributable to individual trainees, taking into account who else could be completing the tasks.	The capabilities and management style of the team leader affect the morale and motivation of the team in a positive or negative way; in the former, there may be more help given to other team members by those not primarily assigned to tasks. ²⁸
3	Gather the existing evidence to support the theory of contribution.	Gather as much concrete evidence as possible (e.g., quantitative, qualitative), including focusing on what pathways look like for residents as they progress and how curricular activities associated with the program contribute to development of these pathways.	Gather evidence demonstrating the linkage of the tasks to individual trainee actions, realizing that context is important: While the superficial elements of the task may change, the core elements remain stable. ^{29,30}	In the real world, most tasks are completed collectively by more than 1 individual member of the team (contribution). There are parts of tasks that may or may not be more attributable to individual members on that team. It is important to consider this overlap.
4	Assess the contribution story and the contestations to it.	Put steps 1 through 3 together, ensuring the context is thoroughly described, the theory of change is plausible, results are explained, and other possible factors are discussed.	Look at the big picture context created to this point and evaluate whether the attribution is valid. Work should include attempting to discern whether the process measure under consideration is truly attributable to an individual trainee rather than to another member of the team or to the team as a collective.	Assess whether activities felt to be attributable to individuals are done by those individuals or by others (either individuals or teams). Determine if activities performed by the team should actually be the responsibility of an individual on the team instead.

5	Seek out additional evidence.	Find additional data as necessary.	Continue to seek out more information to support the claim that the process measure is attributable to the individual trainee.	As noted in step 3, parts of tasks may be more attributable to individual members on a team than to the team as a whole. The importance of consideration of this overlap is reflected in this step, which seeks out additional evidence to consider this overlap and its resultant interactions.
6	Revise and strengthen the contribution story.	Use the new evidence from step 5 to create a stronger contribution story.	Reconsolidate the new information gathered in step 5 to create a clearer alignment of the relationship between the individual trainee and the task measured.	This is the final opportunity in a given iteration to consider the following: (1) Which activities that should be attributable to individuals are actually done by individuals and which are done by others (either individuals or teams)? (2) Which activities performed by the team should actually be performed by teams and which should be owned by an individual on the team?

Focusing on Individuals: Attribution Analysis

We believe that the contribution analysis process can be used to define attribution analysis as well, seeking to build an evidentiary argument for what activities can be largely or entirely attributed to individuals. Indeed, in his conceptualizations of contribution analysis, Mayne made the case that contribution begins with what can be attributed to individual entities; clearly, attribution is seen to be a part of the contribution story.^{20,23} The full contribution story of how an outcome is produced includes individuals, activities, and entities to whom varying degrees of influence in producing that outcome are *attributed*. However, discussions of applying contribution analysis in medical education have approached contribution and attribution as more of a dichotomy, advocating contribution over attribution.^{12,24} This tendency may be because of the focus on outcomes of CBME programs. As discussed previously, we view this conversation of attribution and contribution as a “both, and” rather than an “either, or” for medical education; both lenses can be very useful, depending on the desired focus. Indeed, Mayne conceptualized a continuum between attribution and contribution.

Mayne's Attribution–Contribution Continuum

In his early work, Mayne described a results chain where an “activity” produces 3 types of progressively more distal outcomes.²⁰ This chain, shown on the left side of Figure 1, begins with the “output/reach,” akin to process measures in medicine (e.g., patient has an asthma action plan documented with primary care provider). A little further out are “intermediate outcomes,” akin to medicine’s quality measures of the same name (e.g., steroid courses each year and pulmonary function testing). Most distal are “end outcomes,” akin to true outcome measures in medicine (e.g., number of emergency department visits or hospitalizations for asthma each year). As in medicine, Mayne’s model considers the additional factors beyond the individual that affect results further downstream. For example, intermediate and end outcomes can be affected by patient inhaler technique, access to medications, and the presence of dusts and molds in the home.

The second column of Figure 1 shows an example of the results chain for blood sugar control in a patient with diabetes, an activity that residents (the focus of this example) can learn. When only the resident caring for a patient with diabetes is considered, the following can be used to determine that resident’s performance: process measures such as ordering the proper medication regimen (e.g., metformin) and appropriate screening (e.g., eye exams, urine microalbumin). As Figure 1 shows, these activities can be mostly attributed to the individual resident (i.e., either the resident ordered or did not order a urine microalbumin). Moving down the results chain, intermediate outcome measures include actual patient diet and exercise, medication adherence, and hemoglobin A1c levels; end outcome measures include diabetic neuropathy, retinopathy, stroke, and myocardial infarction. These intermediate and end outcomes include factors that cannot be controlled by an individual resident (e.g., access to medications, patient medication adherence, actual patient diet) *contributing* to whether or not the measures are achieved. Thus, these measures are best positioned to be considered through the lens of contribution rather than attribution.¹³

The measures detailed in Figure 1 are not only relevant for patients but also for residents and their training programs as well as individual practitioners after training and their institutions. As this discussion illustrates, quality measures that can be used in medical education can and should be the same as those used in quality improvement efforts, such as the CMS Merit-based Incentive Payment System. Performance on the process measures reflects the individual provider; performance collectively, across multiple providers, indicates the outcomes of the program or clinic. Performance on the intermediate and end outcome measures reflects the performance of the patient, training program or clinic, and institution combined.

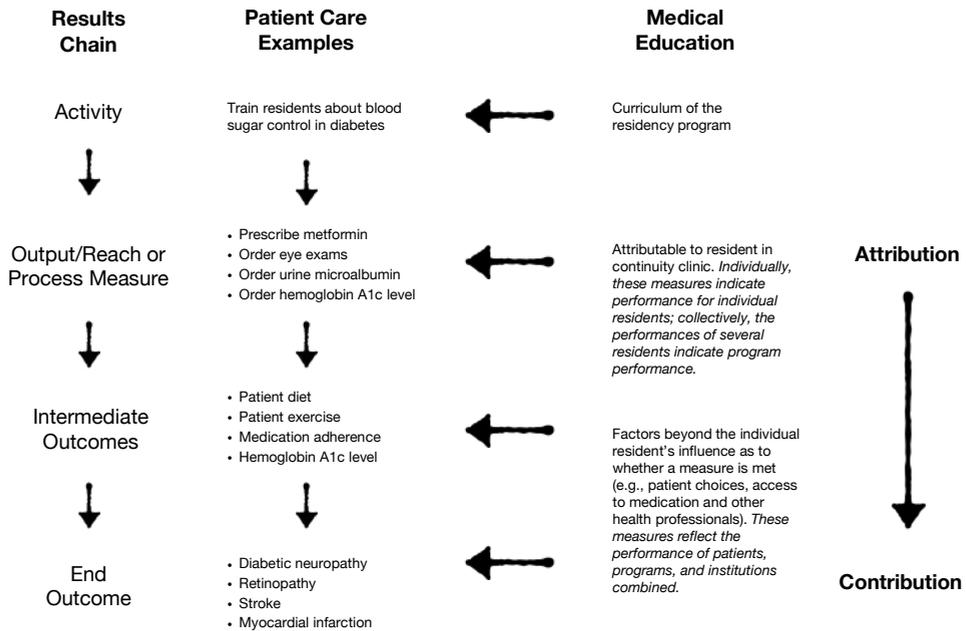


Figure 1: Process and outcome measures for patients, individuals, and programs. This results chain shows attribution and contribution in medical education for the activity of training residents about blood sugar control in diabetes. Starting with output/reach or process measures, which are attributable to individuals, the chain is completed with end outcomes, which result from many individuals' contributions.

Defining Attribution Analysis

Given the continuum between attribution and contribution, we believe contribution analysis can be applied to individuals. Specifically, we believe the same steps used in contribution analysis can be applied to individuals to define attribution analysis and to determine specific components of individuals' performances that can be attributed to them. This approach provides education scholars with a method for collecting evidence when they are attempting attribution studies rather than plugging in multiple variables and seeing "what sticks." We detail the steps for our proposed attribution analysis using residents as an example; we also compare these steps with the parallel steps of contribution analysis in Table 1:

1. *Set out the causal problem to address in individual performance.* For residents this includes tangible tasks that residents—rather than other members of the health care team—regularly complete. At this early step, it is important to consider whether the proposed attributable task should be within the scope of work of the individual of interest. The individual may only perform this task to compensate for a dysfunctional microsystem or interprofessional team in which the person (or people) who should

- be performing the task is not doing so. Such potential attributable factors should be excluded and focus placed on improving the microsystem or team performance.
2. *Develop the postulated theory of attribution and identify risks to the theory.* Explain and justify why these tasks are attributable to residents, taking into account who else could be completing the tasks.
 3. *Gather the existing evidence to support the theory of attribution.* Evidence should demonstrate that selected tasks actually reflect resident actions.
 4. *Assess the attribution story and the contestations to it.* Look at the big picture context created to this point and evaluate whether the attribution is valid in the context of the clinical learning environment. This evaluation should include attempting to understand whether the process measure under consideration is truly highly attributable to the resident rather than to another member of the team or to the team as a collective.
 5. *Seek out additional evidence.* Continue to seek out more information to support the claim that the process measure is attributable to the resident, for example, by conducting chart reviews and/or direct observation combined with discussions with a variety of stakeholders such as families and other members of the health care team.
 6. *Revise and strengthen the attribution story.* Integrate the new information gathered in step 5 to create a clearer alignment of the relationship between the resident and the task measured.

Further building from Mayne's conceptualization of attribution, completing these steps will help to infer "probabilistic causation" between an individual's actions and the measure.²⁵ Mayne asserts that 5 criteria must be met to establish this probabilistic causation: plausibility, implementation according to plan, evidentiary confirmation of key elements, taking other influencing factors into account, and disproving alternate explanations.

In previous work, we have taken some of the 6 steps of this attribution analysis plan in developing resident-sensitive quality measures.^{15,16} By taking this focus, we sought to address the requirements of step 1 and develop the postulated theory of attribution component of step 2. We then gathered existing evidence for this theory (step 3) and identified risks to it (the other component of step 2) by assembling content experts (residents and supervisors) to brainstorm and prioritize potential resident-sensitive quality measures. Our current work with these measures will achieve steps 4 through 6 and include implementing them in the clinical setting to determine if residents or others perform the work under consideration. We will also be able to ascertain whether these measures are indeed important to the quality of care for our initial diagnoses of interest.

The Power of Harnessing *Both* Contribution *and* Attribution

Using both contribution and attribution analyses, we can evaluate a program as a collective while also measuring the progress of the individuals that regulators must certify. Creating and using these measures will help to further resident education but can also be built upon by considering feedback from patients and team members (e.g., nurses) to create measures for more intangible skills, such as communication and patient rapport. This approach creates valuable feedback for individual improvement, with tangible goals while also providing actionable feedback to programs to better facilitate the education of their staffs. This “both, and” rather than “either, or” thinking is evoked in the epigraph.

Conclusion

The challenge of linking educational outcomes to patient care outcomes—our greatest challenge on the education horizon—may well be served by using contribution *and* attribution analyses as the first step toward pulling apart and examining the intertwined performance strands. Such an approach allows us to see how each strand (attribution of individual performance) is intertwined in team performance and how the whole thread (contribution analysis) links team performance to care outcomes for patients. In these efforts, medical education may benefit from following the trajectory of quality improvement, which began as a focus on individual responsibility for outcomes but evolved to focus on team responsibility with shared accountability among individuals and teams.²⁶

References

1. Palmer P. *The Courage to Teach: Exploring the Inner Landscape of a Teacher's Life, 20th Anniversary Edition*. San Francisco, CA: Jossey-Bass; 2017.
2. Weiss AP. Measuring the impact of medical research: Moving from outputs to outcomes. *Am J Psychiatry*. 2007;164(2):206-214.
3. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. *Acad Med*. 2018;93(7):975-978.
4. Chahine S, Kulasegaram K, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Academic Medicine*. 2018;93(6):829-832.
5. McGaghie WC MG, Sajid AW, Telder TV. *Competency-Based Curriculum Development in Medical Education: An Introduction*. Geneva, Switzerland: World Health Organization; 1978.
6. Frank JR, Snell LS, Cate OT, et al. *Competency-based medical education: Theory to practice*. *Med Teach*. 2010;32(8):638-645.
7. Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach*. 2010;32(8):651-656.
8. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: From Flexner to competencies. *Acad Med*. 2002;77(5):361-367.
9. Carraccio C, Englander R, Gihlooly J, et al. Building a framework of entrustable professional activities, supported by competencies and milestones, to bridge the educational continuum. *Acad Med*. 2017;92(3):324-330.
10. Medicare Payment Advisory Commission. *Report to the Congress: Improving Incentives in the Medicare Program*. Washington, DC: Medicare Payment Advisory Commission; June 2009.
11. Eden J, Berwick D, Wilensky G. *Graduate Medical Education That Meets the Nation's Health Needs*. Washington, DC: Institute of Medicine of the National Academies; 2014.
12. Van Melle E, Gruppen L, Holmboe ES, et al. Using contribution analysis to evaluate competency-based medical education programs: It's all about rigor in thinking. *Acad Med*. 2017;92(6).
13. Moreau KA, Eady K. Connecting medical education to patient outcomes: The promise of contribution analysis. *Med Teach*. 2015;37(11):1060-1062.
14. Dauphinee WD. Educators must consider patient outcomes when assessing the impact of clinical training. *Med Educ*. 2011;46(1):13-20.
15. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: A model from pediatric emergency medicine. *Acad Med*. 2018;93(7):1071-1078.
16. Schumacher DJ, Martini A, Holmboe E, et al. Developing resident sensitive quality measures: Engaging stakeholders to inform next steps. *Acad Pediatr*. 2019; 19(2): 177-185.
17. Sandars J CT. Self-regulation theory: Applications to medical education: AMEE Guide No. 58. *Med Teach*. 2011;33:875-886.
18. Schon DA. *The Reflective Practitioner: How Professionals Think in Action*. New York: NY: Basic Books, Inc; 1983.

19. Klein J. Attribution biases in assigning blame for medical error. *Med Educ.* 2017;51:982-991.
20. Mayne J. Addressing attribution through contribution analysis: Using performance measures sensibly. *The Canadian Journal of Program Evaluation.* 2001;16(1):1-24.
21. Weinstein DF. Optimizing GME by measuring its outcomes. *N Engl J Med.* 2017;377(21):2007-2009.
22. Patton MQ. A utilization-focused approach to contribution analysis. *Evaluation.* 2012;18(3):364-377.
23. Dybdal L, Nielsen SB, Lemire S. Contribution analysis applied: Reflections on scope and methodology. *The Canadian Journal of Program Evaluation.* 2010;25(2):29-57.
24. Govaerts MJB, van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: Moving beyond either-or thinking. *Med Educ.* 2019;53(1):64-775.
25. Mayne J. Contribution analysis: Addressing cause and effect. In: Forss K, Marra M, Schwartz R, eds. *Evaluating the Complex: Attribution, Contribution and Beyond.* New Brunswick, NJ: Transaction Publishers; 2011.
26. Dekker S. *Just Culture: Restoring Trust and Accountability in Your Organization.* 3rd ed. Boca Raton, FL: CRC Press; 2016.

References cited in Table 1 only.

27. Lingard L. Paradoxical truths and persistent myths: Reframing the team competence conversation. *J Contin Educ Health Prof.* 2016;36:S19-S21.
28. Baker DP, Salas E, King H, Battles J, Barach P. The role of teamwork in the professional education of physicians: Current status and assessment recommendations. *J Qual Patient Saf.* 2005;31(4):185-202.
29. Horsley T, Regehr G. When are two interventions the same? Implications for reporting guidelines in education. *Med Educ.* 2018;52(2):141-143.
30. Cianciolo AT, Regehr G. Learning theory and educational intervention: Producing meaningful evidence of impact through layered analysis. *Acad Med.* 2019;94(6):789-794.



Chapter 4

Developing Resident-Sensitive Quality Measures: A Model From Pediatric Emergency Medicine



Schumacher DJ, Holmboe E, van der Vleuten CPM, Busari JO, Carraccio C. "Developing Resident-Sensitive Quality Measures: A Model From Pediatric Emergency Medicine." *Academic Medicine* 2018;93(7):1071-1078.

Abstract

Purpose

To begin closing the gap with respect to quality measures available for use among residents, the authors sought to identify and develop resident-sensitive quality measures (RSQMs) for use in the pediatric emergency department (PED) setting.

Method

In May 2016, the authors reviewed National Quality Measures Clearinghouse (NQMC) measures to identify resident-sensitive measures. To create additional measures focused on common, acute illnesses (acute asthma exacerbation, bronchiolitis, closed head injury [CHI]) in the PED, the authors used a nominal group technique (NGT) and Delphi process from September to December 2016. To achieve a local focus for developing these measures, all NGT and Delphi participants were from Cincinnati Children's Hospital Medical Center. Delphi participants rated measures developed through the NGT in two areas: importance of measure to quality care and likelihood that measure represents the work of a resident.

Results

The review of NQMC measures identified 28 of 183 as being potentially resident sensitive. The NGT produced 67 measures for asthma, 46 for bronchiolitis, and 48 for CHI. These were used in the first round of the Delphi process. After two rounds, 18 measures for asthma, 21 for bronchiolitis, and 21 for CHI met automatic inclusion criteria. In round three, participants categorized the potential final measures by their top 10 and next 5.

Conclusions

This study describes a template for identifying and developing RSQMs that may promote high-quality care delivery during and following training. Next steps should include implementing and seeking validity evidence for the locally developed measures.

Large-scale studies in multiple specialties have shown an enduring correlation between residency training and subsequent patient care, as long as one to two decades after completion of training.¹⁻³ These findings call medical educators to consider the link between training and the quality of care delivery, underscoring two critical points. The first is that the imprinting that occurs during residency may have a lasting impact on patient care delivered after training. The second point is that, despite these findings, studies looking at the effects of residency training on patient care outcomes are uncommon.⁴

Beyond these concerns about the patient care delivered following training, evidence suggests that residents may also fail to provide optimal care during training.⁵⁻⁹ Additional studies across a broad range of specialties demonstrate that residency graduates are not prepared to meet the needs of patient populations even for foundational tasks related to common diagnoses such as asthma, diabetes, and upper respiratory infections.¹⁰⁻¹³ These shortcomings in quality care delivery have captured the attention of the Medicare Payment Advisory Committee and National Academy of Medicine. To address these shortcomings, these two organizations have called for changes in reimbursements and transformation in graduate medical education funding, respectively.^{10,14}

The Current State of Resident-Sensitive Quality Measures: A Clear Gap

Given the role residents play in ensuring quality care for current and future patients, it is clear that residents should be taught about quality metrics and how to use them during training. While some quality measures may be suitable for both residents and physicians in practice (e.g., communication with families), there have been cautions against the use of traditional quality measures among residents, including the argument that their fit is not optimal for the work that residents perform (both in terms of what they do and how they do it or are allowed to do it).^{15,16} For example, residents may be challenged to play a role in which they are primarily responsible for a patient's asthma control because of the fragmented care that results from monthly rotational schedules.

The current scarcity of resident-sensitive quality measures (RSQMs) has led Kalet and colleagues¹⁵ and Yin and colleagues¹⁷ to develop educationally sensitive patient outcomes, which they describe as “patient outcomes that are sensitive to provider education, can be measured, and are in the pathway linking medical education interventions to patient outcomes.”¹⁷ An example educationally sensitive patient outcome is activating a patient to make and sustain changes that can positively impact his/her health.¹⁵ Educationally sensitive patient outcomes intentionally map a provider's education to his/her abilities and then to intended outcomes for patients. This effort is one of only a few in the literature to delineate quality measures that are sensitive to the work that residents perform.^{18,19} This paucity of quality measures that are either attributable or

contributed to by residents prevents linking resident performance with many quality indicators of care delivery.^{20,21}

In this study, we defined RSQMs as those measures that predictably require an action by the resident, with the resident possessing a realistic opportunity to do so, that directly impacts patient care in the clinical working and learning environment. Common examples include physician orders, information that is gathered (e.g., history and physical exam), and information that is shared (e.g., instructions given to patients). In other words, RSQMs are those measures that can be *mostly*, but not entirely, attributable to a resident. These measures, therefore, constitute a meaningful and substantial contribution by the resident to the quality and safety of care for patients and/or families.

We believe such measures must be delineated because they are critically important for assessing resident performance; for orienting residents to measures of care that can drive performance improvement; and for ensuring safe, effective, patient-centered care. While collective competence is ultimately what ensures high-quality care, focusing on individual competence is what drives personal improvement through self-regulated learning and reflective practice.^{22,23} To begin closing the gap with respect to quality measures available for use among pediatric residents, we sought to develop RSQMs for use in the pediatric emergency department (PED) setting. While this effort identified measures for our setting of interest, it may also provide a template (i.e., the steps we took in identifying and developing measures) for identifying and developing RSQMs in other settings and specialties.

Method

To develop a list of RSQMs for the PED setting, from May 2016 to December 2016, we used a model described by Haan and colleagues¹⁹ of first identifying existing national measures before developing measures on a local level. We sought measures that met two criteria: The measure had to pertain (1) to acute care illnesses that represent what emergency care is intended to treat and (2) to common illnesses to ensure the opportunity to gather sufficient data to inform resident performance.

National Quality Measures Identification

To identify national measures, we considered measures listed on the Agency for Healthcare Research and Quality National Quality Measures Clearinghouse (NQMC).²⁴ In May 2016, we filtered available NQMC measures by emergency department setting and then by pediatric patients, including measures that did not specify the age of the patient. Of the remaining measures, we excluded adult-, radiology-, and systems-focused measures and measures for

illnesses that were either not acute, not common in the PED, or both. To hone in on RSQMs, we also excluded measures focused on high-acuity illnesses and situations for which residents are unlikely to direct care.

Additional Quality Measures Identification

To develop additional measures, we decided to focus on specific illnesses, as gathering several measures for one illness would provide a performance profile for that illness and also focus data collection amid the volume of diverse encounters residents have with patients. To fit with our focus on common, acute illnesses, we concentrated on acute asthma exacerbation, bronchiolitis, and closed head injury (CHI).²⁵ Unlike more diffuse conditions such as fever, we considered these illnesses to also be discrete, facilitating measure development. Our goal was to develop 10 to 15 measures for each illness, which we felt was a pragmatic number for collecting and collating data for multiple residents and for avoiding cognitive overload of residents when considering data about their performance.

Consistent with the approach delineated by Haan and colleagues,¹⁹ we employed a local focus (i.e., institution conducting study) for developing these measures. We achieved this through a nominal group technique (NGT) followed by a Delphi process. For each of these groups, we intentionally recruited, via e-mail, individuals with experience leading quality improvement work, including defining quality metrics. Recruited individuals could only participate in a single group.

NGT

We recruited 6 pediatric emergency medicine faculty and fellows from the Division of Emergency Medicine at Cincinnati Children's Hospital Medical Center (CCHMC) for the NGT. At the session, held in September 2016, the facilitator (D.J.S.) asked the participants to consider potential quality measures for each of the three illnesses of interest that meet the following criteria: (1) important to quality care in the PED for a patient with the illness, and (2) achievement of the measure would likely primarily represent the resident's care for the patient.

Given that some measures are applicable to several illnesses, participants considered measures that applied to all three illnesses as well as ones that applied to each illness individually. Participants first recorded all of their individual ideas for potential measures. Using a round robin discussion strategy, they then shared their individual ideas with the group for each of the four categories. As ideas were offered, participants asked clarifying questions to ensure their understanding. The full list of measures from the NGT was edited for duplications (some ideas appeared on disease-specific lists as well as the all-illnesses list) and for measures that could not be extracted from the electronic medical record (e.g., timely communication with staff about care plans) to create the illness-specific lists used in the first round of the Delphi process.

Delphi Process

We recruited 18 pediatric emergency medicine faculty and fellows from the Division of Emergency Medicine at CCHMC for the Delphi process. Consistent with Delphi design,²⁶ we prioritized participants that met the following criteria in our recruitment: (1) background and experience (including research) with the illnesses of interest, (2) capable of contributing helpful input, and (3) willing to revise initial or previous judgments to reach group consensus.

Each Delphi round (of three total) was completed via e-mail, using an Excel spreadsheet (version 14.7.7, Microsoft, Inc., Redmond, Washington), within a two-week period between October and December 2016. In the first round, participants rated the measures developed through the NGT, categorized by the three illnesses of interest, on a six-point scale (where 1 = lowest and 6 = highest) in two categories: (1) how important the measure is to quality care for the illness and (2) the likelihood that the measure represents work completed by a resident. We predetermined that at least 80% of participants would need to rate measures as a 5 or 6 in both categories for automatic inclusion; for automatic exclusion, we predetermined that at least 80% of participants would need to rate the measure as a 1 or 2 in both categories. These automatic criteria were chosen because we felt they represented substantial consensus. Measures meeting inclusion or exclusion criteria in one round were not rated again in any subsequent rounds.

After round one, participants received their rating from the previous round as well as the distribution of ratings from the entire group and rated each measure again. We instructed participants to consider the group distribution in providing their updated rating; however, we also told them to maintain their convictions if, after taking the group distribution into consideration, their thoughts still differed from those of the group. Sequential rounds continued until we had a list of at least 10 to 15 candidate measures for each illness that met automatic inclusion criteria. In the final round, we asked participants to prioritize their top 10 measures and then their next 5 measures (11–15) for each illness.

This study was deemed exempt by the CCHMC institutional review board.

Results

National Quality Measures

Filtering NQMC measures by emergency department setting yielded 183 results. We initially excluded the following measures: 16 that did not sort to pediatrics or specify the age of the patient; 12 that were clearly adult focused (e.g., myocardial infarction); 14 related to comprehensive diabetic care that is not part of emergency care (e.g., neuropathy screening); 22 for stroke,

which is an uncommon diagnosis in pediatrics; and 13 specific to radiology services provided to emergency department patients (e.g., mean imaging report turnaround times).

Of the remaining 106 measures, 52 focused on systems-level indicators, with many of these representing activities that extend beyond care in the PED (e.g., proportion of emergency department visits for asthma that meet criteria for the emergency department being the appropriate level of care). Excluding these left 54 candidate measures that were either largely or partially focused on the work of individual physician providers. We next excluded 8 measures for epilepsy that focused on work completed by neurologists or on work completed beyond the emergency department encounter (e.g., magnetic resonance imaging within one year if indicated) as well as 3 measures that did not meet the criteria of belonging to both common and acute illnesses (e.g., diabetic screening in patients started on psychiatric medication). The remaining 43 measures focused on pharyngitis, asthma, procedures, CHI, febrile and nonfebrile seizures, upper respiratory infections, headaches, effective communication, sickle cell disease, flow and efficiency, sepsis, trauma, children with special health care needs, and pain control in patients with long bone fractures. These all met the criteria of common, acute illnesses in the PED.

We next considered the remaining 43 measures from the perspective of whether they were resident sensitive, which led to eliminating 15 pertaining to trauma care, sepsis management, and invasive or high-risk procedures, which are typically led by supervising physicians rather than residents because of the high stakes involved. The remaining 28 measures, which are candidates for being resident sensitive, depending on institutional context, are shown in Table 1.

Additional Quality Measures

All recruited members of the NGT and Delphi groups completed all applicable tasks.

The NGT participants produced 137 measures (46 applicable to all three illnesses, 42 applicable to asthma only, 20 bronchiolitis only, and 29 CHI only). After adding the measures for all illnesses to the illness-specific lists and editing for duplications and measures unable to be extracted from the electronic medical record, 67 measures for asthma, 46 for bronchiolitis, and 48 for CHI remained. We then provided the Delphi group with these 161 measures for the first round.

After the first Delphi round, 3 measures for asthma, 0 for bronchiolitis, and 2 for CHI met automatic inclusion criteria. Following round two, an additional 15 measures for asthma, 21 for bronchiolitis, and 19 for CHI met automatic inclusion criteria, resulting in a total of 18 measures for asthma, 21 for bronchiolitis, and 21 for CHI that met automatic inclusion criteria. No measures met automatic exclusion criteria in the first two rounds.

Table 1 National Quality Measures Clearinghouse (NQMC) Measures That Are Potentially Resident Sensitive, May 2016^a

NQMC no.	Measure summary ²⁴
NQMC:009936	"Appropriate testing for children with pharyngitis: percentage of children 2 to 18 years of age who were diagnosed with pharyngitis, dispensed an antibiotic medication, and received a group A streptococcus (strep) test for the episode."
NQMC:010509	"Appropriate testing for children with pharyngitis: percentage of children 3 to 18 years of age who were diagnosed with pharyngitis, dispensed an antibiotic, and received a group A streptococcus (strep) test for the episode."
NQMC:009937	"Appropriate treatment for children with upper respiratory infection (URI): percentage of children 3 months to 18 years of age who were given a diagnosis of URI and were not treated with an antibiotic medication."
NQMC:010546	"Appropriate treatment for children with upper respiratory infection (URI): percentage of children 3 months to 18 years of age who were given a diagnosis of URI and were not dispensed an antibiotic prescription."
NQMC:010878	"Pediatric asthma: percentage of children, ages 1 through 17 years with persistent asthma, who, during the measurement year, presented to a hospital emergency department (ED) for an asthma exacerbation and were prescribed an inhaled corticosteroid (ICS) at the time of discharge."
NQMC:009940	"Use of appropriate medications for people with asthma: percentage of patients 5 to 64 years of age during the measurement year who were identified as having persistent asthma and who were appropriately dispensed medication during the measurement year."
NQMC:008024	"Diagnosis and management of asthma: percentage of patients with an emergency department visit or inpatient admission for an asthma exacerbation who are discharged from the emergency department or inpatient setting with an asthma discharge plan."
NQMC:010513	"Medication management for people with asthma: percentage of patients 5 to 85 years of age during the measurement year who were identified as having persistent asthma and who were dispensed an asthma controller medication that they remained on for at least 50% of their treatment period."
NQMC:010707	"Overuse of imaging: ratio of the number of MRI scans to the number of CT scans obtained on or within the 30 days after the date of evaluation for atraumatic headache for children, ages 4 through 17 years old, within the measurement year."
NQMC:010705	"Overuse of imaging: percentage of children, ages 4 through 17 years old, for whom imaging of the head (CT or MRI) is obtained for the evaluation of primary headache without indications for neuroimaging, including neurologic deficit lasting longer than 60 minutes, signs and symptoms of increased intracranial pressure, or lumbar puncture."
NQMC:010879	"Overuse of imaging: the percentage of computed tomography (CT) scans obtained without indication on or within 30 days after the date of evaluation for atraumatic headache among children ages 4 through 17 years old."
NQMC:010708	"Overuse of imaging: ratio of the number of MRI scans to the number of CT scans obtained on or within 30 days after the date of evaluation for atraumatic seizure among children, ages 1 through 17 years old, during the measurement year."
NQMC:010706	"Overuse of imaging: percentage of children, ages 2 through 17 years old, with post-traumatic headache who were evaluated in the ED within 24 hours after an injury, and imaging of the head (CT or MRI) was obtained in the absence of documented neurologic signs or symptoms that suggest intracranial hemorrhage or basilar skull fracture."

NQMC:010709	“Overuse of imaging: percentage of children, ages 1 through 17 years old, for whom CT imaging of the head is obtained for the evaluation of a first generalized afebrile, atraumatic seizure without indication for CT imaging.”
NQMC:010704	“Overuse of imaging: percentage of children, ages 6 months through 4 years, diagnosed with simple febrile seizure who are evaluated with imaging of the head (CT or MRI) without indications for neuroimaging, including lumbar puncture and complex febrile seizure.”
NQMC:005609	“Language services: the percent of patient visits and admissions where preferred spoken language for health care is screened and recorded.”
NQMC:010431	“Emergency department (ED): median time from ED arrival to ED departure for discharged ED patients.”
NQMC:010432	“Emergency department (ED): median time from ED arrival to provider contact for ED patients.”
NQMC:010136	“Emergency department (ED): admit decision time to ED departure time for admitted ED patients.”
NQMC:010135	“Emergency department (ED): median time from ED arrival to ED departure for admitted ED patients.”
NQMC:010077	“Hospital inpatients’ experiences: percentage of parents who reported whether they were kept informed about their child’s care in the emergency room.”
NQMC:006162	“Children with special health care needs: percentage of children who meet criteria for having special health care needs according to the Children with Special Health Care Needs Screener (CSHCN Screener).”
NQMC:009626	“Satisfaction of care for children with sickle cell disease (SCD): percentage of parents or guardians of children younger than 18 years of age identified as having SCD who responded, on a scale of 1 to 5, ‘satisfied (4)’ or ‘very satisfied (5)’ to a survey question regarding satisfaction with the care received in the emergency department, during the measurement year.”
NQMC:009568	“Sickle cell disease (SCD): percentage of children younger than 18 years of age identified as having SCD presenting to an ED with an acute pain episode who had a pain assessment within 30 minutes following initial contact during the measurement year.”
NQMC:009572	“Sickle cell disease (SCD): percentage of children younger than 18 years of age identified as having SCD presenting to an ED with fever who received parenteral broad-spectrum antibiotic treatment within 60 minutes following initial contact during the measurement year.”
NQMC:009569	“Sickle cell disease (SCD): percentage of children younger than 18 years of age identified as having SCD presenting to an ED with severe pain who had a parenteral analgesic within 60 minutes following initial contact during the measurement year.”
NQMC:009570	“Sickle cell disease (SCD): percentage of children younger than 18 years of age identified as having SCD presenting to an ED with severe pain who had a pain assessment within 30 minutes following analgesic administration during the measurement year.”
NQMC:009573	“Sickle cell disease (SCD): percentage of children younger than 18 years of age identified as having SCD presenting to an ED with fever, who had a pulse oximetry reading, complete blood count, reticulocyte count, and blood culture within 60 minutes following initial contact during the measurement year.”

Abbreviations: MRI indicates magnetic resonance imaging; CT, computed tomography.

^aAs determined by the authors (see text for details).

During the first two rounds, Delphi participants raised the issue of whether the prevalence of the measure with the given diagnosis should weigh on their decision. The authors agreed that it would be desirable to identify measures that would be present for most encounters with the illnesses of interest. Thus, before moving to round three, we eliminated 2 measures for both asthma and bronchiolitis, leaving 16 potential final measures for asthma and 19 for bronchiolitis. We also collapsed 3 measures for CHI into 1 parent measure (presence or absence of hematoma) with 2 submeasures (location and size of hematoma) if the parent measure was met, leaving 19 potential final parent measures for CHI.

The results of round three, in which Delphi members categorized the potential final measures by their top 10 and then their next 5, are shown in Table 2.

Table 2 Prioritized Resident-Sensitive Quality Measures for Illnesses of Interest, Cincinnati Children's Hospital Medical Center, September to December 2016^a

Measures	No. of Delphi participants (out of 18) who put measure in their top 10	In the top 10 for the majority of Delphi participants?
Acute asthma exacerbation		
Documented response to intervention	17	Yes
Correct medication dose ordered for dexamethasone	16	Yes
Used asthma order set	15	Yes
Correct medication dose ordered for albuterol	14	Yes
Noted the acuity of the patient in documentation	13	Yes
Used dexamethasone as steroid	12	Yes
Documentation of disposition decision	13	Yes
Time from resident assigning him/herself to patient to placing steroid order	11	Yes
Correct medication dose ordered for ipratropium	11	Yes
Home dexamethasone instructions documented in written discharge instructions	11	Yes
Used standardized dosing for discharge medications (albuterol, dexamethasone)	10	Yes
Stated who to follow up with and included contact information in discharge papers	8	
Documentation of needing albuterol more often than every 4 hours as a reason to return in written discharge instructions	8	
Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	8	
Documentation of previous intubation or Bilevel Positive Airway Pressure for asthma	7	
Used ipratropium appropriately	6	

Measures	No. of Delphi participants (out of 18) who put measure in their top 10	In the top 10 for the majority of Delphi participants?
Bronchiolitis		
Assessment of severity documented	18	Yes
Effort of breathing documented	18	Yes
Followed bronchiolitis pathway appropriately	17	Yes
Documentation of response to specific therapeutics (i.e., how patient responded to suctioning, breathing treatment, normal saline bolus, etc.)	15	Yes
Documentation of justification for appropriate disposition (home versus admit)	14	Yes
Documentation of wheezing	11	Yes
Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	10	Yes
Documented quality of air entry (normal, decreased, etc.)	9	
Noted the acuity of the patient in documentation	8	
Used standard/premade discharge instructions for bronchiolitis	8	
Documentation of poor feeding as a reason to return in written discharge instructions	8	
Documented presence or absence of subcostal retractions	7	
Nasal bulb suction teaching for home ordered	7	
Documented presence or absence of intercostal retractions	6	
Stated who to follow up with and included contact information in discharge papers	6	
Documentation of crackles	5	
Documentation of previous wheezing	5	
Stated appropriate number of days for follow-up in discharge papers	4	
Documented presence or absence of suprasternal retractions	4	
Closed head injury (CHI)		
Used PECARN [®] /CHI pathway appropriately	18	Yes
Documentation of presence or absence of loss of consciousness	18	Yes
Documentation of presence or absence of emesis	16	Yes
Documentation of Glasgow Coma Score (GCS)	15	Yes
Documentation of whether back to baseline or not	14	Yes
Full neurological exam documented	14	Yes
Mechanism of injury documented	13	Yes
Reassessments of patient documented	11	Yes

Measures	No. of Delphi participants (out of 18) who put measure in their top 10	In the top 10 for the majority of Delphi participants?
Documentation of time of injury	11	Yes
Appropriate differential diagnosis and medical decision making documented	9	
Return to school/play recommendations in discharge papers	8	
Documentation of presence or absence of hematoma on physical exam (if present, location and size also documented)	8	
Assessment of severity documented	8	
Used standard/premade discharge instructions for diagnosis	4	
Appropriate follow-up (sports medicine, rehabilitation medicine, neurology, etc.) recommended	4	
Management of associated symptoms—headache	3	
Management of associated symptoms—vomiting	3	
Documentation of presence or absence of other injury on physical exam	2	
Documentation of presence or absence of other injury in the history	1	

^aAs prioritized by 18 Delphi participants and developed by 6 nominal group technique participants (see text for details).

^bPECARN: Pediatric Emergency Care Applied Research Network—this research network performed multi-site study on management of closed head injury in pediatrics (Representative citation: Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;374(9696):1160-1170)

Discussion

This study describes a template for identifying and developing RSQMs that may promote high-quality care delivery during and following training. An overview of this process, with considerations gleaned from our experience, is shown in Figure 1.

National Quality Measures

Out of all the NQMC measures we reviewed for our context, only a few (28/183) were potentially attributable to the work residents complete. An additional 15 measures focused on trauma care, sepsis management, and invasive or high-risk procedures, which represent areas where residents could contribute to care but which are generally led by supervising physicians. The remaining 140 measures, while potentially important in resident education and development, are either not common or not likely sensitive to resident attribution or contribution in the PED. This underscores the importance of defining RSQMs moving forward. As noted previously, we believe

such measures are critically important for assessment of resident performance; for resident orientation to measures of care that can drive performance improvement; and for ensuring safe, effective, patient-centered care. Ideally, national quality efforts will focus on identifying these measures. However, for the foreseeable future, it is likely prudent for the medical education community to drive this work.

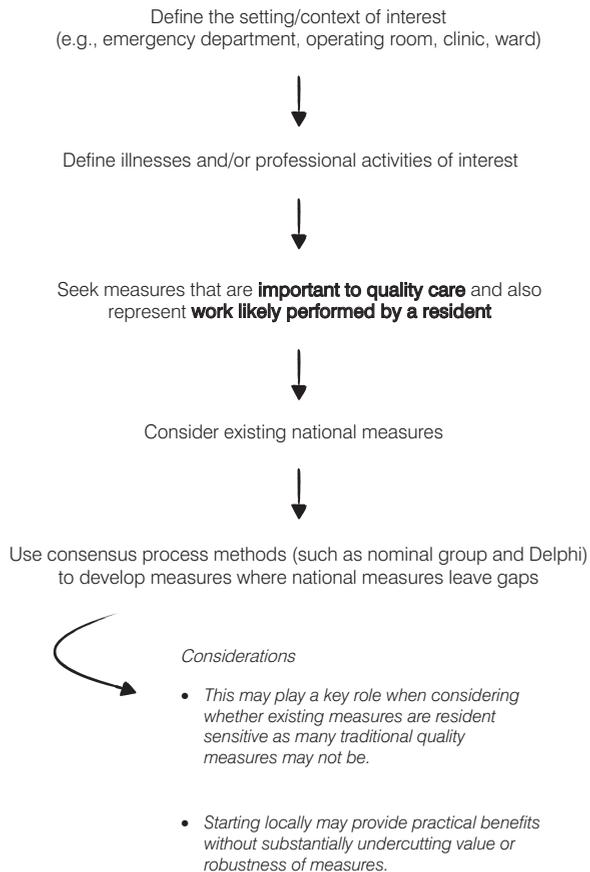


Figure 1: Process for developing resident-sensitive quality measures (RSQMs).

With the 28 NQMC measures we categorized as being potentially resident sensitive, we erred on the side of being more inclusive. Undoubtedly, individual programs and institutions will, and should, choose measures more adaptable to the work that residents perform in their local settings.

Local Quality Measures

As noted above, in developing a list of RSQMs, we used the approach espoused by Haan and colleagues,¹⁹ performing a search for available national measures before turning to a more local focus. At the outset of this study, we believed that our local efforts to identify additional measures would likely generate measures that are important for the three illnesses we were interested in but also measures that represent resident work in the *local* context of our study. An example of the latter is the use of decadron rather than prednisone as the preferred steroid for acute asthma exacerbation, a transition CCHMC initiated just prior to this study. However, most of the measures identified by the NGT and Delphi groups could apply in many local contexts, suggesting that the process of prioritizing measures for *common, acute* illnesses led to the identification of measures that are so germane that universal presentations and management approaches for these illnesses may exist across myriad institutions. The potential for locally developed and prioritized measures to apply beyond the local setting suggests that the local context may be a good starting point for developing RSQMs when reviews of national measures yield gaps. Starting locally may achieve two important goals: (1) the practical and achievable identification of RSQMs where national efforts may be initially prohibitive because of the time and resources required and (2) the development of an available pool of RSQMs that other institutions can adapt while such measures are vetted nationally. Given the current paucity of such pools, even locally developed measures could make practical contributions to efforts in this area. Several measures were developed by the NGT group but not prioritized by the subsequent Delphi group. There is risk in excluding these entirely or labeling them as not being RSQMs. Given the charge to the NGT group, all measures put forth and considered in this part of the study were, by definition, intended to be resident sensitive. It is then important to recall that the intent of our efforts was to prioritize measures that, in part, are *mostly*, but not entirely, attributable to residents. Thus, many of the excluded measures may also be reasonably attributable to residents but were just not thought to be as attributable to residents or as highly prioritized by our participants. Looking beyond this work, it is also important to note that *attribution*, or the likelihood that a resident performs the activity represented by the measure, is not the only marker of value for quality measures that reflect the work of residents. Indeed, recent discussions in the literature have placed more focus on *contribution*, or the degree to which residents have an impact on quality care, than on attribution.^{20,21} The measures that were not prioritized by the Delphi group as being primarily attributable to residents in this study may fall into the category of representing work for which residents contribute toward quality patient care.

Many of the measures developed in this study focus on documentation in notes or discharge instructions for the family. While residents could view these measures as simply focusing on proper documentation rather than proper care, we believe that documentation is a necessary indicator of care provided. Additionally, quality measures that are part of residents' everyday work may also address some of the barriers that have been described in implementing quality

improvement initiatives with residents, such as the perception of the limited value of their work in such initiatives and the lack of understanding about the basic aspects of quality improvement.²⁷ Finally, such measures are likely to provide clear goals for residents' work, which has been found to be a facilitator of quality improvement efforts that are successful with residents.²⁷

National Versus Local Measures

Both national and local measures have their strengths and drawbacks. The national measures have been vetted by national organizations focused on care quality. However, they also tended to be diffuse in topic with many measures that were less likely to be routinely encountered in the daily practice of a resident. This led to us excluding many national measures; our resulting list of potentially resident-sensitive measures was less than one-fifth the size of the starting list. In contrast, we were able to focus the locally developed measures on a few specific illnesses, which allowed for the delineation of multiple measures to form a composite for these illnesses. We believe this is a strength of these measures because it positions them to give a more complete view of a resident's performance in managing patients with these conditions. However, these measures do not have the national vetting of the NQMC national measures. Overall, though, we believe the local measures offer more promise for use with residents as they are clearly focused on the work that residents perform, relevant to resident learning and improvement efforts, and provide multiple measures for the illnesses considered, generating a more complete representation of resident performance in these areas. That said, these measures need to be more broadly studied to establish validity evidence for them.

Limitations

This study has limitations to consider. First, nonemergency and/or nonpediatric medicine NQMC measures may better reflect resident work, limiting the generalization of our findings. Future work should evaluate the remaining NQMC measures for their resident sensitivity. A second limitation is that we developed additional measures through NGT and Delphi groups at a single institution. Vetting these measures through national efforts is an important next step. Third, while developed through group consensus processes, the measures we identified for the three illnesses of interest still need to be implemented and validated. Additional next steps should include implementation and seeking evidence of implementation and validation, steps which plague many current quality measures.²⁸ Fourth, the measures identified by the NGT and Delphi groups were focused on process measures rather than outcome measures; however, "reliance on clinical process measures ... [with] substantial challenges of moving to outcome measures" is common in current quality improvement efforts.²⁸ Finally, only pediatric emergency medicine physicians were engaged in the development of these measures. Future work will need to engage other key stakeholders, including residents, nurses, patients, and families.

Despite these limitations, this study avoids some common pitfalls of NGT and Delphi studies described in the medical education literature, such as not performing a literature review before the consensus-building activity, not providing consensus group participant response rates, and not defining consensus a priori.²⁹

Future Directions

The steps delineated in this paper may provide a template for the future development and prioritization of measures focused on the work of residents, as well as measures focused on the work of medical students and fellows, across a variety of specialties and settings. Once developed and prioritized, such measures should be used and studied as a source of performance data for work-based assessment and improvement efforts.

Conclusions

RSQMs may provide an important metric for driving needed improvements in the quality of care provided by residents both during and following training. Such measures may also lay the groundwork for the goal of linking education outcomes and patient care outcomes. While this goal may feel distant currently, RSQMs may provide a practical strategy for creating an intermediary link in the chain that ties education outcomes to patient care outcomes.¹⁶

References

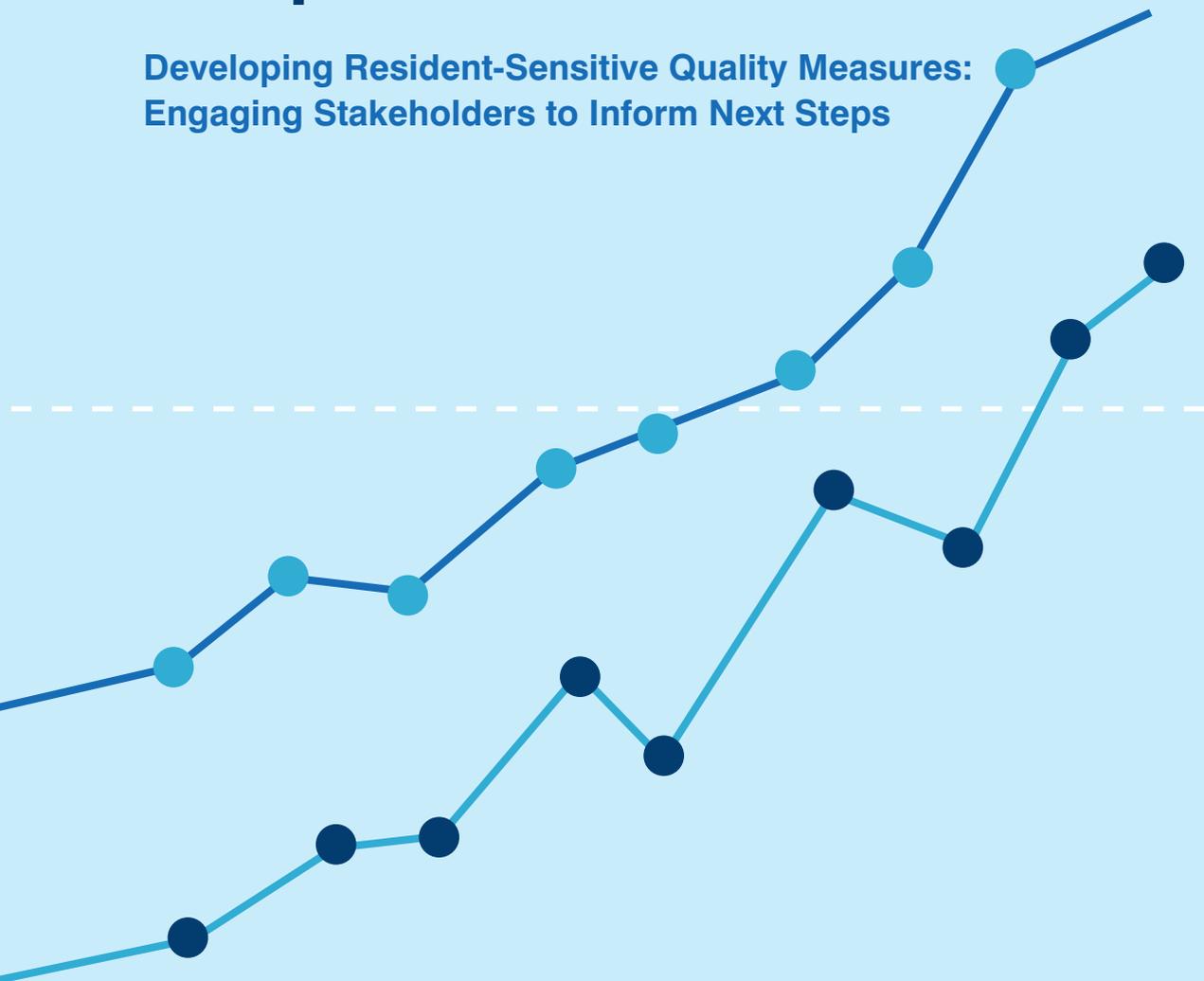
1. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA*. 2009;302:1277–1283.
2. Bansal N, Simmons KD, Epstein AJ, Morris JB, Kelz RR. Using patient outcomes to evaluate general surgery residency program performance. *JAMA Surg*. 2016;151: 111–119.
3. 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;312:2385–2393.
4. 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.
5. Kauvar DS, Braswell A, Brown BD, Harnisch M. Influence of resident and attending surgeon seniority on operative performance in laparoscopic cholecystectomy. *J Surg Res*. 2006;132:159–163.
6. Englesbe MJ, Pelletier SJ, Magee JC, et al. Seasonal variation in surgical outcomes as measured by the American College of Surgeons–National Surgical Quality Improvement Program (ACS-NSQIP). *Ann Surg*. 2007;246:456–462.
7. Phillips DP, Barker GE. A July spike in fatal medication errors: A possible effect of new medical residents. *J Gen Intern Med*. 2010;25:774–779.
8. Resnick AS, Disbot M, Wurster A, Mullen JL, Kaiser LR, Morris JB. Contributions of surgical residents to patient satisfaction: Impact of residents beyond clinical care. *J Surg Educ*. 2008;65:243–252.
9. Inaba K, Recinos G, Teixeira PG, et al. Complications and death at the start of the new academic year: Is there a July phenomenon? *J Trauma*. 2010;68:19–22.
10. Medicare Payment Advisory Commission. Graduate medical education financing: Focusing on educational priorities. In: Report to Congress: Aligning Incentives in Medicare. Washington, DC: Medicare Payment Advisory Commission; 2010:103–126.
11. Cordasco KM, Horta M, Lurie N, Bird CE, Wynn BO. How Are Residency Programs Preparing Our 21st Century Internists? A Review of Internal Medicine Residency Programs' Teaching on Selected Topics. Arlington, VA: RAND Corporation; 2009. http://www.rand.org/content/dam/rand/pubs/working_papers/2009/RAND_WR686.pdf. Accessed November 14, 2017.
12. Medicare Payment Advisory Commission. Report to the Congress: Improving Incentives in the Medicare Program. Washington, DC: Medicare Payment Advisory Commission; 2009. http://www.medpac.gov/docs/default-source/reports/Jun09_EntireReport.pdf. Accessed November 14, 2017.
13. Wiest FC, Ferris TG, Gokhale M, Campbell EG, Weissman JS, Blumenthal D. Preparedness of internal medicine and family practice residents for treating common conditions. *JAMA*. 2002;288:2609–2614.
14. Committee on the Governance and Financing of Graduate Medical Education. Graduate Medical Education That Meets the Nation's Health Needs. Washington, DC: National Academies Press; 2014.
15. 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;85:844–851.

16. Cook DA, West CP. Perspective: Reconsidering the focus on “outcomes research” in medical education: A cautionary note. *Acad Med.* 2013;88:162–167.
17. Yin HS, Jay M, Maness L, Zabar S, Kalet A. Health literacy: An educationally sensitive patient outcome. *J Gen Intern Med.* 2015;30:1363–1368.
18. Swing SR, Schneider S, Bizovi K, et al. Using patient care quality measures to assess educational outcomes. *Acad Emerg Med.* 2007;14:463–473.
19. Haan CK, Edwards FH, Poole B, Godley M, Genuardi FJ, Zenni EA. A model to begin to use clinical outcomes in medical education. *Acad Med.* 2008;83:574–580.
20. Moreau KA, Eady K. Connecting medical education to patient outcomes: The promise of contribution analysis. *Med Teach.* 2015;37:1060–1062.
21. Van Melle E, Gruppen L, Holmboe ES, Flynn L, Oandasan I, Frank JR; International Competency-Based Medical Education Collaborators. Using contribution analysis to evaluate competency-based medical education programs: It’s all about rigor in thinking. *Acad Med.* 2017;92:752–758.
22. Sandars J, Cleary TJ. Self-regulation theory: Applications to medical education: AMEE guide no. 58. *Med Teach.* 2011;33:875–886.
23. Schön DA. *The Reflective Practitioner: How Professionals Think in Action.* New York, NY: Basic Books, Inc.; 1983.
24. Agency for Healthcare Research and Quality. Search results for “Emergency Department.” https://www.qualitymeasures.ahrq.gov/search?f_Measurement_Setting=Emergency+Department&fLockTerm=Emergency+Department. Search performed February 1, 2016. Accessed November 14, 2017.
25. Mittiga MR, Schwartz HP, Iyer SB, Gonzalez Del Rey JA. Pediatric emergency medicine residency experience: Requirements versus reality. *J Grad Med Educ.* 2010;2:571–576.
26. Hsu CC, Sandford BA. The Delphi technique: Making sense of consensus. *Pract Assess Res Eval.* 2007;12(10). <http://pareonline.net/getvn.asp?v=12&n=10>. Accessed December 29, 2017.
27. Butler JM, Anderson KA, Supiano MA, Weir CR. “It feels like a lot of extra work”: Resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med.* 2017;92:984–990.
28. Berenson RA, Pronovost PJ, Krumholz HM. *Achieving the Potential of Health Care Performance Measures: Timely Analysis of Immediate Health Policy Issues.* Princeton, NJ: Robert Wood Johnson Foundation. <https://www.rwjf.org/en/library/research/2013/05/achieving-the-potential-of-health-care-performance-measures.html>. Published 2013. Accessed August 23, 2016.
29. Humphrey-Murto S, Varpio L, Wood TJ, et al. The use of the Delphi and other consensus group methods in medical education research: A review. *Acad Med.* 2017;92:1491–1498.



Chapter 5

Developing Resident-Sensitive Quality Measures: Engaging Stakeholders to Inform Next Steps



Schumacher DJ, Martini A, Holmboe E, Varadarajan K, Busari J, van der Vleuten CPM, Carraccio C. "Developing Resident-Sensitive Quality Measures: Engaging Stakeholders to Inform Next Steps." *Academic Pediatrics* 2019;19(2):177-185.

Abstract

Objective

Despite the need for quality measures relevant to the work residents complete, few attempts have been made to address this gap. Resident-sensitive quality measures (RSQMs) can help fill this void. This study engaged resident and supervisor stakeholders to develop and inform next steps in creating such measures.

Methods

Two separate nominal group techniques (NGTs), one with residents and one with faculty and fellow supervisors, were used to generate RSQMs for 3 specific illnesses (asthma, bronchiolitis, and closed head injury) as well as general care for the pediatric emergency department. Two separate Delphi processes were then used to prioritize identified RSQMs. The measures produced by each group were compared side by side, illuminating similarities and differences that were explored through focus groups with residents and supervisors. These focus groups also probed future settings in which to develop RSQMs.

Results

In the NGT and Delphi groups, residents and supervisors placed considerable focus on measures in 3 areas across the illnesses of interest: 1) appropriate medication dosing, 2) documentation, and 3) information provided at patient discharge. Focus groups highlighted hospital medicine and general pediatrics as priority areas for developing future RSQMs but also noted contextual variables that influence the application of similar measures in different settings. Residents and supervisors had both similar as well as unique insights into developing RSQMs.

Conclusion

This study continues to pave the path forward in developing future RSQMs by exploring specific settings, measures, and stakeholders to consider when undertaking this work.

The impact of trainees' care on patient care outcomes is not well studied¹ despite data showing an enduring correlation between the quality of care delivered during training and that provided in future practice.²⁻⁶ Investigations also demonstrate that many residency graduates are unprepared to meet the needs of patient populations, even for foundational tasks.⁷⁻¹⁰ Given the importance of residency training on current and future patient care, performance feedback using quality measures should be guaranteed during these formative years. Although some quality measures used for physicians in practice are appropriate for trainees as well, many traditional quality measures are poorly aligned with resident contributions to care delivery.^{11,12}

Resident-Sensitive Quality Measures: Meeting a Need

Despite the need for quality measures relevant to the work that residents complete, little has been done to address this need,^{11,13-15} limiting efforts to link resident performance with quality care delivery.^{16,17} This gap led us to develop "resident-sensitive quality measures" (RSQMs), which we define as "those measures that predictably require an action by the resident, with the resident possessing a realistic opportunity to do so, that directly impacts patient care in the clinical working and learning environment."¹⁸ The most effective process for developing RSQMs is not known. Therefore, determining a process for and developing additional RSQMs is important, as such measures are 1) fundamental to assessing resident performance and providing formative feedback that fosters reflection on practice, 2) needed for orienting residents to care markers that can drive improvement in both training and practice, and 3) helpful in defining safe, effective, patient-centered care during training and beyond.

We previously described a method for developing RSQMs among faculty and fellow supervisors for use in the pediatric emergency department (PED) setting through consensus group methodologies for 3 illnesses (asthma, bronchiolitis, and closed head injury [CHI]) commonly seen in the PED.¹⁸ In this study, we replicate this process of developing RSQMs with residents and then explore similarities and differences in measures developed by residents and those developed by supervisors. Furthermore, we explore future settings in which RSQMs could be potentially useful, offering insights into best practices and next steps for developing RSQMs.

Methods

Setting and Design

This mixed-methods study used an explanatory sequential design involving 1) quantitative data collection and analysis, 2) qualitative data collection and analysis, and then 3) data interpretation.¹⁹ In the first study component, we sought to define and prioritize RSQMs for the PED by using the nominal group technique (NGT) and Delphi process with 2 distinct groups: residents and faculty/fellow supervisors. As noted, our work with supervisors has been described previously, with top-rated measures reported.¹⁸ The current study reports these measures in a different format with additional measures for the purposes of drawing comparisons between the 2 stakeholder groups. We followed these initial efforts with a qualitative study (focus groups with residents and supervisors) to explore differences between RSQMs resulting from the NGT and Delphi groups as well as future settings for RSQM development.

Participants for all study components were recruited via e-mail from residents at Cincinnati Children's Hospital Medical Center (CCHMC) as well as faculty and fellows from the divisions of emergency medicine (EM), hospital medicine (HM), and outpatient general pediatrics (GP) at CCHMC. All categorical pediatrics residents were considered eligible for all resident study components, but residents needed to have rotated in the PED to be included in the NGT or Delphi groups. Fellows and faculty in EM were eligible for the supervisor NGT and Delphi groups. Finally, fellows and faculty in EM, HM, and GP (described as best additional specialties to develop RSQMs during resident focus group) were eligible for the supervisor focus group.

Data Collection: NGT and Delphi

Based on prevalence data and the desire to develop RSQMs for common, acute problems in the PED, we focused on acute asthma exacerbation, bronchiolitis, and CHI.²⁰ We followed the advice of Haan and colleagues¹⁴ to develop measures on a local level when existing national measures are insufficient. To accomplish this goal, we developed RSQMs locally using an NGT to generate measures, followed by a Delphi process to prioritize them.^{18,21} We completed this process with 2 groups separately: 1) pediatric residents and 2) faculty and fellow supervisors. By separating measure development and prioritization by group, we were able to explore similarities and differences between measures identified by each group. We recruited participants for the NGT and Delphi groups using guidelines from the literature.^{20,22}

Nominal Group Technique

Eight pediatric residents participated in the June 2017 session. Six pediatric emergency medicine faculty and fellows participated in the September 2016 session. At each session, participants were asked to generate potential quality measures for general care in the PED as well as for the 3 illnesses of interest. They were asked to consider measures that meet the following criteria: 1)

important to quality care for a patient with the illness presenting in the PED, and 2) achievement of the measure likely represents work completed by the primary resident caring for the patient.

Participants individually recorded all ideas for potential measures in the 4 categories. After exhausting their lists, ideas were shared with the group in a round-robin format. NGT measure lists were edited for duplications and measures that could not be extracted from the electronic medical record (e.g., verbal communication between a resident and nurse).

Delphi

Most general PED care measures developed in the NGT groups applied to all 3 of the illnesses of interest. Thus, the relevant general care measures were added to the 3 specific illness measures to comprise the initial set of measures for rating in each Delphi process. A few additional measures also were added by the research team for round 1 of the supervisor Delphi process to expand on ideas generated from the NGT with this group. We believed these additions were acceptable because it is expected that researchers will need to clean and adapt items at the beginning of a Delphi group.²³

We recruited 16 pediatric residents and 18 EM faculty and fellows for the resident and supervisor Delphi, respectively. For the resident Delphi, each round was completed within 2 weeks between June and July 2017. For the supervisor Delphi, each round was completed within 2 weeks between October and December 2016.

In the first round of each Delphi, participants were provided with measures developed by their respective NGT group. They were asked to rate each measure on a 6-point scale, with 1 being the lowest rank and 6 the highest, in 2 categories: 1) importance of the measure to quality care for the specific illness, and 2) likelihood the measure represents work completed by residents. Participants could propose additional measures for rating if they desired. We defined consensus as a minimum of 80% of participants rating measures as a 5 or 6 in both categories for automatic inclusion into our final set. Measures that achieved automatic inclusion after each round were removed from rating in the subsequent round.

After round 1, participants were presented with their previous round ratings as well as the distribution of ratings from the entire group to inform the next round of rating. Sequential rounds continued for each Delphi process until lists of prospective measures met or exceeded our predetermined 10 to 15 measures per illness. In the final round, participants in each Delphi group were asked to prioritize their top 10 measures and next 5 measures (11–15) for each illness.

Data Analysis: NGT and Delphi

NGT and highly rated Delphi measures were compared side by side to determine differences and similarities in measures developed between the groups. In addition, final round Delphi measures were assigned a weighted score by allocating 2 points for each “top 10” rating and 1 point for each “next 5 (11–15)” rating. Thus, a measure could have a maximum score of 36 for the faculty and fellow Delphi (18 participants x 2 points) and 32 for the resident Delphi (16 participants x 2 points). Calculating this score allowed us to account differently for “top 10” and “next 5” ratings in the final Delphi round to produce final rankings of measures.

Data Collection: Focus Groups

Subsequent focus groups explored differences between RSQMs developed by residents and supervisors as well as future settings in which to consider developing RSQMs (see Supplemental Appendix 1 for interview guide). Participants were chosen for each focus group based on consensus of availability of potential participants who completed a poll after receiving the recruitment e-mail. The resident focus group included 8 residents and was held in July 2018. The supervisor focus group was also held in July 2018 and included 10 faculty and fellows (3 EM, 3 GP, and 4 HM). Focus groups were audiorecorded and professionally transcribed.

Data Analysis: Focus Groups

Focus group transcripts were analyzed using conventional content analysis, a method that derives codes from data and defines these codes during the analysis.²² Primary analysis was performed independently by 2 members of the team (D.J.S. and A.M.), who began by reading and rereading the data to gain immersion and then assigning codes. They used Dedoose (Version 8.0.35, SocioCultural Research Consultants, LLC, Los Angeles, Calif.) to help facilitate coding. After independent coding of both focus group transcripts, they met to review codes and agree on a single set of codes. Working together, they grouped codes into categories that linked codes in meaningful ways followed by clusters that provided insights into 1) why differences exist between residents and supervisors in their development of RSQMs and 2) what future settings should be considered for RSQM development. Secondary analysis was performed by another member of the research team (E.H.), who was provided focus group transcripts, primary analysts' codebook, and the primary analysts' categories and clusters. This individual reviewed the transcripts to determine whether the primary analysts' coding, categorizing, and clustering accurately defined results from the data and did not include codes that were not clearly supported by the data. Primary and secondary coders collectively developed and agreed on final themes and results. NGT and focus group participants received \$100 for their time. Delphi participants received \$150, given the need for prolonged engagement and work. This study was determined to be exempt by the CCHMC Institutional Review Board.

Results

All participants in both NGT and Delphi groups completed all tasks for their group.

Nominal Group Technique

The resident NGT produced 150 measures: 57 general for PED, 33 asthma, 28 bronchiolitis, and 32 CHI. In the faculty and fellow NGT, participants generated 115 potential measures: 30 general for PED, 41 asthma, 16 bronchiolitis, and 28 CHI. As indicated in Table 1,²⁴ measures commonly focused on appropriate medication use, documentation, and information provided at discharge.

The NGT groups had 5 general PED measures in common (Table 1). In addition, the supervisor NGT participants had 2 measures focused on 48-hour returns that matched one of the resident measures (ED return/“bounce back” rate). Furthermore, the NGT groups had 4 additional measures that were very similar to one another for general PED measures. Similarities between groups continued for all 3 specific illnesses, as shown in Table 1.

Delphi

The resident Delphi process began with 74 measures for asthma, 68 for bronchiolitis, and 79 for CHI. Table 2²⁴ shows the top 10 measures based on the calculated score for each illness category. The supervisor Delphi process began with 67 measures for asthma, 46 for bronchiolitis, and 48 for CHI. The top 10 measures, based on the calculated score for each illness category, are shown in Table 3.²⁴ As indicated in Tables 2 and 3, measures focused on appropriate medication use, documentation, and information provided at discharge were commonly prioritized among both residents and supervisors.

Focus Groups

Future Settings and Focus for RSQMs

Residents noted that beyond the PED, HM and GP are the best settings in which to consider developing RSQMs. Focus group participants also noted considering these measures in the pediatric intensive care unit, neonatal intensive care unit, newborn nursery, and some inpatient subspecialty services. The commonly noted PED measures of appropriate medication use, documentation, and information provided at discharge were considered to be important in HM and GP as well.

In considering RSQMs beyond the PED, focus group participants felt the relative appropriateness of similar RSQMs varied based on setting. The most striking setting-related differences focused on resident autonomy and team-, compared with resident-, driven care. Residents felt they have high levels of autonomy in both the PED and HM. This autonomy was especially felt when making initial diagnoses and executing initial management plans for patients admitted overnight

Table 1: Similar Measures Produced From Faculty/Fellow and Resident Nominal Group Techniques

Faculty/Fellows	Residents
General Pediatric Emergency Department Care	
Correct medication dose ordered*	Appropriate medication dosing*
Use interpreter if needed	Document use of interpreter if relevant
Pertinent history documented ^	Relevant history of present illness and past medical history documented in note ^
48-hour return to ED for same illness	ED return/"bounce back" rate
48-hour return to ED with same illness with admit	
Orders correct (labs, imaging)	Accuracy of labs/imaging ordered
Finish chart in timely fashion ^	Completion of all notes by end of shift ^
Reassessments of patient documented ^	If time in ED is greater than 4 hours after being assigned to resident, reassessment of patient is documented ^
Documented review of symptoms appropriate for billing level ^	Documentation supports disposition and billing level (e.g., 10-point review of systems documented for admitted patients, number of physical exam components documented for level of visit complexity) ^
Call placed to referring MD during/following encounter	Documentation by resident when resident spoke with PCP ^
Acute Asthma Exacerbation	
Use PRAM appropriately (i.e., initial orders match initial PRAM score)	Resident-assigned PRAM score matches resident-placed orders
Documentation of number of exacerbations per year ^	Documented asthma exacerbation history ^
Flu shot offered	Documentation of flu shot being offered if not up-to-date ^
Documented assessment of work of breathing ^	Documentation of work of breathing ^
Documentation of asthma triggers for patient ^	Documentation of presence or absence of potential triggers in housing environment (e.g., mold) ^
Document response to treatment ^	Reassessment documented in response to treatment ^
Documented justification of imaging use/non-use ^	If chest x-ray ordered, justification is documented in note ^
Avoid unnecessary chest x-ray	
Documented follow-up plan in discharge papers (e.g., PCP in "X" number of days) [§]	Ensure primary care provider follow-up information documented in discharge papers [§]
Documentation of previous steroid use for asthma exacerbation ^	Documentation of last steroid course (or no previous steroid courses) ^
Documentation of relevant social history ^ (smokers in home, pets)	Documented tobacco exposure ^
Documentation of spacer/metered dose inhaler for home (either have it or order it) [§]	Give albuterol prescription and make sure enough albuterol for home & school [§]

Bronchiolitis

Documentation of assessment of hydration status ^	Hydration status clearly documented in note ^
Bulb suction teaching ordered	If disposition set to home, suction teaching for home ordered [§]
Documentation of previous hospitalization for bronchiolitis ^	Document presence or absence of previous admission for bronchiolitis ^
Time to nasal suction order	Time from resident assigning self to patient to placing nasal suction order
Frequent reassessments documented ^	Minimum of 1 reassessment exam documented in note ^
No chest x-ray ordered	If chest x-ray ordered, justification is documented ^

Closed Head Injury

Documentation of presence or absence of loss of consciousness ^	Presence or absence of loss of consciousness documented ^
Appropriate PECARN/closed head injury pathway use	Does PECARN best practice alert recommendation match what was done for the patient
Documentation of presence or absence of emesis ^	Presence or absence of vomiting documented ^
FULL neurologic exam documented ^	Full neurologic exam documented ^
Mechanism of injury documented ^	Mechanism of injury documented ^
Return to school/play recommendations in discharge papers [§]	Guidelines given for return to play [§]
Appropriate follow-up (sports medicine, rehabilitation medicine, neurology, etc.) recommended [§]	Discharge papers document who to follow up with (primary care provider, concussion clinic, etc.) [§]
Follow-up with head injury clinic if concussion diagnosis given [§]	Follow-up in concussion clinic recommended at discharge if appropriate for patient [§]
School/sport note at discharge about restrictions [§]	If diagnosed with concussion, school note given for rest and activity restrictions [§]

ED indicates emergency department; PCP, primary care physician; PECARN, Pediatric Emergency Care Applied Research Network; and PRAM, Pediatric Respiratory Assessment Measure.

* Appropriate medication measure

^ Documentation measure

§ Information provided at discharge measure

PECARN performed multi-site study on management of closed head injury in pediatrics (Reference: Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;374(9696):1160-1170.)

Table 2: Top Measures From Resident Delphi Processes for Acute Asthma Exacerbation, Bronchiolitis, and Closed Head Injury

Measure	Delphi Score (out of 32)
Acute Asthma Exacerbation	
Appropriate medication dosing*	32
Documentation of work of breathing ^	31
Documentation of aeration/air exchange ^	31
Give albuterol prescription and make sure enough albuterol for home & school [§]	30
Presence or absence of wheezing documented ^	30
Ensuring patient sent home with steroid [§]	29
Reassessment documented in response to treatment ^	27
Resident-assigned PRAM score matches resident-placed orders ^	26
Appropriate discharge instructions [§]	26
Steroid ordered at same time or before albuterol/ipratropium ordered	25
Bronchiolitis	
Severity of respiratory distress documented ^	30
Oxygen saturation clearly documented in note ^	27
Appropriate discharge instructions [§]	25
Hydration status clearly documented in note ^	24
Reassessment documented after treatment/intervention ^	24
Diagnosis clearly explained to parents	24
Day of illness clearly documented ^	22
Oral feeding tolerance clearly documented in note ^	22
Discharge papers document specific return to ED instructions (e.g., work of breathing) [§]	20
Document birth history (preemie or not a preemie) ^	20
Closed Head Injury	
Full neurologic exam documented ^	28
Thorough head exam (head, eyes, skull) documented ^	28
Time from arrival to appropriate intervention if increased intracranial pressure suspected	27
Presence or absence of altered mental status documented ^	27
Presence or absence of loss of consciousness documented ^	26
Presence or absence of vomiting documented ^	24
Mechanism of injury documented ^	22
Does PECARN best practice alert recommendation match what was done for the patient	16
Guidelines given for return to play [§]	13

PRAM indicates Pediatric Respiratory Assessment Measure; ED, emergency department; and PECARN, Pediatric Emergency Care Applied Research Network.

* Appropriate medication measure

^ Documentation measure

[§] Information provided at discharge measure

PECARN performed a multisite study on management of closed head injury in pediatrics (Reference: Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;374(9696):1160-1170)

Table 3: Top Measures From Faculty/Fellow Delphi Processes for Acute Asthma Exacerbation, Bronchiolitis, and Closed Head Injury

Measure	Delphi Score (out of 36)
Acute Asthma Exacerbation	
Document response to intervention ^	35
Correct medication dose ordered for dexamethasone*	34
Use asthma order set*	33
Correct medication dose ordered for albuterol*	32
Documentation of disposition decision ^	31
Note the acuity of the patient in documentation ^	30
Correct medication dose ordered for ipratropium*	28
Use of dexamethasone as steroid*	28
Home dexamethasone instructions documented in discharge papers [§]	28
Use of standardized dosing for discharge medications (albuterol, dexamethasone)*	28
Bronchiolitis	
Assessment of severity documented ^	36
Effort of breathing documented ^	36
Effort of breathing documented ^	36
Follow pathway appropriately	35
Documentation of response to specific therapeutics (i.e., how they responded to suctioning, how they responded to breathing treatment, how they responded to normal saline bolus, etc.) ^	32
Documentation of justification for appropriate disposition (home vs admit) ^	31
Documentation of wheezing ^	28
Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions ^	25
Documented quality of air entry (normal, decreased, etc.) ^	24
Documentation of poor feeding as a reason to return in written discharge instructions [§]	22
Closed Head Injury	
Documentation of presence or absence of loss of consciousness ^	36
Appropriate PECARN/closed head injury pathway use	36
Documentation of presence or absence of emesis ^	34
Mechanism of injury documented ^	30
FULL neurologic exam documented ^	31
Documentation of GCS ^	31
Documentation of whether back to baseline or not ^	31
Reassessments of patient documented ^	28
Documentation of time of injury ^	27
Return to school/play recommendations in discharge papers [§]	22

PECARN indicates Pediatric Emergency Care Applied Research Network; and GCS, Glasgow Coma Scale.

* Appropriate medication measure

^ Documentation measure

[§] Information provided at discharge measure

PECARN performed a multisite study on management of closed head injury in pediatrics (Reference: Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;374(9696):1160-1170)

as well as completing care they felt did not need to wait until rounds. Like residents, supervisors also mentioned work done by residents overnight, focusing on the importance of residents discriminating necessary/unnecessary laboratory tests. In GP, there was concern that the team structure of resident continuity clinics served as a barrier to attributing care to an individual resident. For example, nurses may have immunizations ready before a resident orders them and common screenings (e.g., developmental, autism) may be done by a nurse or a resident. Participants felt the acute care delivered in continuity clinics, however, is likely more attributable to a given resident.

Although the common PED measures such as medications, documentation, and discharge instructions were considered to be important in other settings, focus group participants raised concerns about the following exceptions: 1) appropriate medications may not be attributable to residents in HM, 2) since there is less documentation overall in GP, there is less available documentation to attribute to residents, 3) for GP, the established standards for common care provided the opportunity to measure residents' adherence to these standards, and 4) for both HM and GP, use of standard (i.e., pre-made) discharge instructions may lessen additional discharge instructions written by residents. However, for GP, participants thought it would be important for residents to add additional free-text instructions. Several potential foci for RSQMs in other settings were suggested, as detailed in Table 4. A number of these addressed care coordination, communication, and transitions of care.

Table 4: Potential Resident-Sensitive Quality Measures for Other Settings Noted in Focus Groups

Additional Setting	Potential RSQMs
General Pediatrics	Diagnosis entered by resident Preventative care at health supervision visits (e.g., immunizations given, CBC, lead level, social risk screening such as food insecurity, maternal depression, developmental screening, autism screening) Refer patients with developmental delays Medication reconciliation Efficiency/patients seen per clinic Chronic illness management (e.g., asthma action plan) Documenting counseling on vaccines and development Add free-text instructions to discharge papers Provide standard age-based discharge instructions for health supervision visits
NICU	Communication with primary care physician Handover to the PCP at discharge

PICU	<p>PICU component of hospital stay documented in discharge summary as a living document</p> <p>Non-emergent orders (but these are checked closely, so may or may not represent resident action)</p> <p>Handover to floor team</p> <p>Call primary care physician upon admission</p> <p>Effective handovers to hospital medicine team</p>
Newborn Nursery	<p>Newborn checklist items</p> <p>Order hepatitis B immunization</p> <p>Order hearing screen</p>
Hospital Medicine	<p>Effective interdisciplinary communication captured on secure messaging and text paging services</p> <p>Call consult before talking with attending if appropriate</p> <p>Effective handovers</p> <p>Keep primary care physician updated during hospitalization</p> <p>Call primary care doctor at discharge</p> <p>Discharge planning/have discharge prepped</p> <p>Effective communication with consultants to ascertain the input needed</p> <p>Translating consultant recommendations into care</p> <p>Calling consults before talking with the attending</p> <p>Communicating plans of care with the interdisciplinary team members</p> <p>Thorough handovers to other team members</p> <p>Handover to the PCP at discharge</p> <p>Medication reconciliation</p> <p>Concise and effective discharge summary</p> <p>Document contingency planning</p>
GI	<p>Orders</p> <p>Documentation</p>
Cardiology	<p>Orders</p> <p>Documentation</p>
Hematology/Oncology	<p>Orders</p> <p>Documentation</p>
Most Specialties	<p>Ensuring that families are kept updated and educated</p>

RSQM indicates resident-sensitive quality measures; CBC, complete blood count; PCP, primary care physician; NICU, neonatal intensive care unit; PICU, pediatric intensive care unit; and GI, gastrointestinal.

Differences in Proposed Measures Between Groups

The resident NGT produced more measures than the supervisor NGT. Both focus groups felt this may be because residents work on the front line and know more details of care provided there. As one resident noted, “we are the residents so we know,” with other residents quickly offering affirmation.

Elaborating this further, both groups felt that residents tend to have detailed lists of care activities to complete, whereas supervisors focus more on the “big picture” care plan. Part of this was felt to be due to differences in the roles. However, supervisors also noted that residents may not be able to prioritize well or know resources available to help complete their work, and residents noted that supervisors may not know details that go into actually providing frontline care.

They are the ones who are actually doing this stuff. . .and I think for many of us we don't have to do a lot of those steps. (HM faculty)

Supervisors “don't know the system. . .to make sure the patient is taken care of [group laughing].” (resident)

Both focus groups spoke to why residents may provide more specificity in discharge instructions. They agreed that less experience among residents means less knowledge of what may or may not happen after discharge; therefore, residents may strive to provide more detail for families to ensure that all relevant information is covered.

Residents just don't know what is going to happen...they don't have the experience...so they worry [and] think “we have to put in all or our own discharge instructions,” instead of using common ones. (HM faculty)

Both groups also felt that residents may not know all standard discharge instruction templates, with residents adding they may not know their content even if they know they exist, due to rotating from one setting to another. Thus, residents may be more likely to write these details out, compared with supervisors, who may provide a single proposed RSQM of “provide standard discharge instruction template.” However, residents also noted they are often the last physician contact with families before discharge. Therefore, they feel a greater need to cover every detail possible and personalize instructions to ensure an effective discharge, adding that they edit and add to standard discharge instructions when they feel important information is missing.

The templates look nice. . .but. . .looking at them I think, “OK, the family would think this is standard and this doesn't apply to my kid,”... I want to hear about what the doctor thinks personally.”... That's why we take the time to do that. (resident)

Finally, both groups articulated a high level of importance for providing RSQMs to residents. Residents felt RSQMs would provide them with feedback to help drive personal improvement, insights into whether their patients are actually receiving quality care from them, and a meaningful role in quality improvement where they may not otherwise have one. They also noted that objective measures for tracking global improvement throughout training and accurate assessments of their performance are lacking on many rotations. Similar to residents, supervisors noted that RSQMs can provide a more objective way of assessing residents, that feedback on performance drives improvement, and that data drive change. They also noted that the Accreditation Council for Graduate Medical Education requires residents to receive data on their performance and RSQMs can achieve that goal.

If you had objective measures, you...[could] look back...and really see like what went wrong, what could be improved. (resident)

[As an] attending...we get some QI [quality improvement] data back. With resident[s], I believe they get very minimal, if any, and then we expect them to improve. (GP faculty)

Discussion

This study provides insights into differences between RSQMs developed by residents and supervisors and highlights considerations for developing RSQMs in settings beyond the PED, both of which should be considered in the ongoing development of RSQMs.

Measures Important to Residents and Supervisors

Both residents and supervisors, in consensus groups and subsequent focus groups, placed considerable focus on appropriate medication dosing, documentation, and information provided at patient discharge. Medication dosing is foundational to safe care, and medication errors in teaching hospitals are well described, common, and significant.^{25,26} We believe that accurate documentation demonstrates appropriate knowledge, medical decision making, and attention to importance of charting in communication. It also acknowledges the medico-legal importance of documentation. Finally, information provided at discharge is important to not only ensure continued high-quality management and care after discharge but also to prevent unnecessary returns seeking additional care, a known problem.²⁷ With this need, the National Quality Forum has developed quality measures with this focus.²⁸ Given these types of measures are valued by both consensus group and focus group participants, we believe future RSQM development should pay attention to this area.

Considerations for Future RSQM Development

Residents felt that HM and GP are settings of high utility for future RSQM development and should receive prioritization in developing future RSQMs. However, residents and supervisors alike highlighted important nuances in evaluating the fit of RSQMs in different settings and their importance for RSQM development as the process extends into new clinical settings. This finding likely underscores the importance of continuing to develop RSQMs primarily in specific settings rather than seeking global measures for use across settings.

Consensus group participants developed and prioritized similar as well as contrasting measures between residents and supervisors. Subsequent focus groups with these stakeholders showed common rationales for putting forth the measures they did but also offered distinct insights. These findings may indicate “blind spots” possessed by each group for certain areas of resident care processes and work. The presence of these blind spots highlights the relative value of engaging both stakeholder groups in developing future RSQMs. Importantly, however, both resident and supervisor focus group participants felt that residents provided both more potential and more specific RSQMs because they understand the work they are doing at the front line better than supervisors.

Potential Value of RSQMs for Residents

Finally, some of the measures produced by residents in their NGT may emphasize the depth of performance data and feedback they desire about patients they have cared for, including returns to the PED. A recent study underscores that resident experiences with quality improvement are far from ideal, including residents not understanding the vision of quality improvement and feeling they are not valued or valuable in the quality improvement process.²⁹ The types of metrics developed by residents seem to indicate that they are eager for outcome measures related to the care they provide. Although not an intended focus of our study at the outset, these comments were impossible to ignore and therefore formed a focus of investigation in the subsequent focus groups we conducted. This exploration revealed that RSQMs were perceived to be important in providing feedback that can drive improvement and overcome suboptimal performance assessment practices by providing objective data on resident performance.

Limitations

This study has limitations to consider. First, it was conducted at a single institution. Different measures may have been put forth by individuals at another institution. Although we believe that most of the measures developed in this study would apply across institutions, implementation studies will need to investigate whether this is true as well as determine how these measures perform as markers of resident performance. Second, although this work engaged 2 stakeholder groups (supervisors and residents), it did not seek to develop measures among additional stakeholders, such as nurses, patients, and families. Future work should develop and compare

measures that would be set forth by these groups. Third, between the NGT and Delphi groups, we deleted the few proposed measures that could not be extracted from the electronic health record as we did not anticipate such measures being named. Finally, we did not seek the level of clinical experience for residents in this study, which could impact the input they provide. However, it should be noted that all NGT and Delphi resident participants needed to have PED experience to participate, and all resident focus group participants, by the nature of the timing for the focus group, were beyond the intern year. Despite the limitations of this study, it does avoid common pitfalls in consensus group work, such as not describing background information for participants and not defining consensus a priori.²³

Conclusions

RSQMs can continue to advance competency-based medical education efforts where the link between educational outcomes (i.e., what an individual can do at the end of training) and patient care outcomes (i.e., how well patient care needs are met) is both elusive but also key to advancing care quality.^{30,31} This study continues to pave the path forward in developing future RSQMs by exploring specific settings, methods, measures, and stakeholders to consider when undertaking this work.

References

1. van der Leeuw RM, Lombarts KM, Arah OA, et al. A systematic review of the effects of residency training on patient outcomes. *BMC Med.* 2012;10:1.
2. Asch DA, Nicholson S, Srinivas S, et al. Evaluating obstetrical residency programs using patient outcomes. *JAMA.* 2009;302:1277–1283.
3. Bansal N, Simmons KD, Epstein AJ, et al. Using patient outcomes to evaluate general surgery residency program performance. *JAMA Surg.* 2016;151:111–119.
4. Chen C, Petterson S, Phillips R, et al. Spending patterns in region of residency training and subsequent expenditures for care provided by practicing physicians for Medicare beneficiaries. *JAMA.* 2014;312:2385–2393.
5. Sirovich BE, Lipner RS, Johnston M, et al. The association between residency training and internists' ability to practice conservatively. *JAMA Intern Med.* 2014;174:1640–1648.
6. Phillips RL, Petterson SM, Bazemore AW, et al. The effects of training institution practice costs, quality, and other characteristics on future practice. *Ann Fam Med.* 2017;15:140–148.
7. Medicare Payment Advisory Commission. Graduate medical education financing: Focusing on educational priorities. In: Report to Congress: Aligning Incentives in Medicare. Washington, DC: Medicare Payment Advisory Commission. 2010:103-126.
8. Cordasco K, Horta M, Lurie N, et al. How Are Residency Programs Preparing Our 21st Century Internists? A Review of Internal Medicine Residency Programs' Teaching on Selected Topics. RAND Corporation, Working Paper, 2009. Available at: www.rand.org/content/dam/rand/pubs/working_papers/2009/RAND_WR686.pdf. Accessed January 2, 2018.
9. Medicare Payment Advisory Commission. Report to the Congress: improving incentives in the Medicare program. June 2009. Available at: http://www.medpac.gov/docs/default-source/reports/Jun09_EntireReport.pdf. Accessed October 31, 2018.
10. Wiest FC, Ferris TG, Gokhale M, et al. Preparedness of internal medicine and family practice residents for treating common conditions. *JAMA.* 2002;288:2609–2614.
11. 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;85:844–851.
12. Cook DA, West CP. Reconsidering the focus on “outcomes research” in medical education: a cautionary note. *Acad Med.* 2013;88:162–167.
13. Swing SR, Schneider S, Bizovi K, et al. Using patient care quality measures to assess educational outcomes. *Acad Emerg Med.* 2007;14:463–473.
14. Haan CK, Edwards FH, Poole B, et al. A model to begin to use clinical outcomes in medical education. *Acad Med.* 2008;83:574–580.
15. Yin HS, Jay M, Maness L, et al. Health literacy: an educationally sensitive patient outcome. *J Gen Intern Med.* 2015;30:1363–1368.

16. Moreau KA, Eady K. Connecting medical education to patient outcomes: the promise of contribution analysis. *Med Teach*. 2015;37: 1060–1062.
17. Van Melle E, Gruppen L, Holmboe ES, et al. Using contribution analysis to evaluate competency-based medical education programs: it's all about rigor in thinking. *Acad Med*. 2017;92:752–758.
18. Schumacher DJ, Holmboe ES, van der Vleuten C, et al. Developing resident-sensitive quality measures: a model from pediatric emergency medicine. *Acad Med*. 2018;93:1071–1078.
19. Creswell J. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. 6th ed. Upper Saddle River, NJ: Pearson; 2019.
20. Mittiga MR, Schwartz HP, Iyer SB, et al. Pediatric emergency medicine residency experience: requirements versus reality. *J Grad Med Educ*. 2010;2:571–576.
21. Humphrey-Murto S, Varpio L, Wood TJ, et al. The use of Delphi and other consensus group methods in medical education. *Acad Med*. 2016;91:S11.
22. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15:1277–1288.
23. Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval*. 2007;12(10).
24. Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;374:1160–1170.
25. Taylor BL, Selbst SM, Shah AE. Prescription writing errors in the pediatric emergency department. *Pediatr Emerg Care*. 2005;21:822–827.
26. Lewis PJ, Dornan T, Taylor D, et al. Prevalence, incidence and nature of prescribing errors in hospital inpatients. *Drug Safety*. 2009;32:379–389.
27. Navanandan N, Schmidt SK, Cabrera N, et al. The caregiver perspective on unscheduled 72-hour return visits to pediatric acute care sites: a focus on discharge processes. *Acad Pediatr*. 2017;17:755–761.
28. The CTM-15 and CTM-3. 2018. Available at: <https://caretransitions.org/tools-and-resources/>. Accessed August 30, 2018.
29. Butler JM, Anderson KA, Supiano MA, et al. "It feels like a lot of extra work": resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med*. 2017;92:984–990.
30. Carraccio C, Englander R, Holmboe E, et al. Driving care quality: aligning trainee assessment and supervision through practical application of entrustable professional activities, competencies, and milestones. *Acad Med*. 2016;91:199–203.
31. Gruppen L, Frank JR, Lockyer J, et al. Toward a research agenda for competency-based medical education. *Med Teach*. 2017;39:623–630.

Supplemental Appendix 1: RSQM Focus Groups Semi-Structured Interview Guide

We have been working to develop something that we are calling “resident-sensitive quality measures.” What we mean by this is quality measures that meet two requirements (this definition will be posted in the room for all participants to view during the focus group as well):

- 1) Likely attributable to/performed by residents rather than by other individual members of the team or the team in general
- 2) Important to providing quality care to patients

Are there questions about this definition?

Our efforts to date have focused on developing these measures with two distinct groups: 1) faculty and fellow supervisors, and 2) residents. There are areas of overlap in the measures developed by the groups as well as areas of differences. We want to learn more about these similarities and differences to help inform our future efforts to develop additional resident-sensitive quality measures.

Our efforts to date have also focused on developing these measures for the pediatric emergency department setting. Moving forward, we are interested in the types of measures that would likely be best in other settings as well.

Our focus group with you today will focus on exploring the rationale for the similarities and differences between supervisors and residents in developing resident-sensitive quality measures as well as exploring other settings and the types of measures that may be best in those settings.

We will be audiorecording this session. The recording will be transcribed and all data analysis will be done from the transcript of the session.

Do you have any questions before we begin?

1) In the pediatric emergency department setting, both residents and faculty/fellow supervisors placed considerable focus on resident-sensitive quality measures in three areas: appropriate medication dosing, documentation, and information provided at patient discharge. Examples include providing the correct dose of albuterol and decadron to patients presenting with an asthma exacerbation; documenting time since injury, presence or absence of loss of consciousness, presence or absence of vomiting, and a full neurologic exam for patients presenting with closed head injury; and expected course and nasal suctioning instructions for patients being discharged with a diagnosis of bronchiolitis.

[Resident Group Only] Do you think these types of measures (appropriate medication dosing, documentation, and information provided at patient discharge) are the most important types of resident-sensitive quality measures for the pediatric emergency department setting? If so, why? If not, what types of measures do you believe are missing?

Do you think it would be helpful to have resident-sensitive quality measures in other workplace settings? If so, which settings? If not, why not?

In each of those settings, what types of measures do you think would make the best resident-sensitive quality measures?

Drill down probes if needed:

***Do you think the 3 categories above are appropriate? (medication dosing, documentation, and discharge information)

***What other categories would make the best resident-sensitive quality measures?

2) In our work so far residents tended to be more specific in the resident-sensitive quality measures they produce than faculty/fellow supervisors.

Why do you think this is?

3) In our work so far, residents have tended to produce more total measures than faculty/fellow supervisors.

Why do you think this is?

4) In our work so far, residents have also delineated more measures around timeliness of care.

Why do you think this is? Do you think this type of resident-sensitive quality measures is as important in the other settings you mentioned before? Why or why not?

5) In our work so far, faculty/fellow supervisors, compared to residents, focused on providing families with standard or pre-made discharge instructions.

Why do you think this is?

6) In our work so far, supervisors tended to focus on order set use whereas residents tended to focus on orders in general.

Why do you think this is?

7) How important do you think it is to provide residents with quality measures for the work they complete?

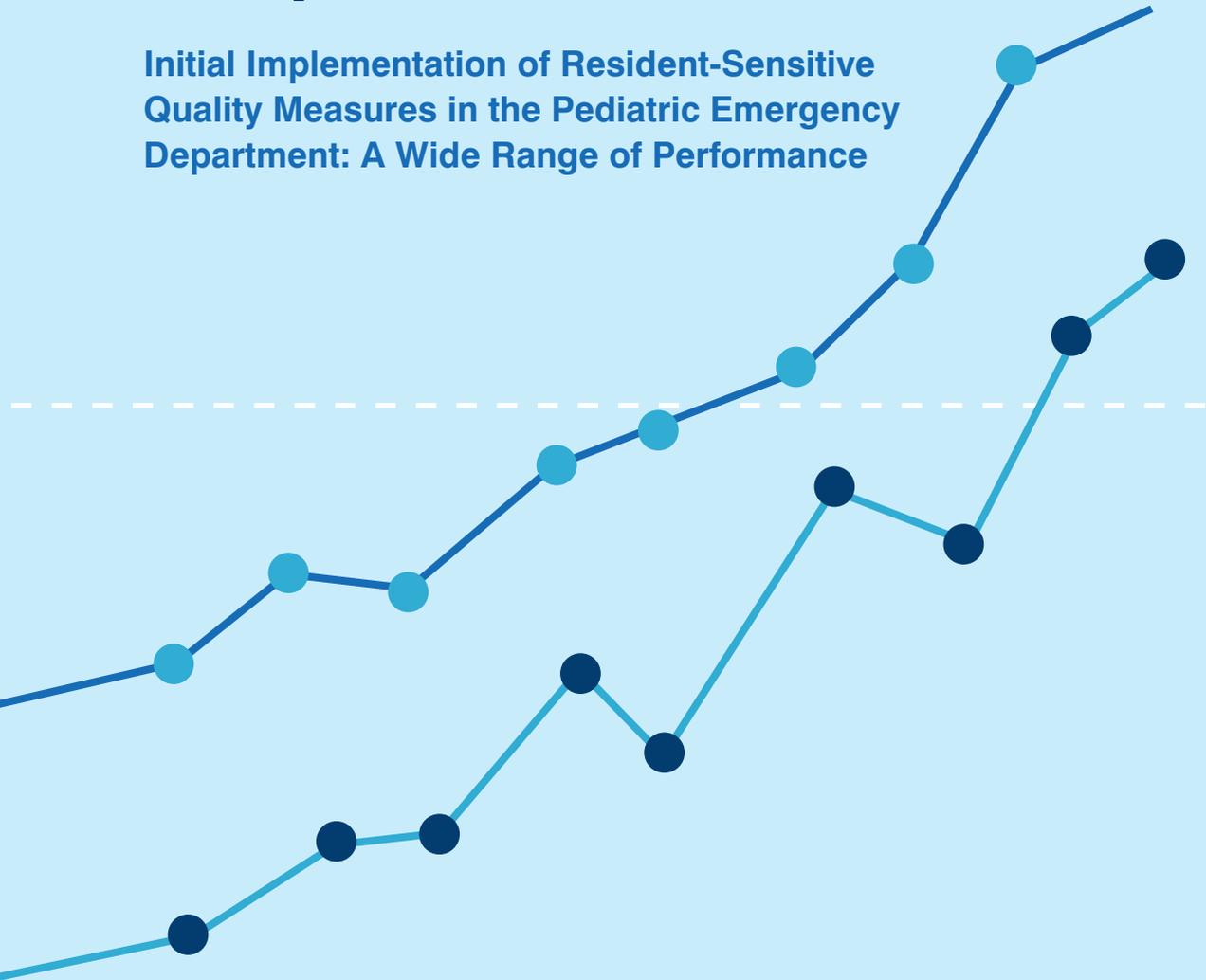
Why is this important or not important?

8) Is there anything else you would like to share with us before we end?



Chapter 6

Initial Implementation of Resident-Sensitive Quality Measures in the Pediatric Emergency Department: A Wide Range of Performance



Schumacher DJ, Martini A, Holmboe E, Carraccio C, van der Vleuten C, Sobolewski B, Busari J, Byczkowski TL. "Initial Implementation of Resident-Sensitive Quality Measures in the Pediatric Emergency Department: A Wide Range of Performance." [published online ahead of print January 7, 2020]. *Academic Medicine*. doi: 10.1097/ACM.0000000000003147.

Abstract

Purpose

A lack of quality measures aligned with residents' work led to the development of resident-sensitive quality measures (RSQMs). This study sought to describe how often residents complete RSQMs, both individually and collectively, when they are implemented in the clinical environment.

Method

During academic year 2017–2018, categorical pediatric residents in the Cincinnati Children's Hospital Medical Center pediatric emergency department were assessed using RSQMs for acute asthma exacerbation (21 RSQMs), bronchiolitis (23 RSQMs), and closed head injury (19 RSQMs). Following eligible patient encounters, all individual RSQMs for the illnesses of interest were extracted from the health record. Frequencies of 3 performance classifications (opportunity and met, opportunity and not met, or no opportunity) were detailed for each RSQM. A composite score for each encounter was calculated by determining the proportion of individual RSQMs performed out of the total possible RSQMs that could have been performed.

Results

Eighty-three residents cared for 110 patients with asthma, 112 with bronchiolitis, and 77 with closed head injury during the study period. Residents had the opportunity to meet the RSQMs in most encounters but exceptions existed. There was a wide range in the frequency of residents meeting RSQMs in encounters in which the opportunity existed. One closed head injury measure was met in all encounters in which the opportunity existed. Across illnesses, some RSQMs were met in almost all encounters, while others were met in far fewer encounters. RSQM composite scores demonstrated significant range and variation as well—asthma: mean = 0.81 (standard deviation [SD] = 0.11) and range = 0.47–1.00, bronchiolitis: mean = 0.62 (SD = 0.12) and range = 0.35–0.91, and closed head injury: mean = 0.63 (SD = 0.10) and range = 0.44–0.89.

Conclusions

Individually and collectively, RSQMs can distinguish variations in the tasks residents perform across patient encounters.

Graduate medical education training is a time for service learning, with trainees impacting the care of the patients they serve. Yet, the relationship between residency training and patient care is only partially understood at best.¹ Evidence suggests residents can be positioned to improve care while in training,^{2,3} and multiple large-scale studies show a lasting relationship between residency programs and the care provided by graduates of those programs.⁴⁻⁸ However, which characteristics of training programs lead to better or worse patient care are not known. Furthermore, the relationship between the quality of care provided by residents in training and patient outcomes has been explored very little. Efforts to disentangle the complexities of the clinical learning environment to understand these relationships have been challenged, likely by the innate messiness of the real world training environment.⁹ Even so these relationships deserve attention if medical educators hope to train a physician workforce that can achieve the Triple (or Quadruple) Aim.¹⁰ Given the impact and scope of resident contributions to patient care outcomes, we believe that the link between the quality of care that residents deliver and patient care outcomes requires exploration. Therefore, this study sought to understand the operational use of quality measures in assessing the work that residents perform as a first step in linking trainee performance with patient care outcomes.

While some quality measures used for physicians in practice are appropriate for residents as well, many traditional quality measures are not aligned with the tasks that residents perform.^{11,12} With limited previous efforts to fill this gap,^{11,13-15} we have committed to developing quality measures that reflect the work performed by residents, called resident-sensitive quality measures (RSQMs).^{16,17} To create an initial set of RSQMs for 3 common, acute illnesses (acute asthma exacerbation [hereafter simply asthma], bronchiolitis, and closed head injury [CHI]) that present to pediatric emergency departments (PEDs), we previously engaged residents as well as faculty and fellow supervisors to develop and prioritize RSQMs for these illnesses that met two criteria: (1) the measure is likely performed by the resident rather than another member of the team and (2) the measure is important to the care of a patient with the illness of interest.^{16,17} Engaging individuals with content expertise in developing these measures helped us build validity evidence into the measures.^{16,17} However, we do not know if these measures are aligned with tasks actually performed by residents and to what extent residents do or do not perform the measures. We believe it is important to understand this relationship, because residency programs have the potential to use RSQMs to both assess and provide feedback to trainees and to provide meaningful information that could help improve the residency program. Thus, this study sought to describe how often residents complete these RSQMs, both individually and collectively, when they are implemented in the actual learning environment. This work has implications for informing RSQM development and implementation in other settings and specialties.

Method

Setting, Participants, and Eligible Patient Encounters

During the 2017–2018 academic year, we recruited categorical (i.e., not in a combined training program, such as internal medicine and pediatrics) pediatric residents rotating through the Cincinnati Children’s Hospital Medical Center (CCHMC) PED to be assessed using previously prioritized RSQMs (described below) for 3 illnesses of interest: asthma, bronchiolitis, and CHI.

Eligible patient encounters, the focus of analysis for this study, were those in which participating residents cared for patients from the beginning of their PED visit to the time of disposition in the electronic health record (EHR), as well as those with a discharge diagnosis of asthma, bronchiolitis, or CHI. Reactive airway disease in younger children that is treated in a manner that mirrors acute asthma exacerbation treatment in older children was included in eligible asthma encounters. For bronchiolitis encounters, patients were required to be 12 months old or younger for inclusion given that bronchiolitis has more overlap with reactive airway disease after this age.

Measures

We used the RSQMs (listed in Appendix 1) to assess resident performance (i.e., performed or not performed). The RSQMs were focused on common, acute diagnoses in the PED¹⁸ and were previously developed and prioritized with both faculty and fellow supervisors¹⁶ as well as residents.¹⁷

Data Collection

Clinical research coordinators working in the PED determined which patient encounters with participating residents were eligible for inclusion in the study using the criteria given above. Following encounters that were included in this study, all individual RSQMs for the illnesses of interest were extracted from the EHR through manual chart review using a REDCap¹⁹ survey (Vanderbilt, Nashville, Tennessee), hosted at CCHMC, to record applicable data. All EHR extractions were independently performed by two individuals (clinical research coordinators, including A.M., and lead investigator, D.J.S.) to ensure accuracy. Discrepancies were most often resolved by reviewing the encounter in the EHR again to verify accuracy of the initial coding of measures.

Data Analysis

First, we determined whether residents had the opportunity to perform the RSQM in the encounter. Then, within encounters where residents had the opportunity to perform the RSQM, we determined whether the resident performed it. These determinations resulted in 3 performance classifications into which a patient encounter might fall: (1) the resident had the opportunity to perform the RSQM and did (opportunity and met), (2) the resident had the opportunity to perform

the RSQM and did not (opportunity and not met), or (3) the resident did not have the opportunity to perform the RSQM (no opportunity). We detailed the frequencies of each individual RSQM for each encounter. Finally, we calculated a RSQM composite score (possible range: 0.00–1.00) for each encounter by determining the proportion of individual RSQMs performed out of the total possible RSQMs that could have been performed for that encounter. We used a proportion rather than the total number for two reasons. First, some individual RSQMs were not applicable to all encounters. Second, the illnesses of interest had different numbers of individual RSQMs: asthma had 21 RSQMs, bronchiolitis had 23 RSQMs, and CHI had 19 RSQMs. Using a proportion ameliorated the issue of varied scales between and within illnesses. Using these composite scores, we then calculated the mean, standard deviation (SD), median, and interquartile range (IQR) for each of the illnesses of interest.

This study was determined to be exempt by the CCHMC Institutional Review Board.

Results

A total of 83 categorical pediatric residents cared for 110 patients presenting with asthma, 112 patients presenting with bronchiolitis, and 77 patients presenting with CHI during academic year 2017–2018.

Frequencies of Individual RSQMs Across Patient Encounters

Appendix 1 details the frequency with which residents had the opportunity to meet a measure and met it, had the opportunity to meet a measure and did not meet it, and did not have the opportunity to meet a measure. Additionally, the last column of Appendix 1 details common reasons why residents did not have the opportunity to meet certain measures. As Appendix 1 shows, residents had the opportunity to meet the RSQMs in most encounters, with asthma presenting the most instances in which residents did not have the opportunity to meet measures.

When given the opportunity, there was a wide range in the frequency of residents meeting measures. For asthma, there were some measures that residents had the opportunity to meet that were met in nearly all encounters, such as noting patient acuity in documentation (A8), documenting work of breathing (A9), documenting aeration or air exchange (A10), documenting the presence or absence of wheezing (A11), and ensuring at least 3 descriptive words were used in respiratory exam documentation (A12). However, for other measures that residents had the opportunity to meet, residents met them in far fewer encounters, such as documenting previous intubation or bilevel positive airway pressure use (A7), which was met in 29/110 (26%) encounters, and documenting their own Pediatric Respiratory Assessment Measure (PRAM) score (A13), which was met in 65/110 (59%) encounters. As Appendix 1 shows, similar variations were also present across bronchiolitis and CHI encounters, with only one measure being met in

all instances in which residents had the opportunity to meet it (C3, the CHI RSQM focused on documenting the mechanism of injury).

Distributions of RSQM Composite Scores

As Figure 1 shows, RSQM composite scores, which combine all eligible individual RSQMs for a given encounter (see above), demonstrated significant range and variation among the 3 illnesses of interest. Asthma had the highest mean (0.81, SD = 0.11) and median (0.81, IQR = 0.74–0.88) composite scores, while bronchiolitis (mean = 0.62, SD = 0.12; median = 0.61, IQR = 0.53–0.71) and CHI (mean = 0.63, SD = 0.10; median = 0.63, IQR = 0.56–0.68) had mean and median composite scores that were lower and similar to one another. Bronchiolitis had the lowest overall composite score, as well as the widest range of composite scores across encounters (bronchiolitis: 0.35–0.91 vs asthma: 0.47–1.00 and CHI: 0.44–0.89). Only asthma had any perfect (i.e., 1.00) composite scores.

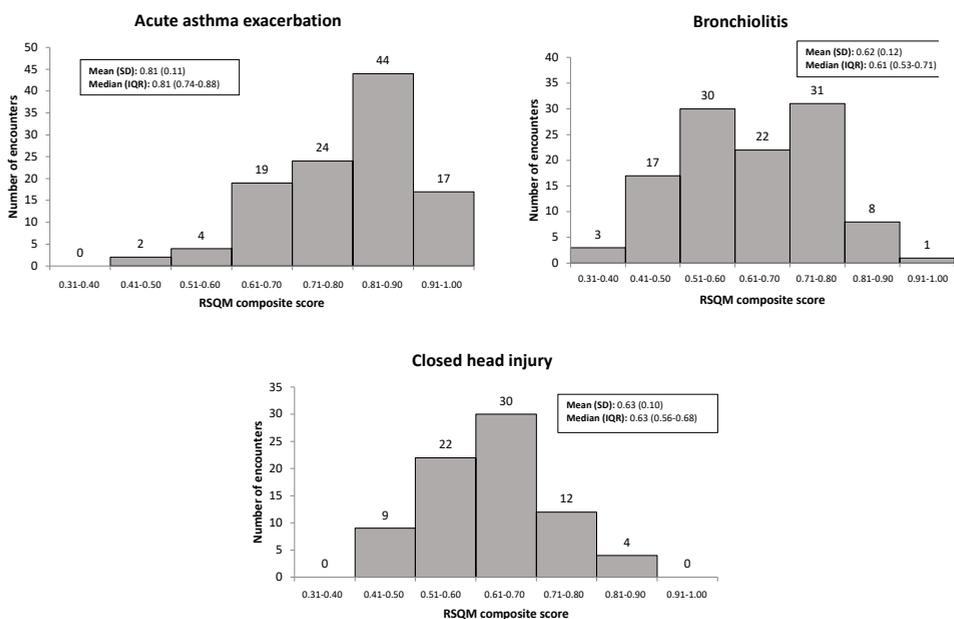


Figure 1: Distribution of RSQM composite scores for acute asthma exacerbation, bronchiolitis, and closed head injury patient encounters, Cincinnati Children’s Hospital Medical Center pediatric emergency department, academic year 2017–2018. The authors calculated a composite score for each encounter by determining the proportion of individual RSQMs performed out of the total possible RSQMs that could have been performed for that encounter. See Appendix 1 for a full list of the RSQMs. Abbreviations: RSQM, resident-sensitive quality measure; SD, standard deviation; IQR, interquartile range.

Discussion

This study explored the initial implementation of RSQMs, both individually and collectively (via composite scores), in the clinical learning environment. Overall, we demonstrate that individual RSQMs capture tasks that residents are performing quite well when implemented in the actual clinical learning environment, as seen by the number of RSQMs that residents did perform in patient encounters when they had the opportunity to do so. Moreover, RSQMs can distinguish variations in the tasks residents perform across encounters, as demonstrated by our findings that some individual RSQMs were not met in many encounters, that there was wide variation across individual RSQMs in whether or not they were met when residents were given the opportunity, and that the RSQM composite scores for each illness also displayed considerable range across encounters. These findings support using individual RSQMs and RSQM composite scores to represent work performed by residents in the PED. However, we also uncovered limitations of RSQMs that could inform future development and implementation of RSQMs in other settings and specialties.

Defining the Benefits of Using Individual RSQMs and RSQM Composite Scores

As work-based assessment efforts continue to evolve, sufficient data to inform defensible summative decisions regarding performance is a common shortcoming.²⁰ Additionally, getting assessors to complete assessment forms can be challenging, faculty development for optimizing assessment is lacking, and direct observation of residents by assessors is difficult to ensure.^{21–23} Furthermore, traditional assessment efforts may suffer from range restriction, where assessors use only a small portion (often the upper portions) of the rating scale or framework provided.^{24,25} Based on the findings in this study, using RSQMs for the purposes of resident assessment may help address some of these known challenges. We showed that residents often had the opportunity to complete the RSQMs in patient encounters, which could provide resident performance data to augment a supervisor's assessment with meaningful clinical data. However, the most significant benefit of RSQMs may be in their ability to distinguish a wide range of differences in the quality of care provided by residents (via the number of possible tasks that were or were not performed) across encounters. For example, the RSQM composite scores for asthma (0.47–1.00) and bronchiolitis (0.35–0.91) spanned more than half of the possible range, while the span for CHI scores was close to half of the range (0.44–0.89). Moreover, for both bronchiolitis and CHI, the IQR of composite scores approached 0.50 on the lower bound. Traditional assessment data displaying this range is uncommon. Thus, RSQM data may be useful for formative and summative assessment decisions as they provide the ability to distinguish variations in the tasks residents perform across encounters. Further work will be needed, though, to assess whether the variation we observed in RSQM data corresponds with other measures of resident performance and to define and study RSQMs in other settings and specialties.

This study did not seek to understand the potential benefits of RSQMs in providing residents with actionable feedback about their performance. Given that only one RSQM was met by all residents in this study, future efforts should seek to understand how, and under what circumstances, RSQMs may benefit residents' personal quality improvement efforts. Furthermore, future efforts should consider how RSQMs with poorer performance across residents might inform curricular changes or systems-level quality improvement initiatives.

Defining the Limits of Using Individual RSQMs and RSQM Composite Scores

While this study supports the use of RSQMs to capture resident performance, limitations of their use were also encountered. Residents did not always have the opportunity to meet all measures in all patient encounters. Measures that cannot be met in given encounters could be treated in 1 of 2 ways. First, they could be treated as we have done, by developing RSQM composite scores, in which a measure can be included if it applies (i.e., if the resident had the opportunity to perform the measure) and, conversely, not included when it does not apply (i.e., if the resident did not have the opportunity to perform the measure). This means that the total possible number of applicable RSQMs will vary from one encounter to another and that a proportion of completed RSQMs will be used to compare performance from one encounter to another. The drawback to this approach is that it prevents direct comparisons of performance across encounters, making comparisons more nuanced. However, these imperfect comparisons may be less important when using RSQM composite scores for formative rather than summative purposes. The benefit to this approach is that it uses all available RSQMs to inform a RSQM composite score for each encounter. The second way to handle RSQMs that cannot be met in an encounter is to simply use RSQMs only in encounters where all measures can be met. Given that residents are typically provided with a paucity of quality metrics relevant to their performance, we believe using any available RSQMs whenever possible is paramount (i.e., that option 1 is preferable to option 2).

Beyond the clinical context precluding residents from meeting an individual RSQM, one asthma RSQM was contingent on completing another RSQM. That is, residents could only complete the RSQM focused on their PRAM score matching the initial medication orders placed for the patient if they had documented their own PRAM score, another asthma RSQM. In 45 encounters, residents did not document their own PRAM score and thus did not receive credit for completing that individual RSQM. However, it also meant that they could not complete the subsequent RSQM regarding the initial medication orders matching their PRAM score for the patient. Should residents also not receive credit for this individual RSQM or should this measure not be included in their RSQM composite score? We believe the latter option is most prudent because it avoids penalizing residents twice if they do not document their own PRAM score.

These limitations, which we encountered when implementing our RSQMs in the PED, can and should inform the development and implementation of RSQMs in other settings and specialties.

Furthermore, identifying which residents provided care to patients may be easier in the PED than in other settings, such as the inpatient environment, where multiple residents may have touchpoints of care on a given day. Future RSQM development will need to consider this as it expands to other contexts.

Study Limitations

There are limitations to consider in this implementation study. First, this study was performed at a single department at a single institution and involved the roles assigned to residents and others at that institution, as well as the site-specific EHR. Thus, our findings may not be generalizable to other settings. However, we believe that our implementation experiences can aid efforts to develop and implement RSQMs in other settings and specialties. Second, the individual RSQMs were obtained via manual chart review. Therefore, human extraction errors are possible. We sought to minimize this through employing two individuals to perform data extraction. Manual extraction was chosen because some of the RSQMs cannot be automatically extracted from the EHR. This could be viewed as a limitation of some of the RSQMs, which may not be easy to extract from the chart and thus might have limited usability beyond the study setting. However, natural language processing is likely a ready solution for obtaining these items from the EHR. Third, we considered patient encounters and not overall resident performance in our implementation analysis. This was by design because resident performance varies by encounter, and we wanted to know how well the RSQMs captured resident performance for each encounter. Yet, there may be certain measures that individual residents consistently do or do not perform, suggesting that residents could influence encounter-based data. Future studies should explore this area.

Conclusions

RSQMs, both individually and collectively, provide some promise for addressing known challenges in work-based assessment and in meaningfully involving residents in quality improvement efforts.²⁰⁻²⁵ Having established that the RSQMs we used largely apply in clinical encounters involving pediatric residents, future efforts should consider the validity and reliability of RSQMs when used for resident assessment, as well as their relationship to other measures of patient care quality and outcomes. Furthermore, individuals undertaking the work of developing RSQMs in other settings and specialties should continue to attend to and share what works and does not work in various contexts to continue advancing this field.

References

1. van der Leeuw RM, Lombarts KMJM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med.* 2012;10:65.
2. Denson JL, Jensen A, Saag HS, et al. Association between end-of-rotation resident transition in care and mortality among hospitalized patients. *JAMA.* 2016;316:2204–2213.
3. Lau BD, Streiff MB, Pronovost PJ, Haider AH, Efron DT, Haut ER. Attending physician performance measure scores and resident physicians' ordering practices. *JAMA Surg.* 2015;150:813–814.
4. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA.* 2009;302:1277–1283.
5. Bansal N, Simmons KD, Epstein AJ, Morris JB, Kelz RR. Using patient outcomes to evaluate general surgery residency program performance. *JAMA Surg.* 2016;151:111–119.
6. 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;312:2385–2393.
7. Sirovich BE, Lipner RS, Johnston M, Holmboe ES. The association between residency training and internists' ability to practice conservatively. *JAMA Intern Med.* 2014;174:1640–1648.
8. Phillips RL, Petterson SM, Bazemore AW, Wingrove P, Puffer JC. The effects of training institution practice costs, quality, and other characteristics on future practice. *Ann Fam Med.* 2017;15:140–148.
9. 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;92:1740–1748.
10. Bodenheimer T, Sinsky C. From Triple to Quadruple Aim: Care of the patient requires care of the provider. *Ann Fam Med.* 2014;12:573–576.
11. 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;85:844–851.
12. Cook DA, West CP. Reconsidering the focus on "outcomes research" in medical education: A cautionary note. *Acad Med.* 2013;88:162–167.
13. Swing SR, Schneider S, Bizovi K, et al. Using patient care quality measures to assess educational outcomes. *Acad Emerg Med.* 2007;14:463–473.
14. Haan CK, Edwards FH, Poole B, Godley M, Genuardi FJ, Zenni EA. A model to begin to use clinical outcomes in medical education. *Acad Med.* 2008;83:574–580.
15. Yin HS, Jay M, Maness L, Zabar S, Kalet A. Health literacy: An educationally sensitive patient outcome. *J Gen Intern Med.* 2015;30:1363–1368.
16. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: A model from pediatric emergency medicine. *Acad Med.* 2018;93:1071–1078.
17. Schumacher DJ, Martini A, Holmboe E, et al. Developing resident-sensitive quality measures: Engaging stakeholders to inform next steps. *Acad Pediatr.* 2019;19:177–185.

18. Mittiga MR, Schwartz HP, Iyer SB, Gonzalez Del Rey JA. Pediatric emergency medicine residency experience: Requirements versus reality. *J Grad Med Educ.* 2010;2:571–576.
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42:377–381.
20. Heaslip V, Scammell JM. Failing underperforming students: The role of grading in practice assessment. *Nurse Educ Pract.* 2012;12:95–100.
21. Holmboe ES, Ward DS, Reznick RK, et al. Faculty development in assessment: The missing link in competency-based medical education. *Acad Med.* 2011;86:460–467.
22. Iobst WF, Sherbino J, Cate OT, et al. Competency-based medical education in postgraduate medical education. *Med Teach.* 2010;32:651–656.
23. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: A conceptual model. *Med Educ.* 2011;45:1048–1060.
24. Regehr G, Bogo M, Regehr C, Power R. Can we build a better mousetrap? Improving the measures of practice performance in the field practicum. *J Social Work Educ.* 2007;43:327–344.
25. Dudek NL, Marks MB, Regehr G. Failure to fail: The perspectives of clinical supervisors. *Acad Med.* 2005;80(10 Suppl):S84–S87.

Appendix 1

Individual Resident-Sensitive Quality Measures (RSQMs) Performance for Acute Asthma Exacerbation, Bronchiolitis, and Closed Head Injury Patient Encounters, Cincinnati Children's Hospital Medical Center Pediatric Emergency Department, Academic Year 2017–2018

RSQM no. ^a	RSQM	Performance classification, ^b no.			Common reasons why residents did not have opportunity to perform RSQM
		Opportunity and met	Opportunity and not met	No opportunity	
Acute asthma exacerbation (110 patient encounters total)					
A1	Use asthma order set	64	20	26	Patient presented through trauma or resuscitation bay and initial orders placed by attending in this context
A2	Correct medication dose ordered for albuterol	59	22	29	Patient presented through trauma or resuscitation bay and initial orders placed by attending in this context
A3	Use of dexamethasone as steroid	71	5	34	Patient presented through trauma or resuscitation bay and initial orders placed by attending in this context
A4	Correct medication dose ordered for dexamethasone	68	3	39	Patient presented through trauma or resuscitation bay and initial orders placed by attending in this context
A5	Time from resident assigning self to patient to resident entering steroid order	44	27	39	Patient presented through trauma or resuscitation bay and initial orders placed by attending in this context
A6	Correct medication dose ordered for ipratropium	51	1	58	<ul style="list-style-type: none"> • Patient presented through trauma or resuscitation bay and initial orders placed by attending in this context • Medication not given during encounter
A7	Documentation of previous intubation or bilevel positive airway pressure for asthma	29	81	0	
A8	Note the acuity of the patient in documentation	109	1	0	
A9	Documentation of work of breathing	104	6	0	
A10	Documentation of aeration or air exchange	104	6	0	
A11	Presence or absence of wheezing documented	108	2	0	
A12	Ensure at least 3 descriptive words used in respiratory exam documentation	109	1	0	
A13	Resident documents own Pediatric Respiratory Assessment Measure (PRAM) score	65	45	0	
A14	Resident-assigned PRAM score matches resident-placed initial medication orders	55	10	45	Resident did not document own PRAM score
A15	Document patient response to intervention	93	10	7	No intervention
A16	Documentation of disposition decision	100	10	0	
A17	Use of standardized dosing for discharge medication (i.e., dexamethasone)	57	2	51	<ul style="list-style-type: none"> • Patient admitted • Patient not given dexamethasone for home

RSQM no. ^a	RSQM	Performance classification, ^b no.			Common reasons why residents did not have opportunity to perform RSQM
		Opportunity and met	Opportunity and not met	No opportunity	
A18	Home dexamethasone instructions documented in written discharge instructions	44	15	51	<ul style="list-style-type: none"> • Patient admitted • Patient not given dexamethasone for home
A19	State who to follow-up with and provide their contact information in discharge papers	53	12	45	Patient admitted
A20	Documentation of needing albuterol more often than every 4 hours as a reason to return in written discharge instructions	24	41	45	Patient admitted
A21	Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	29	36	45	Patient admitted
Bronchiolitis (112 patient encounters total)					
B1	Follow bronchiolitis pathway	66	46	0	
B2	Document birth history (preemie or not a preemie)	69	43	0	
B3	Day of illness clearly documented	111	1	0	
B4	Documentation of previous wheezing	19	93	0	
B5	Assessment of severity documented	108	4	0	
B6	Effort of breathing documented	109	3	0	
B7	Documented quality of air entry (normal, decreased, etc.)	78	34	0	
B8	Documentation of wheezing	90	22	0	
B9	Documentation of crackles	94	18	0	
B10	Documented presence or absence of subcostal retractions	60	52	0	
B11	Documented presence or absence of intercostal retractions	29	83	0	
B12	Documented presence or absence of suprasternal retractions	28	84	0	
B13	Oxygen saturation clearly documented	62	50	0	
B14	Hydration status clearly documented	38	74	0	
B15	Documentation of patient response to specific therapeutics (i.e., how they responded to suctioning, how they responded to breathing treatment, how they responded to normal saline bolus, etc.)	82	23	7	No intervention
B16	Oral feeding tolerance clearly documented	47	65	0	
B17	Documentation of justification for appropriate disposition (sent home versus admitted)	98	14	0	

RSQM no. ^a	RSQM	Performance classification, ^b no.			Common reasons why residents did not have opportunity to perform RSQM
		Opportunity and met	Opportunity and not met	No opportunity	
B18	Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	27	25	60	Patient admitted
B19	Use standard or premade discharge instructions for bronchiolitis	48	4	60	Patient admitted
B20	Documentation of poor feeding as a reason to return in written discharge instructions	22	30	60	Patient admitted
B21	Bulb suction teaching for home ordered	14	38	60	Patient admitted
B22	State who to follow-up with and provide their contact information in discharge papers	43	9	60	Patient admitted
B23	State appropriate number of days to follow-up in discharge papers	30	22	60	Patient admitted
Closed head injury (77 patient encounters total)					
C1	Closed head injury or PECARN ^c best practice advisory used	32	45	0	
C2	Appropriate closed head injury or PECARN ^c pathway used	11	62	4	Patient arrived from outlying hospital with computed tomography scan of head already completed
C3	Mechanism of injury documented	77	0	0	
C4	Documentation of time of injury	67	10	0	
C5	Documentation of presence or absence of loss of consciousness	61	16	0	
C6	Documentation of presence or absence of emesis	70	7	0	
C7	Documentation of whether patient is back to baseline or not	61	16	0	
C8	Documentation of presence or absence of other head injury in the history	31	46	0	
C9	Assessment of severity documented	76	1	0	
C10	Documentation of presence or absence of hematoma on physical exam (if present, location and size also documented)	41	36	0	
C11	Thorough head exam (head, eyes, skull) documented	25	52	0	
C12	Documentation of Glasgow Coma Scale score	25	52	0	
C13	Full neurology exam documented	2	75	0	
C14	Documentation of presence or absence of other non-head injury on physical exam	53	24	0	
C15	Appropriate differential diagnosis and medical decision making documented	73	4	0	
C16	Reassessments of patient documented	40	37	0	

RSQM no. ^a	RSQM	Performance classification, ^b no.			Common reasons why residents did not have opportunity to perform RSQM
		Opportunity and met	Opportunity and not met	No opportunity	
C17	Return to school or play recommendations in discharge papers	20	22	35	Patient younger than school age
C18	Use of standard or premade discharge instructions for diagnosis	62	9	6	Patient admitted
C19	Appropriate follow-up (sports medicine, rehabilitation, neurology, etc.) recommended	62	9	6	Patient admitted

^aFor these numbers: A, acute asthma exacerbation; B, bronchiolitis; C, closed head injury.

^bSee the “Data Analysis” section of the main text for more information on the performance classifications.

^cPECARN is the Pediatric Emergency Care Applied Research Network, which conducted a study that delineated a decision rule for managing closed head injury in Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: A prospective cohort study. *Lancet*. 2009;374:1160–1170.



Chapter 7

Resident-Sensitive Quality Measures in the Pediatric Emergency Department: Exploring Relationships With Supervisor Entrustment and Patient Complexity and Acuity



Schumacher DJ, Holmboe E, Carraccio C, Martini A, van der Vleuten C, Busari J, Sobolewski B, Byczkowski TL. "Resident-Sensitive Quality Measures in the Pediatric Emergency Department: Exploring Relationships with Supervisor Entrustment and Patient Complexity and Acuity." *Academic Medicine* 2020. In press.

Abstract

Purpose

This study sought to understand the association between resident-sensitive quality measures (RSQMs), supervisor entrustment decisions (as measured by supervision level assigned), and patient acuity and complexity for resident encounters in the pediatric emergency department (PED).

Method

Pediatric residents rotating through Cincinnati Children's Hospital Medical Center PED were recruited during the 2017–2018 academic year for the purpose of collecting the following data from their patient encounters for 3 illnesses (asthma, bronchiolitis, and closed head injury [CHI]): supervisor entrustment decision, RSQMs relevant to the care provided, and patient acuity and complexity. The association of RSQM composite scores with the other variables of interest was measured by using mixed models to account for observations nested within residents.

Results

A total of 83 residents cared for 299 patients (asthma, 110; bronchiolitis, 112; CHI, 77). Entrustment decisions were positively associated with asthma RSQM composite scores (beta-coefficient 0.03; $P = .0004$). There was no significant association between RSQM composite scores for bronchiolitis or CHI with entrustment decisions. RSQM composite scores were significantly higher when acuity was also higher and significantly lower when acuity was also lower for both asthma ($P = .0009$) and bronchiolitis ($P = .01$). However, RSQM composite scores were almost identical between levels of acuity for CHI ($P = .94$). There was no association between RSQM composite scores and patient complexity.

Conclusions

This study found limited relationships between RSQM composite scores and entrustment decisions but offers insight into how RSQMs could be used for resident assessment and feedback.

Many traditional quality measures are not aligned with the work residents complete.^{1,2} This reality led us to develop resident-sensitive quality measures (RSQMs), which are quality measures developed and prioritized through expert consensus to meet 2 criteria: important to the care of a patient with an illness of interest and likely performed primarily by a resident and not another team member (or the team collectively).^{3,4} The RSQMs we have developed have largely been shown to both capture resident performance (e.g., most often, residents had the primary opportunity to perform or not perform the clinical measure) as well as demonstrate a wide range of performance across residents and encounters, with some measures being performed often and some measures being rarely performed (as detailed in chapter 6). However, we do not yet fully understand the variation in the demonstrated performance of residents or whether this variation leads to different levels of entrustment ratings by supervisors. Further elucidating this through understanding the relationships between RSQMs and other variables will be important. This study seeks to achieve this goal by exploring the association between RSQMs and supervisor entrustment as well as patient complexity and acuity.

Understanding the relationship between RSQMs and supervisor entrustment is particularly timely and relevant given recent advances in competency-based assessment that have seen the international surge in use of the entrustment framework and specifically entrustable professional activities (EPAs).⁵⁻⁸ Taking a synthetic view of assessment, EPAs seek to define the foundational activities of a profession that individuals in that profession must be entrusted to perform. They further specify how much supervision individuals require to perform the activities until they are entrusted to execute them without supervision. Unlike most assessment frameworks, which focus only on the learner, EPAs and entrustment theoretically place the focus on the learner as well as on the outcome of care that patients receive.^{9,10} For example, in assessing a learner's ability to perform the EPA "provide care for a well newborn," the goal is to determine the level of supervision, including the possibilities of direct or no supervision, that a learner needs to ensure safe and effective care of that newborn. Despite the patient focus of EPAs, the link between entrustment decisions and quality care has not been demonstrated. Based on our method for developing them, RSQMs have built-in clinical validity evidence because one of the criteria used in developing them is importance to the care of patients with the illnesses of interest. If entrustment decisions have a positive association with RSQM performance, that would offer validity evidence for entrustment ratings based on Messick's validity criterion of relationship to other variables.¹¹

More globally, the relationship between care delivered by residents and the impact of this care on patient outcomes is also not well understood.¹² This is likely due to the complexities involved, including variability in supervision approaches (e.g., closeness of clinical oversight,¹³⁻¹⁵ propensity of the supervisor to alter resident care before it reaches the patient, trust propensity of the supervisor¹⁶) and the challenge of measuring the impact of care attributed to the resident

on patient experiences and outcomes.² This latter consideration is complicated both by extrinsic and intrinsic factors. Examples of extrinsic factors that influence care include contributions to care from other team members and patient complexity and acuity, that in turn can affect the degree of trust that a supervisor places in a resident to provide patient care.¹⁷ Stated differently, in the case of extrinsic factors, a supervisor may assign a lower level of entrustment to a highly performing resident to ensure a sicker patient receives appropriate care. In the case of intrinsic factors, such as a more poorly performing resident, a supervisor may also assign a lower level of entrustment to ensure that a routine patient receives appropriate care. Despite these challenges, some findings suggest that resident involvement in care and continuity in care may be beneficial for patients, and large-scale studies show a lasting imprint from residency institutions on care provided by graduates of those institutions.^{18–24} However, characteristics of training programs that contribute to better or worse care are not known, and work to traverse this quagmire is trying.²⁵ While challenging, we believe the relationship between the quality of resident care and the quality of patient care must be better understood.²⁶

To address this gap, our study seeks to understand the association between RSQMs and supervisor entrustment for encounters in the pediatric emergency department (PED) where residents are caring for patients. Given that patient acuity and complexity can modify the relationship between resident performance and supervisor entrustment, we also consider the relationship between these characteristics and RSQMs.

Method

Setting and Participants

During the 2017–2018 academic year, we recruited categorical pediatric residents across all 3 years of training rotating through the Cincinnati Children’s Hospital Medical Center (CCHMC) PED as well as supervising pediatric emergency medicine faculty and fellows to participate in this study. To be eligible, residents were required to have at least 4 weeks scheduled in the PED during the study period in an attempt to gather multiple encounters for each resident. Five of 105 eligible residents and 1 faculty member declined to participate. Additionally, 2 authors (D.J.S. and B.S.) did not participate. For participating residents, we obtained a supervisor entrustment decision and quality measures for encounters they had with patients presenting with 3 common, acute problems: acute asthma exacerbation, bronchiolitis, and closed head injury (CHI). These were chosen as representative diagnoses for the General Pediatrics EPA focused on providing care to patients presenting with common, acute problems.²⁷

Eligible encounters were considered to be those with a participating supervisor where an enrolled resident was the resident of record for the patient from the outset of the encounter to the point of disposition set for the patient in the electronic health record (EHR).

Encounters of patients from 0 to 17 years old were considered eligible for asthma and CHI. For asthma encounters, reactive airway disease in younger children that is treated the same as asthma in older children was included. For bronchiolitis, patients were eligible if not more than 12 months old.

Measures

Supervising pediatric emergency medicine faculty or fellows made entrustment decisions by completing an entrustment assessment based on Chen's supervision scale (question 1 in Supplemental Appendix 1).^{14,28} This supervision scale details supervision levels that include 2 gradations of full supervision, 3 gradations of partial supervision, no supervision, and the ability to supervise others.²⁸

To account for patient acuity and complexity, we asked supervisors to respond to 2 additional questions that rated these variables as standard, higher than usual, or lower than usual (questions 9 and 10 in Supplemental Appendix 1). Acuity is defined as "severity or how sick the patient is" in the question about acuity; whereas, complexity is defined as "challenging to understand and/or manage from a medical, psychosocial, and/or other perspective" in the question about complexity.

After completing the initial version of the questions on supervision level, patient acuity, and patient complexity for the entrustment assessment form, we conducted 15 cognitive interviews with faculty and fellow supervisors. Cognitive interviews seek to determine if individuals completing a questionnaire understand the questions as intended, allowing developers to make adaptations to ensure end-user understanding and promote intended use.^{29,30} Several changes were made as a result of these interviews. An example change was editing the questions about complexity and acuity to give examples of both to ensure that respondents would clearly understand the difference between these 2 factors. Once the questions had been finalized, faculty development sessions for question use were conducted the month prior to initiating the study and once monthly for 6 of the first 7 months of the study. Additionally, a badge card (i.e., card with details of the supervision scale that can be worn on a lanyard with an employee badge) and a tips sheet for question use were provided to all supervisors in the first few months of the study.

Quality measures were also gathered for each encounter. Each of the 3 illnesses of interest had a set of previously prioritized RSQMs that were used in this study (Table 1).³⁴ An RSQM composite score was calculated for each encounter by determining the proportion of RSQMs performed

Table 1: Resident-Sensitive Quality Measures for Asthma, Bronchiolitis, and Closed Head Injury, Cincinnati Children's Hospital Medical Center Pediatric Emergency Department, 2017–2018

Asthma Quality Measures	Bronchiolitis Quality Measures	Closed Head Injury Quality Measures
Use asthma order set	Follow bronchiolitis pathway	Closed head injury/ PECARN ^a best practice advisory use
Correct medication dose ordered for albuterol	Document birth history (preemie or not a preemie)	Appropriate closed head injury/PECARN* pathway use
Use of dexamethasone as steroid	Day of illness clearly documented	Mechanism of injury documented
Correct medication dose ordered for dexamethasone	Documentation of previous wheezing	Documentation of time of injury
Time from resident assigning self to patient to steroid order	Assessment of severity documented	Documentation of presence or absence of loss of — consciousness
Correct medication dose ordered for ipratropium	Effort of breathing documented	Documentation of presence or absence of emesis
Documentation of previous intubation or Bilevel Positive Airway Pressure for asthma	Documented quality of air entry (normal, decreased, etc.)	Documentation of whether back to baseline or not
Note the acuity of the patient in documentation	Documentation of wheezing	Documentation of presence or absence of other injury in the history
Documentation of work of breathing	Documentation of crackles	Assessment of severity documented
Documentation of aeration/air exchange	Documented presence or absence of subcostal retractions	Documentation of presence or absence of hematoma on physical exam (if present, location and size also documented)
Presence or absence of wheezing documented	Documented presence or absence of intercostal retractions	Thorough head exam (head, eyes, skull) documented
Ensure at least 3 descriptive words in respiratory exam	Documented presence or absence of suprasternal retractions	Documentation of Glasgow Coma Scale
Resident documents own Pediatric Respiratory Assessment Measure score	Oxygen saturation clearly documented in note	FULL neuro exam documented

Asthma Quality Measures	Bronchiolitis Quality Measures	Closed Head Injury Quality Measures
Resident-assigned Pediatric Respiratory Assessment Measure score matches resident-placed orders	Hydration status clearly documented in note	Documentation of presence or absence of other injury on physical exam
Document response to intervention	Documentation of response to specific therapeutics (i.e., how they responded to suctioning, to breathing treatment, to normal saline bolus, etc.)	Appropriate differential diagnosis and medical decision making documented
Documentation of disposition decision	Oral feeding tolerance clearly documented in note	Reassessments of patient documented
Use of standardized dosing for discharge medications—dexamethasone	Documentation of justification for appropriate disposition (home versus admit)	Return to school/play recommendations in discharge papers
Home dexamethasone instructions documented in written discharge instructions	Documentation of worsening respiratory symptoms as a reason to return written in discharge instructions	Use of standard/pre-made discharge instructions for diagnosis
State who to follow up with and provide information to contact them in discharge papers	Use standard/pre-made discharge instructions for bronchiolitis	Appropriate follow-up (sports medicine, rehabilitation medicine, neurology, etc.) recommended
Documentation of needing albuterol more often than every 4 hours as a reason to return in written discharge instructions	Documentation of poor feeding as a reason to return in written discharge instructions	
Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	Bulb suction teaching for home ordered	
	State who to follow up with and provide information to contact them in discharge papers	
	State appropriate number of days to follow-up in discharge papers	

^aPECARN is the Pediatric Emergency Care Applied Research Network, which conducted a study that delineated a decision rule for managing closed head injury (Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: A prospective cohort study. *Lancet*. 2009;374(9696):1160–1170).

divided by the total possible RSQMs that could have been performed for each encounter. This approach was taken due to the varied number of RSQMs within and between the 3 conditions. While most RSQMs applied in all encounters, some were not applicable based on the clinical context (e.g., some measures applied only to patients discharged home and not admitted). In addition, the number of RSQMs for each condition varied.

Data Collection

When research coordinators identified eligible encounters, they approached supervisors to complete the entrustment assessment form once a disposition and diagnosis had been set for the patient. Thus, we enrolled a convenience sample of encounters for 83 eligible residents when research coordinators were present in the PED (16 hours a day for 6 days each week and 8 hours a day for the final day of the week). To ensure accuracy, two individuals (clinical research coordinator and lead investigator) worked independently to manually extract RSQMs from the EHR following the encounter.

We compensated supervisors \$5 for each form completed. The CCHMC Institutional Review Board reviewed this study and determined it to be exempt.

Data Analysis

Descriptive statistics were generated for the entrustment decision and the RSQM composite scores for each condition. To measure the association of the entrustment decisions with RSQM composite scores, we used a mixed models approach to account for observations nested within residents.³¹ We also developed a model for each condition with the RSQM composite score as the dependent variable and the entrustment decision as the independent variable of interest. This was due to the nonnormal distribution of the entrustment decisions. We adjusted for patient acuity, patient complexity, and resident postgraduate year (PGY). Similarly, we combined the RSQM composite scores for the 3 conditions and developed a model adding diagnosis as a covariate. In addition, we dichotomized the measure of supervision by including “unsupervised practice” and “ability to supervise others” in one category and all other levels in a second category and conducted the same analysis with the dichotomized measure as the independent variable of interest.

We also used a mixed models approach to measure the association of RSQM composite scores with patient acuity and with patient complexity. First, we developed models for each condition with the RSQM composite score as the dependent variable and acuity as the independent variable of interest, controlling for PGY. We next conducted the same analysis for complexity, again controlling for PGY. Finally, we combined the RSQM composite scores for the 3 conditions and repeated the analyses with diagnosis as an additional covariate.

Results

A total of 83 residents cared for 110 patients presenting with asthma (59 unique residents), 112 patients presenting with bronchiolitis (64 unique residents), and 77 patients presenting with CHI (46 unique residents). Table 2 details descriptive statistics for the entrustment decisions and RSQM composite scores.

Table 2: Descriptive Statistics for Supervisor Entrustment Decision Rating and Pediatric Resident-Sensitive Quality Measure Composite Scores for Asthma, Bronchiolitis, Closed Head Injury, and All Conditions Combined, Cincinnati Children’s Hospital Medical Center Pediatric Emergency Department, 2017-2018^a

	Median (IQR)	Range
Entrustment Decision Rating		
Asthma	5 (4, 6)	1 to 7
Bronchiolitis	5 (4, 6)	1 to 7
Closed Head Injury	4 (4, 6)	1 to 7
Conditions Combined ^b	5 (4, 6)	1 to 7
Resident-Sensitive Quality Measure Composite Score		
Asthma	0.81 (0.74, 0.88)	0.47 to 1
Bronchiolitis	0.61 (0.53, 0.71)	0.35 to 0.91
Closed Head Injury	0.63 (0.56, 0.68)	0.44 to 0.89

Abbreviation: IQR indicates interquartile range.

^aAdjusted for postgraduate year, patient complexity, and patient acuity.

^bCombined conditions adjusted by diagnosis.

Association Between RSQM Composite Scores and Entrustment Decisions

As Table 3 shows, entrustment decisions exhibited a statistically significant positive linear relationship with asthma RSQM composite scores ($P = .0004$) and all conditions combined ($P = .006$). For every 1-point increase in entrustment decision scores, RSQM composite scores for asthma increased by 0.03 (out of 1), and scores for all conditions combined increased by 0.014. The data showed no significant association between RSQM composite scores and entrustment decisions for bronchiolitis or CHI.

Table 3: Association Between Pediatric Resident-Sensitive Quality Measure Composite Score and Supervisor Entrustment Decision Rating for Asthma, Bronchiolitis, Closed Head Injury, and All Conditions Combined, Cincinnati Children's Hospital Medical Center Pediatric Emergency Department, 2017-2018^a

Condition	Beta Coefficient ^b	95% CI	P Value
Asthma	0.030	(0.014, 0.046)	.0004^d
Bronchiolitis	0.0004	(-0.0161, 0.0169)	.96
Closed Head Injury	0.012	(-0.006, 0.031)	.19
Conditions Combined ^c	0.014	(0.004, 0.023)	.006^d

^a Parameter estimates are adjusted by postgraduate year, patient complexity, and patient acuity.

^b The change in RSQM composite score associated with a one unit change in entrustment.

^c Combined conditions are adjusted by diagnosis.

^d Significant at $P < .01$.

As Table 4 shows, when entrustment decisions are dichotomized to encounters where residents are entrusted with unsupervised practice or the ability to supervise others (the highest 2 levels of the supervision scale) compared with being assigned any supervision level short of entrustment with unsupervised practice (the lowest 5 levels of supervision), there are statistically significant differences in RSQM composite scores between groups for asthma ($P = .01$), CHI ($P = .04$), and all conditions combined ($P = .02$). Bronchiolitis, which did not have significantly different RSQM composite scores between groups, had the lowest average RSQM composite score for residents entrusted with unsupervised practice or the ability to supervise others (0.61, 95% CI: 0.55–0.66) and the second lowest average RSQM composite score for residents entrusted below unsupervised practice (0.62, 95% CI: 0.56–0.67).

Table 4: Difference in Resident-Sensitive Quality Measure (RSQM) Composite Score for Pediatric Residents Between Those Entrusted With at Least Unsupervised Practice Compared to Entrustment Below This Level for Asthma, Bronchiolitis, Closed Head Injury, and All Conditions Combined, Cincinnati Children's Hospital Medical Center Pediatric Emergency Department, 2017-2018^a

Condition	Average RSQM Composite Score for Entrustment with Unsupervised Practice or Ability to Supervise Others (95% CI)	Average RSQM Composite Score When Assigned a Lower Supervision Level Than Entrusted with Unsupervised Practice or Ability to Supervise Others (95% CI)	P Value
Asthma	0.83 (0.78-0.87)	0.76 (0.73-0.80)	.01^c
Bronchiolitis	0.61 (0.55-0.66)	0.62 (0.56-0.67)	.65
Closed Head Injury	0.63 (0.58-0.69)	0.58 (0.53-0.62)	.04
Conditions Combined^b	0.70 (0.67-0.73)	0.66 (0.64-0.69)	.02

^a Adjusted for postgraduate year, patient complexity, and patient acuity.

^b Combined conditions are adjusted by diagnosis.

^c Significant at $P \leq .01$.

Association Between RSQM Composite Scores, Patient Acuity, and Patient Complexity

As Figure 1 shows, when controlling for PGY, RSQM composite scores were significantly higher when acuity was also higher and significantly lower when acuity was also lower for both asthma ($P = .0009$) and bronchiolitis ($P = .01$). However, RSQM composite scores were almost identical between levels of acuity for CHI ($P = .94$). Combining illnesses and controlling for PGY and illness, RSQM composite scores were significantly higher when acuity was also higher and significantly lower when acuity was also lower ($P = .0005$).

No significant differences were seen in RSQM composite scores when complexity varied for any individual illness or all illnesses combined (Figure 1).

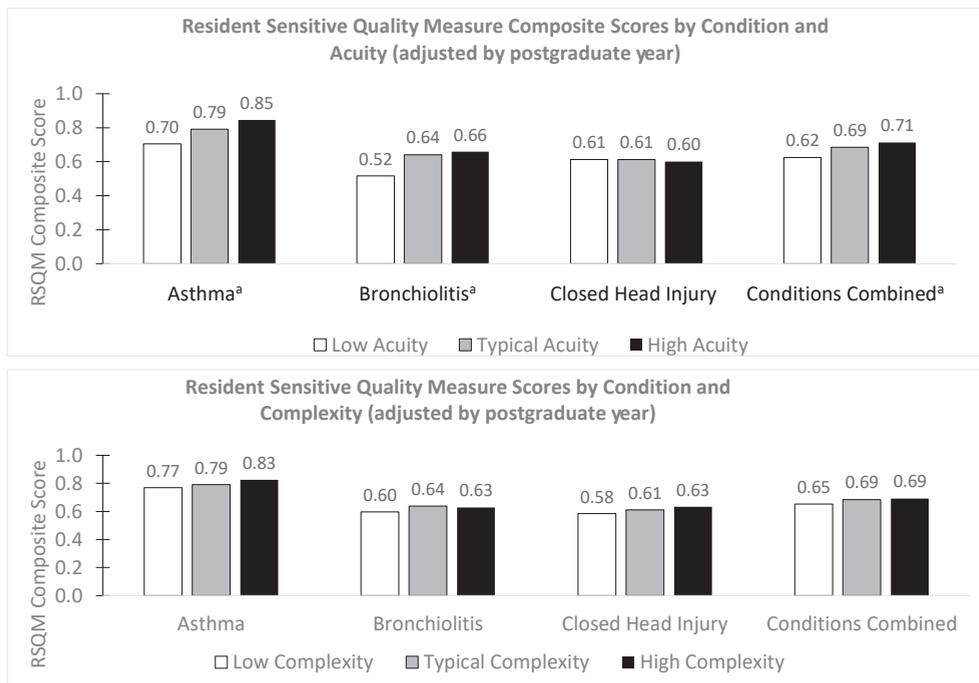


Figure 1: Pediatric resident-sensitive quality measure (RSQM) composite score by condition (asthma, bronchiolitis, closed head injury, and all conditions combined) and acuity and complexity, Cincinnati Children's Hospital Medical Center Pediatric Emergency Department, 2017–2018 (^asignificant difference at $P < .05$).

Discussion

This study takes an important step in pursuing the link between educational outcomes and patient care outcomes through understanding the associations between RSQM composite scores and supervisor entrustment decisions, patient acuity, and patient complexity.²⁶

Association Between RSQM Composite Scores and Entrustment Decisions

We found a statistically significant, but modest positive relationship between entrustment decisions and RSQM composite scores for asthma. However, asthma RSQM composite scores only increased 0.03 (out of 1) for every 1 (out of 7) point increase in entrustment decision scores. While statistically significant, this change is likely not educationally meaningful. We also found a statistically significant positive relationship between entrustment decisions and RSQM composite scores for all conditions combined, but this phenomenon seems to be driven by the asthma data because we did not find significant relationships between RSQM composite scores and entrustment decisions for the other 2 diagnoses we considered.

RSQM composite scores were significantly higher for residents entrusted with unsupervised practice or the ability to supervise others than for residents entrusted to lower supervision levels for all conditions (asthma, CHI, and all conditions combined) other than bronchiolitis. This finding offers validity evidence for RSQM composite scores; however, the absolute difference is modest.

Laying a Foundation for Future Work: RSQMs as Assessment Data?

While we would expect a positive relationship between RSQM composite scores and entrustment decisions across all entrustment decisions and conditions considered, our findings do not demonstrate this. It may be that such a relationship currently does not exist. In this case, one of two considerations may be important to consider. First, RSQMs may be better suited than entrustment decisions to assess resident performance in a manner that matters for patients. Second, a combination of RSQMs as well as entrustment or other assessment decisions around aspects of care not measured by RSQMs may be needed. Unlike entrustment decisions, which are intended to be holistic decisions based on the integration of competencies to deliver care, RSQMs only consider circumscribed tasks thought to be linked with care quality. They do not reflect more complex aspects of care, such as clinical reasoning, thoroughness, and dependability.

At the outset of this paper, we noted that entrustment should *theoretically* place focus on learners as well as on the outcome of care that patients receive. However, assessment efforts using an entrustment framework have been shown to suffer from the same “frame-of-reference” problems that other assessment efforts face. Namely, supervisors use self as the frame of reference in determining what level of supervision the learner needs rather than what the patient needs to receive safe, effective, patient-centered care.³² The “self” that is used for comparison could be what the assessor is capable of now, what the assessor was capable of at the same level of training as the learner being assessed, or some other reference. If entrustment decisions do not align with the delivery of expert-informed optimal care, which this study calls into question, other measures to determine whether patients receive such care must be sought. With the noted

caveat that RSQMs alone may not be sufficient for assessing resident performance, the range of RSQM composite scores suggests their potential promise for better discerning performance than traditional assessment efforts that suffer from more range restriction. The significantly higher RSQM composite scores for asthma, CHI, and all conditions combined for residents entrusted with unsupervised practice or the ability to supervise others than residents entrusted to lower supervision levels provide additional support for their use as assessment tools given this expected relationship between these 2 variables. Furthermore, given their primary focus on the provision of optimal care to patients, RSQMs could perhaps be an important, or even primary, form of assessment for residents to inform entrustment decisions. This use could lead to entrustment decisions that are associated with measures of quality care, which is unquestionably the desired outcome.

Stated differently, the paradox of the entrustment framework, which is intended to place the patient at its core, is that care may worsen if trainees are given more trust than deserved. In this study, residents may be receiving higher than deserved entrustment decisions while patients are not receiving optimal care, based on RSQM composite scores. The question remains as to why this is the case. Is it because of the need for frame-of-reference training, as suggested above, or may it also call into question the supervisory skills of faculty and fellows if considering components of quality care is not part of their entrustment decisions?

Educational Value Despite Non-Educationally Significant Findings

In addition to potential utility as direct measures of resident performance, we believe RSQMs also have educational value despite their large lack of association with other variables we considered in this study. Residents report suboptimal training in and understanding of quality improvement as well as the perception that they do not contribute meaningfully to quality improvement efforts.^{33,34} As with other areas of suboptimal training, this finding may reflect the lack of faculty skills in competencies that were not a focus of medical training in the precompetency-based medical education era. Furthermore, many traditional quality measures are not aligned with the work residents complete, and few efforts have been tried to address this gap.^{1,2,35-37} We believe these circumstances send a strong message that we must embrace any efforts to provide residents with quality metric feedback about their performance and then subsequently determine if such feedback leads to improved RSQM performance. Feedback can be hard to come by in training, even in the advent of competency-based assessment that places focus on the importance of direct observation and real-time feedback.³⁸⁻⁴⁰ Therefore, even if RSQM composite scores do not have strong relationships with other measures of resident performance, they do capture whether residents are completing components of their work that have been deemed important to patient care by fellow residents and supervisors. Moreover, these measures demonstrate a wide range of performance (0.47–1 for asthma, 0.35–0.91 for bronchiolitis, and 0.44–0.89 for CHI). Almost no residents in this study attained all possible RSQMs in their patient encounters, and thus

RSQMs give residents feedback about care components they are not providing that have been deemed important to patient care.

There were higher RSQM composite scores for asthma than for the other 2 illnesses. This is likely because quality improvement efforts at the study site were targeting the care of patients with asthma in the PED during this study period. However, it is striking to note that these efforts still left large gaps in the asthma care provided by residents during the study.

Association Between RSQM Composite Scores, Patient Acuity, and Patient Complexity

For asthma and bronchiolitis, RSQM composite scores were higher when residents were caring for patients with higher acuity than when they were caring for patients with lower acuity. This finding is perhaps the opposite of what might be expected because higher acuity patients could lead to worse care given the additional cognitive burden they may elicit, and lower acuity patients may lead to more thorough or complete care given the lower cognitive burden they may elicit. However, our findings suggest that high acuity triggers better performance and more attention to doing the right things in patient care. Acuity may focus the mind by engaging alertness and attention to detail. Conversely, perhaps residents view lower acuity cases as opportunities to streamline their documentation and care in a manner that helps their overall efficiency while managing multiple patients but perhaps lacks the desired depth of care for these lower acuity patients. Future study should explore this area further as it is important to building further validity evidence for RSQMs and their composite scores.

While we found significant relationships between RSQM composite scores and patient acuity, we did not find significant relationships between RSQM composite scores and patient complexity. Acuity was defined as “severity or how sick the patient is” on the entrustment decision form whereas, complexity was defined as “challenging to understand and/or manage from a medical, psychosocial, and/or other perspective” on the entrustment decision form (Supplemental Appendix 1). It appears that the learner may be activated by patients presenting in a manner that is more severe or ill than other patients but not by patients who are more challenging to understand or manage.

Limitations

There are limitations to consider in this study. First, the manual chart extraction of RSQMs makes extraction errors possible. We sought to minimize errors by employing 2 individuals to perform data extraction. Manual extraction was chosen because some of the RSQMs cannot be automatically extracted from the EHR, limiting their usability beyond the study setting. However, advances in natural language processing may soon provide a solution to this problem. Second, while we completed cognitive interviews during development of assessment items to reduce unwanted variability in questionnaire responses, raters likely varied in their trust propensity and supervision style.^{15,16} These variations almost certainly led to variations in assigning supervision levels, which could have affected the association with RSQM composite scores. Third, for this study we constructed RSQM composite scores, which are novel and have not been the focus of previous validity work. Therefore, our findings should be considered with this in mind. However, it should be noted that the RSQMs were developed to have validity evidence built into them as they all met the criteria of being rated highly in terms of importance to care for the illnesses of interest by supervisors and residents working in the PED.^{3,4} As a result, we did not think it was necessary, or appropriate, to weight RSQMs differently in determining the RSQM composite score because all individual RSQMs for all illnesses were rated highly for importance to the illness of interest in consensus group methods prior to this study. Finally, in our analyses, we treat supervision level assignments, which are ordinal, as continuous data. However, this approach is commonplace, and psychometric scholars have noted that “parametric methods can be utilized [in these cases] without concern for ‘getting the wrong answer.’”⁴¹

Conclusion

The relationship between traditional assessment frameworks, such as entrustment, and care quality is critical but elusive. As authors of one of the first studies seeking to tackle this relationship, we did not succeed in taming this wicked problem. However, we found interesting relationships that must be considered in the current landscape. Entrustment is being increasingly adopted as an assessment framework, and harnessing the EHR to assess and evaluate care quality and patient outcomes as they relate to medical education is being increasingly called for as well.⁴²⁻⁴⁴ The concurrent juxtaposition of multiple, promising assessment advancements that place a focus on ensuring patients receive the care they need is a leap forward. Addressing population health needs is the foundation of competency-based medical education, which was first advocated more than 40 years ago.^{45,46} The question today is whether we are committed to determining how these new and emerging patient-centered assessment efforts interact with one another to ensure education outcomes meet patient outcomes. If we do not make this commitment, will patients and quality improvement scholars wait another 40 years for us?

References

1. 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.
2. Cook DA, West CP. Reconsidering the focus on “outcomes research” in medical education: A cautionary note. *Acad Med.* 2013;88:162–167.
3. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: A model from pediatric emergency medicine. *Acad Med.* 2018;93(7):1071–1078.
4. Schumacher DJ, Martini A, Holmboe E, et al. Developing resident-sensitive quality measures: Engaging stakeholders to inform next steps. *Acad Pediatr.* 2019;19(2):177–185.
5. ten Cate O, Scheele F. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice? *Acad Med.* 2007;82(6):542–547.
6. Hicks PJ, Margolis M, Poynter SE, et al. The pediatrics milestones assessment pilot: Development of workplace-based assessment content, instruments, and processes. *Acad Med.* 2016;91(5):701–9.
7. Schumacher DJ, Bartlett KW, Elliott SP, et al. Milestone ratings and supervisory role categorizations swim together, but is the water muddy? *Acad Pediatr.* 2019;19(2):144–151.
8. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: Outlining their usefulness for competency-based clinical assessment. *Acad Med.* 2016;91(2):186–190.
9. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014;89(5):721–727.
10. Carraccio C, Englander R, Holmboe ES, Kogan JR. Driving care quality: Aligning trainee assessment and supervision through practical application of entrustable professional activities, competencies, and milestones. *Acad Med.* 2016;91(2):199–203.
11. Messick S. Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice.* 1995;14(4):5–8.
12. 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(1):1.
13. Kennedy TJ, Regehr G, Baker GR, Lingard LA. ‘It’s a cultural expectation...’ The pressure on medical trainees to work independently in clinical practice. *Med Educ.* 2009;43(7):645–653.
14. Kennedy TJ, Regehr G, Baker GR, Lingard L. Point-of-care assessment of medical trainee competence for independent clinical work. *Acad Med.* 2008;83(10 Suppl):S89–92.
15. Gingerich A, Daniels V, Farrell L, Olsen SR, Kennedy T, Hatala R. Beyond hands-on and hands-off: Supervisory approaches and entrustment on the inpatient ward. *Med Educ.* 2018;52(10):1028–1040.
16. ten Cate O. Managing risks and benefits: Key issues in entrustment decisions. *Med Educ.* 2017;51(9):879–881.
17. Hauer KE, ten Cate O, Boscardin C, Irby DM, Iobst W, O’Sullivan PS. Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ Theory Pract.* 2014;19(3):435–456.

18. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes *JAMA*. 2009;302(12):1277–1283.
19. 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.
20. 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.
21. 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.
22. Phillips RL, Petterson SM, Bazemore AW, Wingrove P, Puffer JC. The effects of training institution practice costs, quality, and other characteristics on future practice. *Ann Fam Med*. 2017;15(2):140–148.
23. Denson JL, Jensen A, Saag HS, et al. Association between end-of-rotation resident transition in care and mortality among hospitalized patients. *JAMA*. 2016;316(21):2204–2213.
24. Lau BD, Streiff MB, Pronovost PJ, Haider AH, Efron DT, Haut ER. Attending physician performance measure scores and resident physicians' ordering practices. *JAMA Surg*. 2015;150(8):813–814.
25. 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;92(12):1740–1748.
26. Schumacher DJ, van der Vleuten CPM, Carraccio C. The future of high quality care depends on better assessment of physician performance. *JAMA Pediatr*. 2016;170(12):1131–1132.
27. Carraccio C, Englander R, Gilhooly J. General Pediatrics Entrustable Professional Activities. American Board of Pediatrics. Published 2015. <https://www.abp.org/entrustable-professional-activities-epas>. Accessed September 8, 2015.
28. Chen HC, van den Broek WS, ten Cate O. The case for use of entrustable professional activities in undergraduate medical education. *Acad Med*. 2015;90(4):431–436.
29. Willis GB, Artino AR. What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. *J Grad Med Educ*. 2013;5(3):353–356.
30. Willis GB. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications; 2004.
31. *Mixed Models Analyses Using the SAS System: Course Notes*. Cary, NC: SAS Institute; 2006.
32. Apramian T, Cristancho S, Sener A, Lingard L. How do thresholds of principle and preference influence surgeon assessments of learner performance? *Ann Surg*. 2018;268(2):385–390.
33. Butler JM, Anderson KA, Supiano MA, Weir CR. "It feels like a lot of extra work": Resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med*. 2017;92(7):984–990.
34. Schumacher DJ, Leslie LK, Van KD, Freed GL. Pediatric residents' reports of quality improvement training and experiences: Time for an improvement cycle? *Acad Pediatr*. 2019;19(4):399–403.
35. Swing SR, Schneider S, Bizovi K, et al. Using patient care quality measures to assess educational outcomes. *Acad Emerg Med*. 2007;14(5):463–473.

36. Haan CK, Edwards FH, Poole B, Godley M, Genuardi FJ, Zenni EA. A model to begin to use clinical outcomes in medical education. *Acad Med.* 2008;83(6):574–580.
37. Yin HS, Jay M, Maness L, Zabar S, Kalet A. Health literacy: An educationally sensitive patient outcome. *J Gen Intern Med.* 2015;30(9):1363–1368.
38. Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. *Med Teach.* 2013;35:569–574.
39. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: From Flexner to competencies. *Acad Med.* 2002;77(5):361–367.
40. Holmboe E. Realizing the promise of competency-based medical education. *Acad Med.* 2015;90(4):411–413.
41. Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ Theory Pract.* 2010;15(5):625–632.
42. Chahine S, Kulasegaram K, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Acad Med.* 2018;93(6):829–832.
43. Triola MM, Hawkins RE, Skochelak SE. The time is now: Using graduates' practice data to drive medical education reform. *Acad Med.* 2018;93(6):826–828.
44. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. *Acad Med.* 2018;93(7):975–978.
45. 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.
46. McGaghie WC, Miller GE, Sajid AW, Telder TV. Competency-based curriculum development on medical education: An introduction. *Public Health Pap.* 1978(68):11–91.

Supplemental Appendix 1: Faculty / Fellow Entrustment Decision Assessment Form

Record ID

Please answer the following questions about **[resident name]** for this encounter ONLY.

1. Based on the encounter you just had with this resident caring for the patient in room [room number] with a diagnosis of [diagnosis name], how much supervision do you think this resident needed?

Note: Please choose the response you believe corresponds with his/her abilities and not what you are required to do because they are still in training.

- Needs full, proactive supervision (*If Yes, Please answer 1a.*)
- Needs partial supervision (*If Yes, Please answer 1b.*)
- Did not need supervision (i.e., if the resident were not still in training, you think he/she would not have needed supervision.) (*Skip 1a & 1b, go to question 2.*)
- Did not need supervision and could have even been the supervisor for someone else caring for the patient (i.e., if the resident were not still in training, you think he/she could have supervised others) (*Skip 1a & 1b, go to question 2.*)

1a. Do you believe this resident:

- May jointly complete work with you
- May complete work on their own with you in the room ready to step in as needed

1b. Do you believe this resident requires:

- You to be immediately available and double checking ALL findings
- You to be immediately available and double checking what you consider to be KEY findings only
- You to be available only if needed

2. To what extent do you believe this resident possessed the necessary knowledge and skill to care for this patient?

- Had **all** the necessary knowledge and skill
- Had **almost all** the necessary knowledge and skill
- Had **most** of the necessary knowledge and skill
- Had **some** of the necessary knowledge and skill
- Had **little** of the necessary knowledge and skill

3. Did this resident ask for help when needed?

- Did not ask for help, but **I did not observe an area where help was needed**
- Asked for help with **all** components he/she needed help with
- Asked for help with **most** components he/she needed help with
- Asked for help with **some** components he/she needed help with
- Did not ask for help** when needed

4. To what extent was this resident thorough in presenting the information he/she gathered for this encounter?

- Presented **all** of the necessary pertinent information
- Presented **most** of the necessary pertinent information
- Presented **some** of the necessary pertinent information
- Presented **a limited amount** of the necessary pertinent information

5. How dependable was this resident in following through with tasks?

- Followed through on **all** tasks **without needing** prompting
- Followed through on **all** tasks **but needed** prompting
- Followed through on **most** tasks
- Followed through on **some** tasks
- Followed through on **few** tasks

6. Do you feel this resident ever neglected to give you accurate information in order to keep from disclosing that he/she did not do something that should have been done?

(e.g., not asking about immunizations but telling you they are up to date, not seeing the tympanic membrane but reporting he/she saw it, telling you he/she had paged a consultant that had not yet been paged)?

- Yes
- No

7. Consider a situation where the resident would have cared for this patient without a supervisor present. Please place the slider bar at the point that indicates how much you would have trusted this resident to provide safe, effective care for this patient without supervision:

Would not
have trusted
at all

Would have
completely
trusted



Please indicate a number (0-100) where you would place the pointer on the slider bar:

(Place a mark on the scale above.)

8. To ensure safe care for this patient, how much did you need to alter the plans and/or actions of the resident?

- I **did not need to alter** the plans/actions of the resident to ensure safe care.
- I needed to make **minimal changes** to the plans/actions of the resident to ensure safe care
- I needed to make **moderate changes** to the plans/actions of the resident to ensure safe care.
- I needed to make **significant changes** to the plans/actions of the resident to ensure safe care.
- I needed to essentially **develop the plan myself** to ensure safe care.

These final two questions ask you to think about the patient in the encounter you just had rather than the resident.

9. How would you classify the COMPLEXITY (i.e., challenging to understand and/or manage from a medical, psychosocial, and/or other perspective) of the patient's presentation for this encounter?

- More complex than usual for this illness
- Typical complexity for this illness
- Less complex than usual for this illness

10. How would you classify ACUITY? (i.e., severity or how sick the patient is) of the patient's presentation for this encounter?

- Higher acuity than typical presentation for this illness
- Typical acuity for this illness
- Lower acuity than typical presentation for this illness



Chapter 8

Contribution of Resident-Sensitive Quality Measure Data to Making an Entrustment Decision in a Pediatric Residency



Schumacher DJ, Martini A, Sobolewski B, Carraccio C, Holmboe E, Busari J, Poynter S, van der Vleuten C, Lingard L. "Contribution of Resident-Sensitive Quality Measure Data to Making an Entrustment Decision in a Pediatric Residency."

Academic Medicine 2020. In press.

Abstract

Purpose

Explore how individual clinical competency committee (CCC) members interpret, use, and prioritize resident-sensitive quality measure (RSQM) data added to their usual review processes. RSQMs are characterized by (1) their importance to care quality and (2) the likelihood they are performed by residents.

Method

In this constructivist grounded theory study, 19 members of the pediatric residency CCC at Cincinnati Children's Hospital Medical Center were purposively and theoretically sampled. Participants were provided a resident assessment portfolio; asked to review performance ratings and narrative comments for 5 rotations, along with RSQM data for one of these rotations; and make a decision about the resident's ability to care for patients presenting with common, acute problems (a general pediatric entrustable professional activity). Data collection consisted of 2 phases: (1) observation and think out loud while participants reviewed the portfolio and (2) semistructured interviews to probe participant reviews. Analysis moved from close reading to coding and theme development, followed by the creation of a model illustrating theme interaction.

Results

Five dimensions for how participants view and use RSQMs were identified: (1) ability to orient to RSQMs: confusing to self-explanatory, (2) propensity to use RSQMs: reluctant to enthusiastic, (3) RSQM interpretation: requires contextualization to self-evident, (4) RSQMs for assessment decisions: not sticky to sticky, and (5) expectations for residents: potentially unfair to fair to use RSQMs. The interactions among dimensions generated 3 RSQM data user profiles: eager incorporation, willing incorporation, and disinclined incorporation.

Conclusions

Most participants used RSQMs to varying extents, supporting their inclusion as resident assessment data for CCC review.

Contemporary clinical competency committees (CCCs) in graduate medical education have a foundational task of reviewing resident assessment data.¹⁻³ Literature to date on CCCs has largely focused on 2 areas: (1) detailing how to design a CCC⁴⁻⁹ and (2) potential best practices based on expert opinion, evidence from fields other than medicine, and personal experience at individual residency programs.^{7,10-20} However, extending research beyond structure and process to seek an understanding of how CCCs view, use, and prioritize different types of assessment data in their decision making is needed.²⁰ This additional information can help inform the types of assessment data that are provided to CCCs. Preliminary work in this area suggests that individuals prioritize qualitative assessment data.^{21,22} Furthermore, data that are difficult to interpret can lead to lengthy, labored discussions in an effort to understand them.²³ However, to our knowledge, no study has considered the relative value that CCC members ascribe to different types of assessment data.

Health care quality measures are rarely used in assessment, despite recent calls and efforts to determine how quality measures can be used in graduate medical education.²⁴⁻²⁷ Such measures that are sensitive to the work that residents complete not only enhance meaningful involvement of residents in quality improvement work but can also serve as a patient-centered assessment approach, indicating the quality of care provided by residents.²⁸ To this end, Warm and Mathis have suggested that Accreditation Council for Graduate Medical Education requirements for residency training should include measuring and improving patient care outcomes.²⁹

Furthermore, Jaffe and colleagues have advocated educators partnering with institutional quality improvement experts to design resident quality and safety reports.³⁰ However, many traditional quality metrics may not align well with the work that residents perform.³¹ Paradoxically, quality measures sometimes attributed to supervising physicians may be attributable to residents as well.³² These challenges led our research team to develop resident-sensitive quality measures (RSQMs), which are quality measures that meet 2 criteria: (1) likely performed by an individual resident and not another member of the team or the team collectively and (2) important to the care of a patient with a given illness of interest.^{27,33}

While initial implementation evidence for RSQMs has been established, how these measures may be used for assessment purposes has not been examined. To probe this area, we explored how individual CCC members interpret, use, and prioritize RSQMs alongside traditional assessment data when making a summative entrustment decision.

Method

Design

We conducted a constructivist grounded theory study to explore how CCC members interpret, use, and prioritize RSQM data when these data are inserted into their usual review processes. We were interested in the views, values, and perspectives of CCC members, aligning with a constructivist paradigm. Furthermore, we anticipated that a theory of how CCC members interpret, use, and prioritize quality measure data in entrustment decision making can inform next steps in how such information can or should be used for entrustment decision making among CCCs.

We began by purposively sampling CCC members from the pediatric residency at Cincinnati Children's Hospital Medical Center (CCHMC) to ensure a representative range of roles in the program (e.g., associate program directors, faculty without formal leadership roles in the program), gender, generalist versus subspecialist, and program leadership perceived leniency or stringency in making assessment decisions. The CCHMC pediatric residency CCC has 32 members (22 directors of resident rotations and 10 program leaders, including chief residents).

This study focused on individual CCC members. Many CCCs in pediatrics have each member review portfolios of resident assessment data and bring a summary of findings along with a proposed advancement or entrustment decision to a larger group of CCC members for vetting. The recommendations of individual CCC members may often stand. In some programs, there is little or no additional vetting of these initial decisions for residents without performance concerns. Given these contextual considerations, individual CCC members who review decisions play a significant role in summative assessment decisions regarding pediatric residents. Therefore, we focused on individual CCC members.

After redacting name, gender, and training year, we provided participants with a real resident portfolio that contained the typical assessment data CCC members review before a CCC meeting. The resident presented to them was, by the judgment of the research team, a typically developing resident in the middle of training who was generally performing well but had some concerns noted, a few of which were consistent across rotations and the remainder isolated within specific rotations. A synopsis of the resident portfolio is presented in Supplemental Appendix 1. In addition to milestones and/or entrustable professional activity (EPA) ratings as well as narrative comments for the 5 rotations included, participants were provided RSQM performance data for 3 diagnoses in the pediatric emergency department (asthma, bronchiolitis, and closed head injury). These novel data, shown in Appendix 1, are a product of our previous study^{27,33} and reflect the average performance data on these measures.^{34,35}

Data Collection

We held sessions with 19 participants between February and July 2019. Data collection consisted of 2 phases: (1) observation and think out loud, in which participants reviewed the resident portfolio, and (2) semistructured interviews to probe participant reviews. The first phase lasted approximately 7.5 to 20 minutes and the second phase 8.5 to 21 minutes across participant sessions. The entirety of the sessions (both phases) was audiorecorded and professionally transcribed.

Phase 1: Assessment Data Review

At the outset of each participant session, CCC members were provided the resident portfolio and instructed to review as they normally would, verbalizing their thoughts, ideas, and decisions. We told them we were seeking to understand how they interpret, use, and prioritize assessment data, without specific mention of RSQM data. The decision to withhold this information was made in an effort to avoid participant reactivity,³⁶ that is, participants behaving other than they normally would—in this case, unduly focusing on RSQMs due to our interests. To provide focus, we asked participants to review the resident portfolio with the goal of providing a summative entrustment decision for the general pediatrics EPA “manage patients with acute, common diagnoses.”³⁷ We chose this EPA because of its relevance to the RSQMs presented.

Two authors (D.J.S. or A.M.) conducted the sessions and recorded field notes to capture nonverbal nuances, review times, and key thoughts from participants.

Phase 2: Brief Semistructured Interview

Once participants had completed their review, we conducted a brief, semistructured interview to probe how they viewed, used, and prioritized the RSQMs in their review and decisions. We also probed how they did or did not differentially use other assessment data to mask our focused interest on RSQMs and thus minimize response bias. Data collection and analysis were iterative. Once initial themes were identified, theoretical sampling was used in subsequent sessions to confirm, disconfirm, and elaborate on the evolving results.

The CCHMC Institutional Review Board determined this study to be exempt.

Data Analysis

Constructivist grounded theory³⁸ posits that meaning is developed through the interaction of participants and researchers. Therefore, we believed that an analytic team of varied experience with resident education and performance review was important. Our team consisted of clinicians and nonclinicians, current and former residency program leadership, current and former CCC members, a former CCC chair, and education researchers with and without previous scholarly work in the area of CCCs, some of whom have held leadership roles in medical education. Two

members of the analytic team were current CCHMC residency program leaders at the time of the study (S.P., program director; B.S., assistant program director) and members of the CCHMC pediatric residency CCC. These individuals helped to ground our analysis in the context of local nuances when applicable. Our team composition allowed us to negotiate varied perspectives, placing us in a position to recognize the influence of our individual perspectives; to benefit from productive debate when we viewed things differently; and to develop a richer set of insights into the studied phenomenon through this dialectic, co-constructive process.

Dedoose, Version 8.2.14 (SocioCultural Research Consultants, LLC: Los Angeles, CA) was used to aid analysis. Analysis began with individual, close reading of 2 transcripts from participant sessions by 3 members of the research team (D.J.S., A.M., L.L.) and then discussion of initial codes that were used to analyze another 4 participant sessions. The group then met to resolve differences, identify early themes, and determine initial strategies for theoretical sampling. Following the analysis of an additional 7 participant sessions, we coalesced the findings into a preliminary model through which participants would engage with the RSQM data. When presented to the full research team, we identified additional issues for theoretical sampling. One such issue was the influence of benchmarks on participants' interactions with the RSQM data. Expected performance for the RSQMs had not been included in the data for the first 16 participant sessions; to theoretically sample this issue, we provided a benchmark of 80% for all measures in the resident portfolio for the final 3 participant sessions. D.J.S., A.M., and L.L. continued iterative data collection and analysis in an effort to fully characterize the identified dimensions as well as to understand the relationships among them. The latter effort resulted in the articulation of 3 recurring profiles illustrating how the dimensions interact. As our results approached theoretical sufficiency, the analysis was again discussed with the full research team.

As analysis proceeded following the initial participant sessions, transcripts from subsequent sessions were used to challenge initial codes and themes. We also repeatedly returned to previous sessions' transcripts in analysis, ensuring constant comparison.

Results

Of the 19 participants, 8 were men and 11 women. Nine were pediatric subspecialists (from separate subspecialties); the remainder were general pediatricians, hospitalists, and chief residents. Some were in their first year serving as CCC members, and others had served for multiple years.

Participants tended to view and use all types of data (narrative comments, milestone and EPA ratings, and RSQMs) substantively in their decision making. However, they demonstrated

preferences for narrative comments, faculty ratings, and sometimes data from particular rotations (continuity clinic, emergency medicine, and hospital medicine).

Engagement With the RSQMs

We identified 5 dimensions for how participants view and use RSQMs. We classified these as dimensions because this term has a notion of measurable extent built into it, and each of the dimensions has a spectrum.

Ability to Orient to RSQMs: Confusing to Self-Explanatory

While most participants were able to correctly and quickly orient themselves to the RSQMs in an efficient manner, one participant did note: “Well, I barely understand that paper [with the quality measures on it], so I’m probably not going to use it very much, to be honest with you” (Participant 9). Some noted they had not seen such data in their previous reviews, but they were still able to interpret the RSQMs after some time for processing: “Oh man, how do I read this? [7 second pause] Okay, these are like the metrics of [whether] they followed certain guidelines of practice. Okay” (Participant 15).

Propensity to Use RSQMs: Reluctant to Enthusiastic

Some participants were reluctant to use the RSQMs. One barrier to using them was the volume of data they comprise: “It’s just a lot of information on one page for me to parse through. . .and try to understand. I think if I was more familiar with it, I would probably find it more helpful” (Participant 9). Another source of reluctance arose from viewing the RSQMs as simply “checking a box” of giving quality measure feedback to residents. In this case, the RSQMs were not seen as providing sufficient value to use for assessment purposes. A final source of reluctance arose from viewing the RSQMs as too specific for one setting (emergency department), creating a barrier for giving weight to the quality measure data.

While some participants expressed varying degrees of reluctance, others displayed variable levels of enthusiasm in using RSQMs, feeling they were helpful in their review and noting qualities such as RSQM data being objective and providing an overall picture of performance: “I think the quality data is really helpful because I think it’s just very specific to say in these types of patients this is this resident’s typical practice” (Participant 7).

RSQM Interpretation: Requires Contextualization to Self-Evident

Some participants felt that properly interpreting the RSQMs required more details about the context in which they were collected, such as having some assurances that a measure should have been met based on the clinical context and understanding a resident’s reason for not meeting it.

When I'm looking through the different measures, I want more context, I guess. . . . Follows bronchiolitis pathways—6 times they did this, 3 times they did not. . . .but I don't know why. . . . If that was the correct thing to do, why didn't they [do it]? Or what was their thinking? (Participant 6)

Having a peer comparison or benchmark for how often RSQMs should be met was also noted as an important piece of additional context for interpreting RSQM performance:

Did not seem like this person consistently hit all those measures. But, I'm not sure that I know how this stacks up to someone else in this comparable situation and, so, I had a hard time putting them into perspective about whether or not they did well or not well. (Participant 14)

However, while some participants thought a benchmark would help interpret RSQMs, only some of the participants in the sessions in which we presented a benchmark thought it was helpful or necessary.

Some participants commented that not being an emergency medicine physician who understands the importance of some of the RSQMs was a contextual issue for interpreting RSQM performance. For example, a nephrologist noted: "This is not something that you encounter on the regular floor. So, this is unique for emergency room[s]. . . . I don't know what is expected" (Participant 16). Additionally, an endocrinologist said: "If [diabetic ketoacidosis] were on here, I probably would have perked up" (Participant 19).

Documentation of RSQMs was another area where participants sometimes wanted more context. Some participants were inclined to give residents the benefit of the doubt; that is, they believed that even though the resident did not document something, he or she likely had done it. This stance was rationalized by the idea that knowing to do something and documenting doing it may be 2 different skills to learn, and perhaps a resident needed to learn the latter rather than the former. The trend in the data was for participants with this view to discard the RSQMs focused on documentation to avoid this influence on the performance picture they were developing of the resident.

The quality measures paint a very loose picture of [the following]: This resident did some things by the book and did some things that weren't quite by the book. But that doesn't necessarily mean that they made bad decisions. They just didn't have documented stuff. So that didn't mean as much to me. (Participant 10)

Giving the benefit of the doubt to oneself with quality measures was also noted as a reason for giving the same benefit of the doubt to residents.

However, other participants did not give residents the benefit of the doubt on these measures; they believed that if the action was not documented then the resident had not done it. In other words, these measures are self-evident.

I would say (a) this person either needs to work on their documentation or (b) just needs to work on their history and physical taking. Which, again being objective, it probably has to be the latter versus the former. (Participant 4)

Some of the participants who considered the RSQMs, as a whole, as objective data also viewed the RSQMs as self-evident. One participant noted: "This [set of RSQMs] is very helpful because it tells exactly what they [the residents] do or do not do" (Participant 7). Sometimes this sense of objectivity was related to the perception that RSQMs arose from care guidelines, which seemed to give additional weight to those measures: "I think that's very useful to say that you don't meet those guidelines [represented in the RSQMs]. There's nothing more objective than that" (Participant 12).

RSQMs for Assessment Decisions: Not Sticky to Sticky

Some participants used the RSQMs in the summative assessment task they were assigned similarly to how they used the other assessment data they were reviewing. For these participants, the RSQMs may have been helpful for them, but they were not "sticky," meaning the data did not stand out, were not prioritized, or both, when they made their entrustment decisions. For other participants, the RSQMs were sticky, most frequently when the participants felt the RSQMs represented deficiencies. Some of these performance gaps concerned areas that participants viewed as care fundamentals.

Let me look at bronchiolitis here. . . . Interesting. . . as I go more down this list, I see lots of things that I would not have expected. . . a bulb suction order for home [not done], . . . didn't do documentation, poor feeding as a reason [to return] not in the discharge instructions. I almost wonder if this is a resident that can get a little easily overwhelmed, forgetting to kind of do some of the basic stuff when they get busy. (Participant 1)

However, some of these deficiencies were viewed as more significant gaps in providing safe, appropriate patient care: "I have serious concerns about the safety of the patient because there's not prompt treatment, there's not a good review of when to return to the ED [emergency department], and, again, clinical assessment; it's just not there" (Participant 12).

Expectations for Residents: Potentially Unfair to Fair to Use RSQMs

Participants raised concerns about the potential fairness of using some of the measures to assess residents. An example is the specificity of the measures; some participants did not see the importance of holding residents to the level of detail included in some of the RSQMs, such as documenting discharge instructions in addition to verbally discussing them with families. In these instances, participants may not include that level of detail in their own work or might not document something they know another provider is expected to document in the chart (e.g., the Pediatric Respiratory Assessment Measure score in patients presenting with an asthma exacerbation, which is also documented by the respiratory therapist). Concerns were also raised about whether residents knew these measures were expected of them. If not, participants felt it might be inappropriate to hold residents to meeting measures they were unaware of.

I would never say ignore the measure. I would say, "Are we doing enough to make sure that it is easy to meet this measure?". . . I would say, "Maybe nobody ever taught you that, but you should be doing it anyway. The reason this is here is because this is good practice. The fact that you didn't know [what] it means, maybe it's my fault we didn't teach you." (Participant 11)

Looking beyond the resident, some participants felt that the RSQMs included measures that should not be expected of individual residents and should instead be automated in the electronic medical record. For example, documenting important parts of the history or physical exam could become components of standardized notes, and some of the measures pertaining to discharge instructions could become part of standardized discharge instructions.

While concerns were raised with certain aspects of using the RSQMs for performance expectations, participants largely felt that using the RSQMs among residents was both useful and appropriate. Beyond using them for assessment purposes, participants noted the utility of RSQMs for providing formative feedback to residents about areas for improvement.

Profiles for Using RSQMs to Make Entrustment Decisions

The dimensions identified in this study interact with one another. When considering these interactions, we theorized that 3 profiles regarding use of RSQM data can be defined: eager incorporation, willing incorporation, and disinclined incorporation.

Eager Incorporation

Eager incorporation was reflected when an individual's initial actions were to use the RSQMs in their entrustment decisions. In this profile, there was an enthusiastic propensity to use RSQMs, and RSQMs were seen as self-explanatory. Individuals displaying this profile felt that RSQMs sometimes needed additional context to interpret and sometimes were self-evident, but neither

stance precluded them from being eager adopters. Finally, individuals displaying this profile tended to find RSQMs to be sticky and did not raise major concerns about the fairness of their use.

Willing Incorporation

Willing incorporation was demonstrated when participants did not describe the role that RSQMs played in their entrustment decision, but when asked specifically about the RSQMs subsequently, they described their willingness to use RSQMs and even the perceived utility in using them. Similar to the first profile, participants displaying this profile found the RSQMs to be self-explanatory and had some propensity to use them. They sometimes felt the RSQMs required additional contextualization to interpret and sometimes felt the RSQMs were self-evident. They also sometimes found the RSQMs sticky for summative assessment and other times did not. Finally, individuals displaying this profile raised concerns about the RSQMs reflecting both fair and unfair expectations for residents.

Disinclined Incorporation

Individuals with a disinclined profile did not use the RSQMs when making a summative entrustment decision in the interview, even when specifically probed about the RSQMs. However, they could envision using different RSQMs in the future. Some individuals displaying a disinclined profile noted the RSQMs were confusing. They were also likely to raise concerns about the fairness of holding residents accountable for meeting RSQMs. Finally, participants displaying a disinclined profile were reluctant to use RSQMs and did not find the RSQMs to be sticky.

One participant tended toward the more highly disinclined end of this category. This person's responses, while close to creating a fourth profile of "dismissive," did not reach this level.

Discussion

This study sought to understand how CCC members view, use, and prioritize RSQMs alongside other assessment data when making an entrustment decision. Participants often gave some types of assessment data more weight than others, such as prioritizing faculty data over peer data and viewing narrative comments as the most useful data they reviewed. These findings corroborate previous research.^{22,39} However, participants tended to view and use all types of data, including RSQMs, substantively in their decision making. We found a range of engagement with RSQMs in 5 dimensions. While participants gravitated toward the end of the spectra where dimensions regarding RSQMs were viewed more positively, as evidenced by the eager or willing incorporation profiles being more common than the disinclined, examples from the other end of the spectra were present for all dimensions. The defined profiles have similarities to Rogers'

stages of adoption, with eager incorporation aligning with early adopters, willing incorporation corresponding to early majority, and disinclined incorporation perhaps corresponding to late adopter.⁴⁰ These findings present important implications for the potential future use of RSQMs because this study did not place focus on helping participants become comfortable with RSQMs, but participants still tended toward using them.

While participants largely used the RSQMs, their comments have implications for improving the use of the measures in the future. First, participants' discussions about whether or not residents should be documenting some of the information represented in several RSQMs warrant exploration. RSQMs were developed and prioritized through consensus group methods (nominal group technique and Delphi) to meet the criteria of both being done by a resident and being important to care for patients with a given illness of interest.^{27,33} In light of some of the issues raised by participants, perhaps the consensus group results should be revisited. Finding that some participants, all of whom are CCC members, believed certain RSQMs might not be important could be indicative of the need to orient and train CCC members on RSQMs if these measures are to be used for assessment. We intentionally did not include such orientation and training in this study. The profiles developed in this study may help program leaders tailor their training efforts for the varying ways CCC members tend to incorporate RSQMs in their resident data reviews. Such training may also address some of the challenges raised by participants who were unfamiliar with measures that would be important in the pediatric emergency department. Finally, our participants described giving residents the same benefit of the doubt they give themselves for the documentation measures, aligning with the findings of Apramian and colleagues,⁴¹ who showed that surgeons' individual procedural preferences may be the largest factor driving intraoperative assessment of residents.

Some participants noted that the RSQM data felt overwhelming. This problem can be addressed by developing a visual presentation of the RSQMs that minimizes the cognitive load the reviewers experience. Another consideration is whether presenting performance on all measures is important or if presenting a composite score for all measures and/or subscores for types of measures (e.g., documentation, appropriate medications) is sufficient.

Future Directions

Some participants were reluctant to use RSQMs because they viewed them as too specific to the emergency department, and they did not want to give undue weight to that setting in their entrustment decisions. Not using RSQMs because more RSQMs do not exist actually suggests how useful these measures could be and underscores the value of and need for developing RSQMs for other settings moving forward. As this work is carried out, it will be important to explore differences between views of RSQMs among pediatric faculty participants at CCHMC and those of pediatric faculty at other institutions. Furthermore, it will be important to ascertain

how views on using RSQMs differ between faculty in pediatrics, other medical specialties, and surgical specialties.

The findings of this study give insight into information to include in CCC member faculty development regarding RSQMs. We suggest the following:

- A discussion of how CCC members' residency programs would like them to approach their reviews of resident performance data.⁴²
- Adequate orientation to the definition of RSQMs and how these measures are developed (i.e., by engaging supervisors and residents with experience in the settings for which RSQMs have been developed).^{27,33}
- A high-level summary of why the measures apply in the settings for which they have been developed for CCC members reviewing RSQMs from clinical areas outside their expertise.
- A reminder that CCC members should not make assumptions based on information they do not possess or on how they would perform if they were in the resident's place (e.g., even though an action was not documented, assuming that the resident had actually done it). To help avoid this situation, context for RSQM performance should be provided to CCC members alongside data on performance for the measures themselves.

Such faculty development comes at a cost, most notably in time. However, regular CCC member training is a core component of high-value CCCs.⁴² The visual presentation of quality measure data to CCC members will also be important. For institutions with extensive experience gathering and presenting dashboard data for quality measures, the time and people cost for preparing visual presentations may be negligible. It should also be noted that some of our initial RSQMs were gathered via manual chart review. However, at our site, and likely at others as well, some of these RSQMs can be extracted from the electronic health record automatically. Furthermore, many others can be gleaned by using natural language processing of written documentation in the electronic health record.

While there are costs associated with implementing RSQM-informed assessment, there may also be costs associated with not implementing this innovation: namely, lower-quality patient care. Our previous work demonstrates the wide range of resident performance on these measures, which have been prioritized as important for care for the illnesses of interest.³⁵

Finally, while the focus of this study was on how CCC members use RSQMs in completing an assessment task, participants did raise the perceived value in presenting RSQM data to residents for formative purposes as well. While a more detailed exploration of this idea was

beyond the scope of this study, future efforts should engage residents in characterizing how RSQMs can be used to inform and guide their personal improvement efforts.

Limitations

This study has limitations to consider. First, it was done at an institution where quality improvement is integrated into the system, making it possible that our participants view and use RSQMs differently than individuals in other settings would, thus limiting the transferability of our findings. Second, while we designed this study to mimic authentic resident portfolio reviews by a CCC member, it was conducted in a study setting, potentially influencing participant responses. Components of our process that differed from authentic reviews included using deidentified data (without resident name or training year) and providing RSQMs (which is normally not done). The latter was the intended focus of the study. It is known that the presence of a resident name and/or training year can influence assessment decisions, but we were unable to capture this aspect of reviews based on the need for confidentiality.⁴³ Third, we did identify a discrepant case in the dataset that bordered on viewing the RSQMs in a dismissive manner that could not be fully characterized as an isolated case. Future work should seek to determine if cases such as this one are more common than it seems in our initial data. Fourth, RSQM benchmarks were not included initially; however, we were able to characterize them as helpful and important to some participants and not for others through theoretical sampling in the final interviews. Fifth, we were able to determine how participants used or did not use RSQMs to triangulate assessment with other data presented to them and vice versa; however, given our intent not to focus on RSQMs too heavily in our probing, we were not able to fully characterize this triangulation. We also were not able to determine the role that CCC discussion with other members of the CCC plays in how RSQMs are viewed and used because we focused only on individual CCC members. This interaction should be the focus of future RSQM study. Finally, the dimensions and profiles identified in this study may apply to other types of assessment data that CCCs review as well. Determining these possible relationships was beyond the scope of this study but should be the focus of future work.

Conclusions

In this study, participants used RSQMs to varying extents in their review of resident data to make entrustment decisions and found such data helpful to varying degrees. Importantly, participants largely found RSQMs to be of some utility despite receiving no orientation before seeing the measures for the first time in this study. These findings support the inclusion of RSQMs as resident assessment data based on participants' receptivity to them.

References

1. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—Rationale and benefits. *N Engl J Med*. 2012;366:2324–2325.
2. Mink RB, Schwartz A, Herman BE, et al. Validity of level of supervision scales for assessing pediatric fellows on the common pediatric subspecialty entrustable professional activities. *Acad Med*. 2018;93(2):283–291.
3. Mink RB, Myers AL, Turner DA, Carraccio CL. Competencies, milestones, and a level of supervision scale for entrustable professional activities for scholarship. *Acad Med*. 2018;93(11):1668–1672.
4. Promes SB, Wagner MJ. Starting a clinical competency committee. *J Grad Med Educ*. 2014;6(1):163–164.
5. French JC, Dannefer EF, Colbert CY. A systematic approach toward building a fully operational clinical competency committee. *J Surg Educ*. 2014;71(6):e22–27.
6. Hauer KE, Kohlwes J, Cornett P, et al. Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ*. 2013;5(1):54–59.
7. Chahine S, Cristancho S, Padgett J, Lingard L. How do small groups make decisions? A theoretical framework to inform the implementation and study of clinical competency committees. *Perspect Med Educ*. 2017;6(3):192–198.
8. Hauer KE, Chesluk B, Iobst W, et al. Resident competence review in graduate medical education: A qualitative study. *Acad Med*. 2015;90(8):1084–1092.
9. Hauer KE, ten Cate O, Boscardin CK, et al. Ensuring resident competence: A narrative review of the literature on group decision making to inform the work of clinical competency committees. *J Grad Med Educ*. 2016;8(2):156–164.
10. Ross FJ, Metro DG, Beaman ST, et al. A first look at the Accreditation Council for Graduate Medical Education anesthesiology milestones: Implementation of self-evaluation in a large residency program. *J Clin Anesth*. 2016;32:17–24.
11. Sklansky DJ, Frohna JG, Schumacher DJ. Learner-driven synthesis of assessment data: Engaging and motivating residents in their milestone-based assessments. *Med Sci Educ*. 2017;27(2):417–421.
12. Ketteler ER, Auyang ED, Beard KE, et al. Competency champions in the clinical competency committee: A successful strategy to implement milestone evaluations and competency coaching. *J Surg Educ*. 2014;71(1):36–38.
13. Shumway NM, Dacus JJ, Lathrop KI, Hernandez EP, Miller M, Karnad AB. Use of milestones and development of entrustable professional activities in 2 hematology/oncology training programs. *J Grad Med Educ*. 2015;7(1):101–104.
14. Hong R. Observations: We need to stop drowning—A proposal for change in the evaluation process and the role of the clinical competency committee. *J Grad Med Educ*. 2015;7(3):496–497.
15. Mount CA, Short PA, Mount GR, Schofield CM. An end-of-year oral examination for internal medicine residents: An assessment tool for the clinical competency committee. *J Grad Med Educ*. 2014;6(3):551–554.

16. Donato AA, Alweis R, Wenderoth S. Design of a clinical competency committee to maximize formative feedback. *J Community Hosp Intern Med Perspect*. 2016;6(6):335-338.
17. Schumacher DJ, Sectish TC, Vinci RJ. Optimizing clinical competency committee work through taking advantage of overlap across milestones. *Acad Pediatr*. 2014;14(5):436-438.
18. Johna S, Woodward B. Navigating the next accreditation system: A dashboard for the milestones. *Perm J*. 2015;19(4):61-63.
19. Friedman KA, Raimo J, Spielmann K, Chaudhry S. Resident dashboards: Helping your clinical competency committee visualize trainees' key performance indicators. *Med Educ Online*. 2016;21(1):298-303.
20. Hauer KE, Chesluk B, Lobst W, et al. Reviewing residents' competence: A qualitative study of the role of clinical competency committees in performance assessment. *Acad Med*. 2015;90(8):1084-1092.
21. Pool AO, Govaerts MJB, Jaarsma DADC, Driessen EW. From aggregation to interpretation: How assessors judge complex data in a competency-based portfolio. *Adv Health Sci Educ Theory Pract*. 2018;23(2):275-287.
22. Schumacher DJ, Michelson C, Poynter S, et al. Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. *Med Teach*. 2018;40(1):70-79.
23. Pack R, Lingard L, Watling CJ, Chahine S, Cristancho SM. Some assembly required: Tracing the interpretative work of Clinical Competency Committees. *Med Educ*. 2019;53(7):723-734.
24. Chahine S, Kulasegaram K, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Acad Med*. 2018;93(6):829-832.
25. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. *Acad Med*. 2018;93(7):975-978.
26. Arora VM. Harnessing the power of big data to improve graduate medical education: Big idea or bust? *Acad Med*. 2018;93(6):833-834.
27. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: A model from pediatric emergency medicine. *Acad Med*. 2018;93(7):1071-1078.
28. Smirnova A, Sebok-Syer SS, Chahine S, et al. Defining and adopting clinical performance measures in graduate medical education: Where are we now and where are we going? *Acad Med*. 2019;94(5):671-677.
29. Warm EJ, Mathis BR. Ambulatory education: Time to move from process to outcome. *J Grad Med Educ*. 2019;11(2):143-145.
30. Jaffe R, Diemer G, Caruso J, Metzinger M. Creating provider-level quality reports for residents to improve the clinical learning environment. *J Grad Med Educ*. 2017;9(3):381-382.
31. 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.
32. Lau BD, Streiff MB, Pronovost PJ, Haider AH, Efron DT, Haut ER. Attending physician performance measure scores and resident physicians' ordering practices. *JAMA Surg*. 2015;150(8):813-814.
33. Schumacher DJ, Martini A, Holmboe E, et al. Developing resident-sensitive quality measures: Engaging stakeholders to inform next steps. *Acad Pediatr*. 2019;19(2):177-185.

34. Schumacher DJ, Martini A, Holmboe E, et al. Initial implementation of resident-sensitive quality measures in the pediatric emergency department: A wide range of performance. *Acad Med*. In press.
35. Schumacher DJ, Holmboe E, Carraccio C, et al. Resident-sensitive quality measures in the pediatric emergency department: Exploring relationships with supervisor entrustment, patient complexity and acuity. *Acad Med*. In press.
36. Paradis E, Sutkin G. Beyond a good story: From Hawthorne Effect to reactivity in health professions education research. *Med Educ*. 2017;51(1):31–39.
37. American Board of Pediatrics. Entrustable Professional Activities. American Board of Pediatrics. Published 2017. <https://www.abp.org/entrustable-professional-activities-epas>. Accessed May 10, 2017.
38. Charmaz K. *Constructing Grounded Theory*. London, UK: SAGE Publications Ltd; 2014.
39. Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning from clinical work: The roles of learning cues and credibility judgements. *Med Educ*. 2012;46:192–200.
40. Rogers E. *Diffusion of Innovations*. 5th ed. New York, NY: Free Press; A Division of Simon & Schuster, Inc.; 1962.
41. Apramian T, Cristancho S, Sener A, Lingard L. How do thresholds of principle and preference influence surgeon assessments of learner performance? *Ann Surg*. 2018;268(2):385–390.
42. Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Med Teach*. 2018;40(11):1110–1115.
43. Stroud L, Herold J, Tomlinson G, Cavalcanti RB. Who you know or what you know? Effect of examiner familiarity with residents on OSCE scores. *Acad Med*. 2011;86(10 Suppl):S8–11.

Appendix 1

Resident-Sensitive Quality Measure (RSQM) Performance From Resident Portfolio, Contribution of Resident-Sensitive Quality Measure Data to Competency Entrustment of Pediatrics Residents Study, Cincinnati Children's Hospital Medical Center, 2019

Benchmark for All Measures: 80%

ASTHMA QUALITY MEASURES	Number of encounters resident met measure	Number of encounters resident did NOT meet measure	Number of encounters that measure did not apply (e.g., measure related to discharge and patient was admitted)
Total asthma encounters since last report = 11			
Use asthma order set	8 (72%)	3	0
Correct medication dose ordered for albuterol	8 (72%)	3	0
Use of dexamethasone as steroid	11 (100%)	0	0
Correct medication dose ordered for dexamethasone	11 (100%)	0	0
Time from when resident assigned self to patient to steroid order less than 30 minutes	4 (36%)	7	0
Correct medication dose ordered for ipratropium	8 (72%)	3	0
Documentation of previous intubation or BIPAP for asthma	3 (27%)	8	0
Note the acuity of the patient in documentation	11 (100%)	0	0
Documentation of work of breathing	11 (100%)	0	0
Documentation of aeration/air exchange	11 (100%)	0	0
Presence or absence of wheezing documented	11 (100%)	0	0
Ensure at least 3 descriptive words in respiratory exam	11 (100%)	0	0
Resident documents own PRAM score	4 (36%)	7	0
Resident-assigned PRAM score matches resident-placed orders	3 (27%)	1	7
Document response to intervention	10 (91%)	1	0
Documentation of disposition decision	10 (91%)	1	1
Use of standardized dosing for discharge medications—dexamethasone	7 (64%)	0	4
Home dexamethasone instructions documented in written discharge instructions	0 (0%)	7	4
State who to follow up with and information to contact them is in discharge papers	5 (45%)	2	4
Documentation of needing albuterol more often than every 4 hours as a reason to return in written discharge instructions	0 (0%)	7	4
Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	5 (45%)	2	4

BRONCHIOLITIS QUALITY MEASURES	Number of encounters resident met measure	Number of encounters resident did NOT meet measure	Number of encounters that measure did not apply (e.g., measure related to discharge and patient was admitted)
Total bronchiolitis encounters since last report = 9			
Follow bronchiolitis pathway	6 (67%)	3	0
Document birth history (preemie or not a preemie)	5 (56%)	4	0
Day of illness clearly documented	8 (89%)	1	0
Documentation of previous wheezing	0 (0%)	9	0
Assessment of severity documented	8 (89%)	1	0
Effort of breathing documented	9 (100%)	0	0
Documented quality of air entry (normal, decreased, etc.)	7 (78%)	2	0
Documentation of wheezing	7 (78%)	2	0
Documentation of crackles	8 (89%)	1	0
Documented presence or absence of subcostal retractions	3 (33%)	6	0
Documented presence or absence of intercostal retractions	1 (11%)	8	0
Documented presence or absence of suprasternal retractions	0 (0%)	9	0
Oxygen saturation clearly documented in note	0 (0%)	9	0
Hydration status clearly documented in note	0 (0%)	9	0
Documentation of response to specific therapeutics (i.e., how they responded to suctioning, how they responded to breathing treatment, how they responded to NS bolus, etc.)	5 (56%)	4	0
Oral feeding tolerance clearly documented in note	5 (56%)	4	0
Documentation of justification for appropriate disposition (home versus admit)	8 (89%)	1	0
Documentation of worsening respiratory symptoms as a reason to return in written discharge instructions	6 (67%)	3	3
Use standard/pre-made discharge instructions for bronchiolitis	4 (44%)	5	3
Documentation of poor feeding as a reason to return in written discharge instructions	0 (0%)	6	3
Bulb suction teaching for home ordered	0 (0%)	6	3
State who to follow up with and information to contact them in discharge papers	5 (56%)	1	3
State appropriate number of days to follow-up in discharge papers (within 2 days)	3 (33%)	3	3

CLOSED HEAD INJURY QUALITY MEASURES	Number of encounters resident met measure	Number of encounters resident did NOT meet measure	Number of encounters that measure did not apply (e.g., measure related to discharge and patient was admitted)
Total closed head injury encounters since last report = 5			
CHI/PECARN best practice advisory use	5 (100%)	0	0
Appropriate CHI/PECARN pathway use	5 (100%)	0	0
Mechanism of injury documented	5 (100%)	0	0
Documentation of time of injury	3 (60%)	2	0
Documentation of presence or absence of LOC	4 (80%)	1	0
Documentation presence or absence of emesis	5 (100%)	0	0
Documentation of whether back to baseline or not	4 (80%)	1	0
Documentation of presence or absence of other injury in the history	2 (40%)	3	0
Assessment of severity documented	5 (100%)	0	0
Documentation of presence or absence of hematoma on PE (if present, location and size also documented)	3 (60%)	2	0
Thorough head exam (head, eyes, skull) documented	0 (0%)	5	0
Documentation of GCS	1 (20%)	4	0
FULL neuro exam documented	0 (0%)	5	0
Documentation of presence or absence of other injury on PE	1 (20%)	4	0
Appropriate differential diagnosis and medical decision making documented	5 (100%)	0	0
Reassessments of patient documented	2 (40%)	3	0
Return to school/play recommendations in discharge papers	0 (0%)	2	3
Use of standard/pre-made discharge instructions for diagnosis	5 (100%)	0	0
Appropriate follow-up (sports med, rehab, neuro, etc.) recommended	3 (60%)	2	0

Abbreviations: BIPAP indicates bilevel positive airway pressure; PRAM, Pediatric Respiratory Assessment Measure; NS, normal saline; CHI/PECARN, closed head injury/Pediatric Emergency Care Applied Research Network; LOC, loss of consciousness; PE, physician exam; GCS, Glasgow Coma Scale.

Supplemental Appendix 1: Resident Portfolio Synopsis

To create the resident assessment portfolio, we received permission from a CCMHC resident to use personal assessment data over an approximately six-month time period with all identifying information (name, gender, training year) removed. To this, we added RSQM data to the emergency medicine rotation data (such data were being studied during the timeframe of the actual assessment data but were not in use beyond the study). The goal of developing the mock resident portfolio was to represent a “typically developing” resident in the first part of residency. By typically developing, we intended for the resident to have performance data that would likely be viewed as similar to that of many residents who have limited areas of concern rather than pervasive performance issues. Assessment data from five rotation settings with the following type of data, by setting, were included in the mock resident portfolio:

- Continuity clinic (general pediatric clinic where residents care for patients each week): supervisor milestones and entrustable professional activity (EPA) assignments, supervisor narrative comments
- Emergency medicine: supervisor milestones assignments; RSQMs for asthma, bronchiolitis, and closed head injury; and supervisor narrative comments
- Pediatric intensive care unit: supervisor milestones and EPA assignments, supervisor narrative comments
- Neonatal intensive care unit: supervisor milestones and EPA assignments, supervisor narrative comments
- Pediatric intensive care unit: supervisor milestones and EPA assignments, supervisor narrative comments, peer milestone assignments, and peer narrative comments

Across rotations, supervisor milestones and EPA assignments placed the resident at or near level 3 (out of 5 levels for both constructs, with raters allowed to place residents between levels for milestones). For EPA items, level 3 was “may act under reactive supervision (supervision readily available on request).”

In continuity clinic, written comments focused on outstanding bedside manner and communication skills, efficiency in clinical reasoning, and the need to work on general efficiency with encounters. In the pediatric emergency department (PED), comments noted the need to focus on admitting what the resident did not know and seeking help when needed, being more thoughtful and thorough with patient presentations (including clinical reasoning that is “sometimes quite far off”), and not leaving after a shift before staffing all patients. PED comments also noted that the resident reported things that were not done and dismissed corrective feedback. Finally, two comments from the PED specifically noted difficulty trusting the resident. In the pediatric intensive care unit, comments noted that the resident was very dedicated to patients, spending

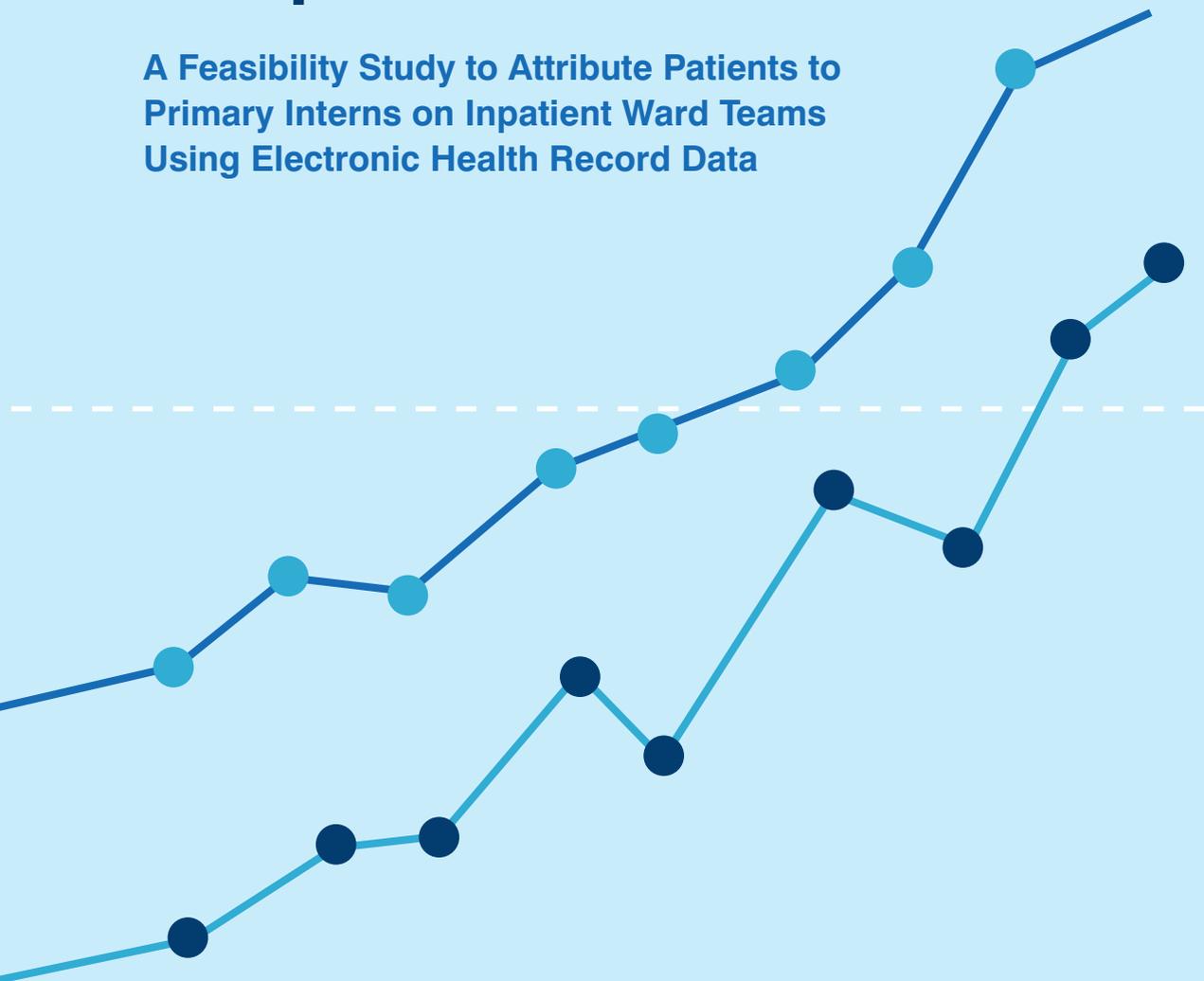
a lot of time with families outside of rounds. They also noted that patient presentations, including the ability to present a well-developed plan, improved significantly over the course of the rotation. Finally, comments from the pediatric intensive care unit noted that the resident minimized concerns or illness severity of patients. Neonatal intensive care unit comments noted conscientiousness, courtesy, and the need to deepen knowledge base. Supervisor comments from hospital medicine noted that the resident had a good knowledge base and was dedicated to other learners but that the resident also tended to minimize the illness severity of patients and had issues with efficiency and completeness. Peer assessments from hospital medicine were generally better than any of the supervisor assessments (levels 4 to 5 for milestones/EPAs), with written comments on these forms lauding interactions with patients, families, and other learners.

RSQM data for the PED were presented by illness with details of each RSQM provided (see Appendix 1 in this chapter). Information provided for each measure included number of times the measure should have been met and was, number of times the measure should have been met and was not, and number of times the resident did not have the opportunity to meet the measure. The RSQM data were created based on average resident performance on these measures during a previous study (see chapter 6).



Chapter 9

A Feasibility Study to Attribute Patients to Primary Interns on Inpatient Ward Teams Using Electronic Health Record Data



Schumacher DJ, Wu DTY, Meganathan K, Li L, Kinneer B, Sall D, Holmboe E, Carraccio C, van der Vleuten CPM, Busari J, Kelleher M, Schauer D, Warm E. "A Feasibility Study to Attribute Patients to Primary Interns on Inpatient Ward Teams Using Electronic Health Record Data." *Academic Medicine* 2019;94(9):1376-1383.

Abstract

Purpose

To inform graduate medical education (GME) outcomes at the individual resident level, this study sought a method for attributing care for individual patients to individual interns based on “footprints” in the electronic health record (EHR).

Method

Primary interns caring for patients on an internal medicine inpatient service were recorded daily by five attending physicians of record at University of Cincinnati Medical Center in August 2017 and January 2018. These records were considered gold standard identification of primary interns. The following EHR variables were explored to determine representation of primary intern involvement in care: postgraduate year, progress note author, discharge summary author, physician order placement, and logging clicks in the patient record. These variables were turned into quantitative attributes (e.g., progress note author: yes/no), and informative attributes were selected and modeled using a decision tree algorithm.

Results

A total of 1,511 access records were generated; 116 were marked as having a primary intern assigned. All variables except discharge summary author displayed at least some level of importance in the models. The best model achieved 78.95% sensitivity, 97.61% specificity, and an area under the receiver-operator curve of approximately 91%.

Conclusions

This study successfully predicted primary interns caring for patients on inpatient teams using EHR data with excellent model performance. This provides a foundation for attributing patients to primary interns for the purposes of determining patient diagnoses and complexity the interns see as well as supporting continuous quality improvement efforts in GME.

Recently, calls to determine the outcomes of graduate medical education (GME)¹⁻⁷ and align the goals and purposes of medical education with those of health care delivery⁸ have been increasing. These appeals are consistent with the intent of competency-based medical education (CBME) to ensure training curricula and outcomes that prepare graduates to meet the needs of patients.⁹ Naturally, achieving this aim requires defining the goals of GME and measuring achievement of those goals.¹ The medical education community has suggested that “big data” can help achieve this ambition.^{2,4} Educators have further noted that big data may be the answer to linking data on education with data on care.³ Although publicly available big data from local, state, and federal governments are increasingly common in medicine, the GME community is not taking advantage of the opportunities afforded.^{2,4} This may be a result of the challenges with harnessing big data for these purposes.³ Indeed, the limited efforts that have used big data to inform medical education outcomes have noted the substantial resources required to complete this work.¹⁰

Recent calls to determine GME outcomes have also placed more focus on programs, institutions, and the GME system on the whole rather than on individual residents or fellows.^{5,11,12} Although this focus is important, individual residents are equally imperative to consider given that providing individuals with data related to their performance on the team may be the best way to drive their personal improvement.^{13,14} Furthermore, we graduate, certify, and credential individuals in medicine rather than teams—and competent individuals form the basis of functional teams—making it crucial to be able to determine an individual’s performance. Finally, discerning the performance of individuals on teams can be used collectively to inform team performance.

Efforts to provide resident-level quality feedback are challenged by issues of attribution,¹⁵ with some calling the ability to attribute performance to an individual into question.¹⁶ However, other work offers promise in this area. For example, some of us have developed resident-sensitive quality measures, which attempt to capture work that is likely attributable to individual residents.¹⁷ Furthermore, Levin and Hron¹⁸ have harnessed the electronic health record (EHR) to provide data on patient volumes and diagnoses seen by individual residents. Finally, Herzke and colleagues¹⁹ describe a method for attributing patient-level metrics to attending physicians through the type, timing, and number of charges for patient hospitalizations. To explore the use of big data in the EHR at the individual resident level, we sought to determine a method for attributing care for individual patients to individual interns based on “footprints” in the EHR (i.e., activities logged in the EHR). We believe such modeling represents a first step toward disentangling overall care and attribution of that care. If it is not possible to identify a primary intern for a patient, it is also not possible to disentangle the level of contribution of more than one intern or other member of the team.

In this study, we intended to demonstrate the feasibility of predicting primary interns caring for patients using selected EHR data. Although this preliminary effort does not consider patient care outcomes attributed to these interns or consider the role that supervisors and other members of the health care team play in modifying those outcomes, it does provide the foundation to take those next steps. Furthermore, it informs opportunities for automation of case logs to track factors such as diagnoses and ranges of patient complexity seen by interns. It also allows for connection of interns to quality data for their patients, creating opportunities for reflective continuous quality improvement exercises and, ultimately, relative contribution of these interns to improving care quality and outcomes. All of these purposes serve to answer the call to determine the outcomes of GME training to ensure that they indeed serve to meet the needs of patient populations.

Method

Clinical Setting

This study was conducted at the University of Cincinnati Medical Center (UCMC). All residents rotating on the inpatient general medicine wards during the time frames of interest were considered eligible. The internal medicine residency at UCMC is a three-year residency program with 89 categorical (i.e., noncombined training program) and preliminary year (e.g., single-year residents destined for another specialty after that year) trainees. UCMC has two types of general medicine inpatient teams: attending, senior resident, intern, third-year medical student, and often a fourth-year medical student; and attending and intern only. Both teams see general internal medicine patients on the same units.

We considered eligible patient records to be those that had an eligible intern as the primary intern on each of the days of the study. We conceptually defined “primary intern” as a trainee who was assigned to a patient and primarily responsible for delivery of that patient’s care on a given day (e.g., consulting other services, writing progress notes, communicating with the patient and other members of the health care team) as determined by the attending physician of record.

Model Predictor Selection

Primary interns often write daily progress notes and discharge summaries as well as enter orders for their primary patients, so we chose these activities as variables to include in modeling.

We also chose information about users’ interactions with the EHR system for inclusion in modeling. Such interactions are recorded automatically in the form of event logs or audit trails (“EHR clicks” hereafter). Our hypothesis was that EHR clicks can play a significant role in predicting primary interns. A full list of EHR clicks is shown in Supplemental Appendix 1. We believed that many of

these EHR clicks may not be germane to the work that interns often do. Therefore, we created two additional categories of EHR clicks based on the perceived importance of types of clicks to identify a primary intern. To determine this, we convened a group of 14 current internal medicine residents to provide feedback on the perceived importance of click types to likely represent a resident caring for a hospitalized patient on a resident team. Residents were provided a list of clicks and asked to circle those that they believed were commonly done by interns and upper-level residents (i.e., circle everything that applies) and star those that they strongly believed were commonly done by interns and upper-level residents (i.e., star what is most important). Example clicks prioritized by residents included medications activity accessed, note viewed in chart review, inpatient sign-out activity accessed, and inpatient orders section accessed. All included clicks, with resident voting, are shown in Supplemental Appendix 2.

In addition to the resident-generated categories, most of the physician authors (D.J.S., B.K., D.R.S., M.K., and E.W.), all experienced GME administrators, selected click behaviors they felt were common and likely performed by interns and upper-level residents to form a separate category in the dataset. These are shown in Supplemental Appendix 3.

The goal of gathering feedback from residents and our author team was simply to crudely eliminate click types that were unlikely to be useful in modeling (e.g., inpatient education activity accessed, barcode scanned, and edit claim information window accessed). All “eliminated” click types were considered as part of the total number of clicks in the analyses. The hope was that eliminating some of these in future modeling would pick up more signal in the noise. We deemed crude grouping appropriate, given the desire to develop a model rather than refine one.

Data Collection

On a daily basis, five attending physicians (including authors B.K. and D.R.S.) independently recorded the primary intern for each patient each day during their service time from August 1 to 12, 2017; August 17 to 25, 2017; and January 8 to 13, 2018. As such, we considered each day as a discrete primary intern–patient pair to allow for changes in primary interns with activities such as cross-covering.

This dataset was expanded both in terms of columns and rows, by combining it with the data from the UCMC EHR (Epic Systems Corporation; Verona, Wisconsin). For row expansion, because multiple EHR users were involved in patient care on each day, each EHR user who had “touched” the patient had a row in the dataset. That is, we included all EHR users (i.e., intern and nonintern, as well as physician and nonphysician) with any interaction in the EHR on a given day in this expansion. For column expansion, for each EHR user of each patient on each day, we added the following information:

- patient admission/discharge time and length of stay;
- provider type and postgraduate year (PGY) of residency, if applicable;
- whether this person was the progress note author for the day;
- whether this person was the discharge summary author for the encounter;
- whether this person ever placed an order for the patient on that day as well as the total number of orders placed;
- total EHR clicks; and
- EHR clicks in different categories (e.g., notes viewed, flowsheets activity accessed, orders viewed).

The data collection and study conduction were reviewed and determined to be exempt by the UCMC Institutional Review Board (no. 2016-8982).

Data Manipulation

We further manipulated the expanded dataset to generate informative variables. First, residents from the fourth year of the medicine–pediatrics combined training program were coded as postgraduate year 3 (PGY3). Because there was a small number of these residents, their data behaved the same as PGY3 categorical residents, and they (like PGY3 categorical residents) were rarely expected to be identified as a primary resident. Second, the total numbers of clicks were ranked among all EHR users and within all the residents (interns and noninterns). For example, a resident may have been ranked first with the highest number of clicks among other residents but may have been ranked fifth overall because other users (e.g., nurses) may have touched the patient more in the EHR. Our pilot experiment showed that the former was a better predictor, so the latter was dropped. Third, we recategorized the clicks because the original categories prescribed by the EHR vendor were too granular. This new categorization was made based on the perceived importance of types of clicks to identify a primary intern, as detailed above.

Data Modeling

Before modeling, we eliminated all nonresident EHR users from the dataset based on provider types in our EHR system. The final dataset was split into two subsets for training and validation, respectively. Training data were used to train the model, and validation data were used to validate that model. We considered two data split proportions to examine the differences between model performance. The first split (80–20) used 80% of the data for training and 20% for validation, as is commonly used in machine learning tasks.²⁰ The second split (50–50) used half of the data for training as well as for validation. This split may have increased the model performance because of the larger validation dataset. We executed data splits at the patient level. All of the following variables were considered, and no systematic feature selection was performed.

1. Resident PGY
2. Did this resident write the progress note of this patient on this day?
3. Was this resident the author of the progress note of the patient on this day?
4. Did this resident write the discharge summary of the patient?
5. Was this resident the author of the discharge summary of the patient?
6. Did this resident place an order on this patient on this date?
7. Number of orders placed by this resident on this patient on this date
8. Rank of number of clicks among residents
9. Number of event logs of this resident on this patient on this date
10. Number of clicks deemed to be of medium importance among residents for this patient on this date
11. Number of clicks deemed to be of high importance among residents for this patient on this date
12. Number of clicks deemed to be of high importance by physician authors with GME administrative experience

However, we assigned variables differently to generate four models. As shown in Table 1, Model 1 had variables 1–9 as its input, and Model 2 replaced variable 9 with variables 10–12. Models 3 and 4 had a similar setting as Models 1 and 2 except that they did not have variable 1 (PGY). This variable assignment was expected to better demonstrate the performance gain/ loss when PGY or EHR click information was available.

Table 1 Attribution of Primary Interns on Internal Medicine Inpatient Teams Variable Assignment to Models, From a Study of Attribution of Patients to Primary Interns via EHR, University of Cincinnati Medical Center, 2017 and 2018

Model	Input variables ^a			
	#1 (PGY)	#2–8 (progress note, discharge summary, orders, and click rank among residents)	#9 (number of clicks)	#10–12 (click categories)
1	x	x	x	
2	x	x		x
3		x	x	
4		x		x

Abbreviation: EHR indicates electronic health record.

^aSee Table 3 for a full listing of the variables.

The data were modeled using SAS 9.4 software's high-performance procedure HPSPLIT (a decision tree algorithm) in SAS/STAT statistical software, version 14.3 (SAS Institute Inc., Cary, North Carolina). This decision algorithm determined the decisive variables in each layer by how

much information was gained when a new variable was added. The tree depth was set at five, and up to three leaves were permitted at each node. We selected a decision tree algorithm because of its easily interpretable output to understand the role of the predictor variables. Comparing decision trees with other modeling techniques was beyond the scope of this exploratory study.

We used the following metrics to evaluate the model performance: sensitivity (recall), specificity, precision, F score (a measure of accuracy, where 1 is best and 0 is worst), and the area under the receiver-operator curve (ROC), or AUC. Specifically, F score, AUC, and specificity of the validation dataset were used to determine the model with the best performance.

Results

A total of 369 recordings of primary interns were made by the five attending physicians. The dataset for patients related to these records was expanded to 23,242 access logs (all clicks/activities in the EHR) for individual EHR users who touched these patients in the EHR. On a given day, an average of 7 (range: 1–28) medical providers (medical students, residents, fellows, and attending physicians) touched a patient in the EHR. The dataset was also manipulated to generate the 12 variables described above. To focus on predicting primary interns, only internal medicine interns and upper-level residents were kept in the dataset, and all other clinicians were dropped. This led to a subset of 1,511 daily access logs (internal medicine interns and upper-level residents touching the patient) belonging to 120 patients, with 116 access logs (7.68%) marked by the attending physicians as primary interns.

The 80–20 data split at the patient level resulted in a training set of 1,193 (78.95%) records and 318 (21.05%) validation records; the 50–50 split resulted in a training set of 869 (57.51%) records and 642 (42.49%) records. Each split trained the same four models. The model performance is shown in Table 2. The four models in the 80–20 data split did not achieve better performance based on F score and AUC. In the 50–50 data split, Models 1 and 2 outperformed Models 3 and 4, indicating that PGY is a critical piece of information in determining a primary intern. In addition, recategorizing clicks based on perceived importance improved model performance. However, none of the models had desirable results for precision (all below 80%).

Table 3 summarizes the important variables of the four models in the 50–50 data split. Across all models, a key variable was whether a given clinician wrote the progress note for the day. The relative rank of number of clicks among residents was also consistently at the top of the list.

Table 2 Attribution of Primary Interns on Internal Medicine Inpatient Teams' Model Performance, From a Study of Attribution of Patients to Primary Interns via EHR, University of Cincinnati Medical Center, 2017 and 2018

Split and model	Training (80-20 Split: 1193 records; 50-50 Split: 869 records)				Validation (80-20 Split: 318 records; 50-50 Split: 642 records)					
	Sensitivity	Specificity	Precision	F-score	AUC	Sensitivity	Specificity	Precision	F-score	AUC ^b
80-20										
1	0.867	0.987	0.859	0.863	0.993	0.778	0.970	0.609	0.683	0.922
2	0.898	0.990	0.889	0.893	0.995	0.722	0.977	0.650	0.684	0.893
3	0.745	0.974	0.716	0.730	0.975	0.722	0.930	0.382	0.500	0.901
4	0.765	0.995	0.926	0.838	0.985	0.389	0.950	0.318	0.350	0.835
50-50										
1	0.915	0.993	0.900	0.907	0.997	0.754	0.966	0.683	0.717	0.896
2 ^a	0.881	0.995	0.929	0.904	0.997	0.790	0.976	0.763	0.776	0.910
3	0.763	0.979	0.726	0.744	0.984	0.526	0.961	0.566	0.545	0.800
4	0.881	0.989	0.853	0.867	0.993	0.597	0.957	0.576	0.586	0.806

Abbreviations: EHR indicates electronic health record; AUC, area under the receiver-operator curve.

^aModel with the best performance.

Table 3 Attribution of Primary Interns on Internal Medicine Inpatient Teams' Variable Selection of Four Models in 50–50 Data Split, From a Study of Attribution of Patients to Primary Interns via EHR, University of Cincinnati Medical Center, 2017 and 2018

Variable number	Variable description	Ranks in model			
		Model 1	Model 2	Model 3	Model 4
1	Resident postgraduate year	2	2	--	--
2	Did this resident write the progress note of this patient on this day?	1	1	1	1
3	Was this resident the author of the progress note of the patient on this day?	-- ^a	-- ^a	-- ^a	-- ^a
4	Did this resident write the discharge summary of the patient?	4	4	4	3
5	Was this resident the author of the discharge summary of the patient?	-- ^a	-- ^a	-- ^a	-- ^a
6	Did this resident place an order on this patient on this date?	-- ^a	-- ^a	-- ^a	-- ^a
7	Number of orders placed by this resident on this patient on this date	-- ^a	-- ^a	5	5
8	Rank of number of clicks among residents	3	3	2	2
9	Number of event logs of this resident on this patient on this date	-- ^a	-- ^b	3	-- ^b
10	Number of clicks deemed to be of medium importance among residents for this patient on this date	-- ^b	-- ^a	-- ^b	7
11	Number of clicks deemed to be of high importance among residents for this patient on this date	-- ^b	-- ^a	-- ^b	4
12	Number of clicks deemed to be of high importance by physician authors with GME administrative experience	-- ^b	-- ^a	-- ^b	6

Abbreviations: EHR indicates electronic health record; GME, graduate medical education.

^aVariable not selected as important in the model.

^bVariable not assigned to the model.

A trainee's PGY was the second most important variable if it was included (Models 1 and 2). However, this variable seems to provide strong information to identify primary interns such that the click variables (total clicks and categorized clicks, or variables 9–12) were not as important. When PGY was not included in a model, the click variables played a more vital role in predicting primary interns (Model 4).

The best model in our study was Model 2 in the 50–50 data split, which achieved 78.95% sensitivity and 97.61% specificity. Figure 1 demonstrates the ROC curve of this model. The AUC was approximately 91% on the validation dataset. Figure 2 further illustrates the decision tree of Model 2. This decision tree has five levels. The first decision was whether the clinician wrote the progress note on the patient on that day. If yes (ID:1) and the clinician's PGY is 2 or 3 (ID:3), this clinician was not a primary intern. If this clinician's PGY was 1 (ID:4) and he/she ever placed an order on this patient on the specified date (ID:9), this clinician was likely to be a primary intern. On the other hand, if this clinician was not a progress note author (ID:2) and his/her PGY was 2 or 3 (ID:5,6), this clinician was not a primary intern. However, if his/her PGY number was 1 (ID:7), his/her ranking of total clicks among residents would be a decisive factor (ID:A,B,C).

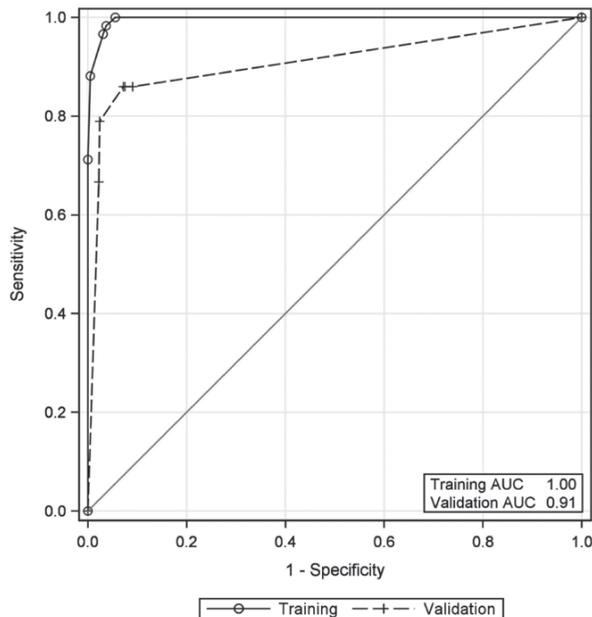


Figure 1: Attribution of primary interns on internal medicine inpatient teams receiver-operator curve for Model 2 in 50–50 split (best performance). From a study of attribution of patients to primary interns via EHR, University of Cincinnati Medical Center, 2017 and 2018. Abbreviations: EHR indicates electronic health record; AUC, area under the receiver-operator curve.

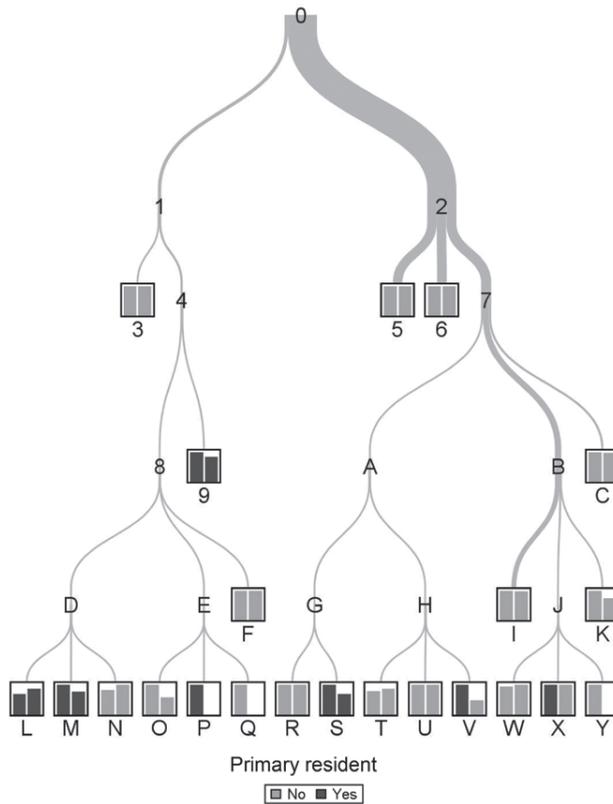


Figure 2: Attribution of primary interns on internal medicine inpatient teams decision tree for Model 2 in 50–50 split (best performance). From a study of attribution of patients to primary interns via EHR, University of Cincinnati Medical Center, 2017 and 2018. Abbreviation: EHR indicates electronic health record.

Supplemental Appendix 4 lists all the decision points and the counts of each subdataset, which can be used to implement this model to predict primary interns using EHR data. Click variables were largely used in levels 3–5 to identify primary interns. As can be seen in Table 3, these click variables were less important than other variables in the upper levels, although they still provided some information to identify primary interns.

Discussion

In this study, we successfully predicted primary interns caring for patients on inpatient internal medicine teams using EHR data with excellent model performance. This proof-of-concept study demonstrates that individual resident attribution in the EHR is possible and can be automated through computerized algorithms. Modeling and efforts such as those undertaken in this study

provide the foundation for attributing patients to primary interns and expanding to attributing patient care to other residents and team members as well.

Model Performance

In our study, PGY and being the daily progress note author were decisive factors in determining primary intern. These findings are expected because interns are PGY1s and write the daily progress notes for those patients at our institution. However, interns also provide care for patients who are not their primary patient, and multiple interns on an inpatient team are each the primary intern for only a subset of the patients cared for by that team when multiple interns are present on a team. Therefore, although decisive, PGY has limitations, highlighting the importance of additional modeling.

Although progress note author was a decisive factor in our best model, it was most decisive in excluding individuals as primary interns when they were not primary interns but not when identifying primary interns when they wrote daily progress notes. We did not consider other authors of daily progress notes, but future work should consider this.

Our study also found that EHR clicks contained critical information about residents' behaviors and patient touches in the EHR and had great potential to predict primary interns. These data already exist in the EHR, but they are largely ignored and may not be stored properly for analysis. For example, in our system the log data are archived within a year, so conducting a retrospective attribution study can be very resource consuming. Our model stresses the value in harnessing these data to identify primary interns. Once commandeered, such data may serve a role for understanding residents' workflow, providing insights into improving efficiency and organization.

Practical Considerations and Future Work

The best model in our study (i.e., Model 2) achieved 78.95% sensitivity, 97.61% specificity, and an AUC of about 91% for the validation dataset. This AUC is very close to perfect and indicates excellent model performance.

These results are likely best viewed through the lens of the intended use of automation to determine a primary intern in the EHR. Our goal was to *correctly* attribute patients to primary interns. Therefore, specificity is most important, and our model was excellent in this task. Given our results, *many* encounters could likely be yielded for the purpose of engaging residents in quality improvement efforts, which is currently suboptimal.²¹ Therefore, identifying *any*, let alone many, encounters where quality measures from a resident's primary patients can be provided for reflective continuous quality improvement efforts, including identifying and addressing critical deficiencies and following developmental progress over time, can make an important contribution. However, it is important to bear in mind that this study considered interns, and

thus the focus should likely remain on continuous quality improvement rather than any attempts to infer higher-stakes attribution of care, which have influences from supervisors and other members of the team. It is also important to bear in mind that our model considered primary interns on a daily basis, and thus an intern who served as the primary intern for the same patient for several days could correctly be ascribed more weight in continuous quality improvement activities but incorrectly counted more than once if using the model to identify diagnoses and complexity seen by the intern.

Although our results offer benefit for primary interns, much work remains. First, future modeling efforts will need to discern attribution of patients to other interns as well as upper-level residents. Second, optimizing the model to gain sensitivity will likely be important. For continuous quality improvement purposes, tilting the ROC toward specificity and trying to avoid false positives is likely acceptable. However, using this process to make higher-stakes performance assessment decisions, which would have tremendous value in the CBME era with a focus on educational and patient outcomes,²²⁻²⁴ will require better sensitivity as well as better understanding of relative contribution toward overall care provided by primary interns. Moving forward, it will be important to learn more about the encounters that this model misses. For example, is less footprint in the EHR associated with worse care, better care, or no difference in care provided by a primary intern? Do those missed by the model have worse performance, and thus are activities in the model performed by other members of the team to compensate?

Future work should also focus on developing a model that seeks to discern the relative attribution of care for all health care professionals, or at least those in the most central frontline roles (e.g., physicians, advanced practice providers, nurses, pharmacists) of patient care. This was beyond the scope of our current study but important. Determining relative attribution for care would enable answering important questions about which actions and outcomes for patients are more or less likely to be attributed to primary interns compared with other providers, including supervising and other residents as well as attending physicians. For example, are adverse events and outcomes more or less likely to be attributed to primary interns, and what attribution to patient care resides with providers caring for patients when the primary intern is not present (e.g., night float, cross-covering, on-call residents)? Furthermore, how are outcomes shared among team members caring for patients at the same time? The breadth of future work to define attribution of all members of the team through EHR data underscores the importance of the modeling work we described in this article. Although such modeling could be obviated by simply having residents assign themselves a role in patient care each day in the EHR, such self-assignment is riddled with challenges, including the need to reliably assign oneself each day and/or unassign oneself when no longer in that role. In addition, it emphasizes the need to develop myriad definitions of roles that all residents on the care team (e.g., nonprimary intern, supervising resident, consult

service resident, continuity clinic resident whose patient is hospitalized) can reliably assign themselves to each day.

Limitations

This study has limitations to consider. First, our modeling is based on data collected by five attending physicians at a single institution. Future work should include applying various modeling techniques on larger datasets from multiple institutions. Second, we only considered primary interns and not other members of the team. Future work should explore other resident and nonresident members of the team. Third, this study sought to attribute *patients* to primary interns but not necessarily *patient care* to those interns. Given our focus on interns, much of the decision making that leads to patient care is likely a collaborative process in conjunction with other members of the health care team, perhaps most importantly senior residents and attending physicians. Future work should seek to tease out, to the extent possible, which EHR entries resulted from a decision made by primary interns before supervisor review, which entries are the result of a directive from a supervisor, and which entries resulted from a collaborative dialogue among team members. Although we cannot make such determinations with our current data, our data can be a starting place for such a line of inquiry. Additionally, this work is time-consuming. We estimate that our data extraction, manipulation, and modeling required 200 person-hours for completion of this study. However, implementing our model in a new setting would take considerably less time because of the foundation already laid. Finally, by design we only considered the individual in this study, but data such as those collected can also inform team performance and should be explored for such purposes.

Conclusions

Prediction models for health care decisions and patient outcomes deserve attention.²⁵ EHR data modeling can be a pathway to medical education's obligation to determine both educational outcomes as well as their impact on patient outcomes.

References

1. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. *Acad Med.* 2018;93:975–978.
2. Arora VM. Harnessing the power of big data to improve graduate medical education: Big idea or bust? *Acad Med.* 2018;93:833–834.
3. Chahine S, Kulasegaram KM, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Acad Med.* 2018;93:829–832.
4. Triola MM, Hawkins RE, Skochelak SE. The time is now: Using graduates' practice data to drive medical education reform. *Acad Med.* 2018;93:826–828.
5. Caverzagie KJ, Lane SW, Sharma N, et al. Proposed performance-based metrics for the future funding of graduate medical education: Starting the conversation. *Acad Med.* 2018;93:1002–1013.
6. Weinstein DF. Optimizing GME by measuring its outcomes. *N Engl J Med.* 2017;377:2007–2009.
7. Cohen E, Kuo DZ, Agrawal R, et al. Children with medical complexity: An emerging population for clinical and research initiatives. *Pediatrics.* 2011;127:529–538.
8. Sklar DP, Hemmer PA, Durning SJ. Medical education and health care delivery: A call to better align goals and purposes. *Acad Med.* 2018;93:384–390.
9. 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;376:1923–1958.
10. Carraccio CL, Englander R. From Flexner to competencies: Reflections on a decade and the journey ahead. *Acad Med.* 2013;88:1067–1073.
11. Silkens MEWM, Arah OA, Wagner C, Scherpbier AJJA, Heineman MJ, Lombarts KMJM. The relationship between the learning and patient safety climates of clinical departments and residents' patient safety behaviors. *Acad Med.* 2018;93:1374–1380.
12. 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;92:1740–1748.
13. Sandars J, Cleary TJ. Self-regulation theory: Applications to medical education: AMEE guide no. 58. *Med Teach.* 2011;33:875–886.
14. Schon DA. *The Reflective Practitioner: How Professionals Think in Action.* New York, NY: Basic Books, Inc.; 1983.
15. Berman S. Training pediatricians to become child advocates. *Pediatrics.* 1998;102(3 pt 1):632–636.
16. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: Implications for assessment and entrustment. *Med Educ.* 2018;52:970–980.
17. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: A model from pediatric emergency medicine. *Acad Med.* 2018;93:1071–1078.
18. Levin JC, Hron J. Automated reporting of trainee metrics using electronic clinical systems. *J Grad Med Educ.* 2017;9:361–365.

19. Herzke CA, Michtalik HJ, Durkin N, et al. A method for attributing patient-level metrics to rotating providers in an inpatient setting. *J Hosp Med.* 2018;13:470–475.
20. Whitenack D. *Machine Learning With Go: Implementation Regression, Classification, Clustering, Time-Series Models, Neural Networks, and More Using the Go Programming Language.* Birmingham, UK: Packt Publishing; 2017.
21. Butler JM, Anderson KA, Supiano MA, Weir CR. "It feels like a lot of extra work": Resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med.* 2017;92:984–990.
22. 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;85:844–851.
23. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014;89:721–727.
24. Carraccio C, Englander R, Holmboe ES, Kogan JR. Driving care quality: Aligning trainee assessment and supervision through practical application of entrustable professional activities, competencies, and milestones. *Acad Med.* 2016;91:199–203.
25. Shah ND, Steyerberg EW, Kent DM. Big data and predictive analytics: Recalibrating expectations. *JAMA.* 2018;320:27–28.

Supplemental Appendix 1: All Electronic Health Record Click Types for Attribution of Primary Interns on Internal Medicine Inpatient Teams

Metric ID Number	Metric Name
34005	Notes viewed
17304	Related encounters requested
34010	Notes activity and navigators accessed
20620	Visit Navigator template loaded
14200	Printing occurred
17104	Patient Encounter opened
34111	Flowsheet activity accessed
34101	Inpatient MAR activity accessed
20040	SmartText used
17227	Note viewed in Chart Review
20030	SmartLink used
17234	Order viewed in Chart Review
34200	Barcode scanned
17108	Order list changed
34707	Flowsheet accepted
34122	Inpatient Sign Out Report activity accessed
34150	Order Reconciliation Section accessed
34706	Flowsheet viewed
34141	MAR accessed
20045	SmartPhrase used
49008	Visit Navigator Flowsheet section is accessed
20001	In Basket message viewed
17106	Medications activity accessed
48000	Order Verification Activity accessed
48003	Verified order dispensed
34146	MAR administration form accessed
17338	BestPractice Advisories displayed
49077	Narrator Event Log accessed
49031	Narrator accessed
34105	Inpatient Education activity accessed
34142	MAR administration accepted
49017	Visit Navigator Flowsheet section viewed
34127	Inpatient Orders section accessed
17015	Problem List activity accessed
17336	BestPractice Advisories viewed
24550	Patient Station accessed
60012	Sign clinical note
34100	Inpatient Care Plan activity accessed
49002	ED Visit Navigator Disposition section accessed
48004	i-Vent edited or viewed (pre-2016)
34119	Orders section accessed

49005 Viewed Discharge Instructions
52055 Imaging Schedule Orders Report Row accessed
15651 Form viewed
24058 Patient Summary viewed from Bed Planning
49018 Visit Navigator Flowsheet section accepted
17250 Problem List accessed
49004 Visit Navigator charge capture section viewed
60010 Pend clinical note
24299 Expected Discharge section accessed
17518 BPA follow-up action applied
17017 Envelope form in Letters activity accessed
34009 Viewed Sign Out Report
17121 Demographics activity selected
35030 Patient-based record exported
49003 Visit diagnoses modified
34123 Inpatient Intake/Output activity selected
17228 Encounter viewed in Chart Review
49015 Vist Navigator SmartForm Flowsheet section viewed
17165 Multistep and multistep collection tasks entry form accessed
34117 Opened Inpatient Order Set Activity
35020 Patient based RW report exported
17280 Chart searched
15136 Registration/ADT workflow initiated
72009 Home tab accessed in Care Everywhere activity
17188 Multistep and multistep collection tasks completed
34022 Saved New Sign Out Report Comment
15137 Registration/ADT workflow finished
20093 PCP History accessed
60005 View clinical notes section
49001 ED Visit Navigator Arrival section accessed
19125 Patient Chart Advisories viewed
49010 ED Visit Navigator Follow-Up section accessed
48001 Order History Activity accessed
17333 Actions taken from a BestPractice Advisory
51091 Order inquiry accessed
60000 Print notes from inpatient notes activity
49032 Orders accessed from Narrator
34175 Inpatient Work List task edited
34120 Inpatient Fast Note activity accessed
34137 Patient Education information viewed
17046 Charge Capture section accepted
17158 Visit diagnoses viewed
17152 OB and Lactation Status section accessed
60015 Cosign clinical note with attestation
34167 Patient Education documented on
17060 Patient Implants section or activity accessed
49061 Orders acknowledged in Narrator
20621 Visit Navigator section added to navigator

Supplemental Appendix 1: continued

17102	Progress Notes navigator section accessed
34704	Intake/Output activity accessed
20500	Successful Break-the-Glass
49076	Flowsheet Filed in Popup/Entry activity
34051	OB and Lactation Status accessed
60006	View clinical note from copy previous
49078	Narrator Event Log accessed
17281	Chart search results viewed
15604	Patient Emergency contacts grid accessed
41038	Account Note form accessed
17251	Problem List modified
15502	Coverage Additional Items subcomponent accessed
15505	Coverage Level Copay subcomponent accessed
15540	Coverage FC form Condensed Summary HTML subcomponent accessed
72011	Documents tab accessed in Care Everywhere activity
34118	Order sets accessed
17126	Allergies activity accessed
48020	i-Vent viewed
17259	Patient SnapShot viewed
15542	Coverage FC form Single Member Info subcomponent accessed
49054	ED Workup Activity viewed
48021	i-Vent created or edited
24032	ADT View Only Admission activity accessed
17148	Communication Management navigator section accessed
20075	Prelude Documents list accessed
34110	Inpatient Synopsis activity accessed
24205	Provider Info subcomponent accessed
49067	Flowsheet filed in Narrator
17124	Chief Complaint navigator section accessed
72013	Other Results tab accessed in Care Everywhere activity
41078	Hospital Billing Coverage Info accessed
17040	Diagnosis association updated
49033	Flowsheets accessed from Narrator
49030	ED Narrator Note accessed
72012	Lab Results tab accessed in Care Everywhere activity
24552	Form Reprints report viewed
15047	Patient Address subcomponent accessed
15048	Patient Demographics subcomponent accessed
15606	Patient Employer Information subcomponent accessed
34997	Time for Inpatient Medication adjusted
15167	Hospital Accounts form accessed from HAR Advisor
34012	Care Plan modified
24015	Event Management activity accessed
49037	ED Visit Navigator Clinical Impressions section accessed
49064	Note filed in Narrator
34700	LDA properties accessed

15501 Coverage Subscriber Employment subcomponent accessed
15506 Coverage Demographics subcomponent accessed
34014 Care Plan progress modified
17664 Synopsis activity accessed
17239 Note printed from Chart Review
20510 Break-the-Glass canceled
17821 Care Teams viewed
17822 Care Teams exited
15525 Coverage FC form Plan and Group Number subcomponent accessed
34701 LDA properties accepted
15544 Coverage FC form Auth Phone and Group Name subcomponent accessed
41087 Hospital Billing Account Maintenance accessed
52054 Radiology Protocol Work List Row accessed
15541 Coverage FC form Group number and Subscriber ID subcomponent accessed
17142 SmartForms accessed
15602 Patient Documents/Notices and Misc Info subcomponent accessed
72010 Summary tab accessed in Care Everywhere activity
34152 MAR administration edited
15611 Patient PCP Display Only HTML subcomponent accessed
15067 Hospital Accounts form accessed
17215 Release Orders activity accessed
35015 Detailed report viewed in Reporting Workbench
15532 Coverage FC form Covered Through subcomponent accessed
24246 Discharge Date and Time subcomponent accessed
24248 Discharge Disposition and Destination subcomponent accessed
24249 Discharge Comments subcomponent accessed
49050 ED Disposition Activity viewed
34143 MAR administration aborted
41077 Hospital Billing Account Summary accessed
41109 Hospital Account Note accessed
15621 Patient Relationships subcomponent accessed
24250 Discharge Bed Status and Service Priority subcomponent accessed
24251 Discharge Provider subcomponent accessed
24252 Discharge HTML Summary subcomponent accessed
24253 Discharge Code and Means of Departure subcomponent accessed
24553 ADT Transfer Accommodation Code & Reason subcomponent accessed
24558 ADT Transfer Destination subcomponent accessed
24566 ADT Transfer Patient Class & Service subcomponent accessed
49025 Edited Discharge Instructions
24560 ADT Transfer Date and Time subcomponent accessed
24561 ADT Transfer HTML Display subcomponent accessed
24563 ADT Transfer Old Bed Status and Service subcomponent accessed
24565 ADT Transfer Old Bed Pend with Check Box subcomponent accessed
24567 ADT Transfer Providers subcomponent accessed
17233 Media viewed in Chart Review
17380 Automatic actions performed by a BestPractice Advisory
15402 Guarantor Employer Information subcomponent accessed
15411 Guarantor Demographics subcomponent accessed

Supplemental Appendix 1: continued

17325	Episode viewed
24059	Bed Planning request opened
49012	After Visit Summary Viewed
49034	Medications accessed in Narrator
24201	Diagnoses subcomponent accessed
24202	Procedures subcomponent accessed
24227	Expected Date and Time subcomponent accessed
24520	ADT L&D Delivery Method/Multiple Birth subcomponent accessed
15632	Documents list accessed
34135	Patient device removed
72002	Care Everywhere Request Outside Records activity accessed
49016	Visit Navigator SmartForm Flowsheet section accepted
15619	Registration Research subcomponent accessed
24206	Other Provider subcomponent accessed
15655	Benefit Collection opened
34133	Patient device added
34157	Patient Education title added
60017	Delete clinical note
60013	Cosign clinical note
17824	Care Teams modified
17225	Order Entry activity accessed
15405	Guarantor Account Coverage Summary subcomponent accessed
15406	Guarantor Account Miscellaneous Info subcomponent accessed
17623	Communication Management accessed
17624	Communication Management exited
56004	Procedure event log accessed
17321	New alternative orders accepted
49024	Visit Navigator Follow-Up section accepted
48007	Prescription history activity accessed
34994	Inpatient Adjust Times activity viewed
24214	Admission Date and Time subcomponent accessed
24216	Room Assignment subcomponent accessed
24229	Service subcomponent accessed
24263	Room and Bed Needs subcomponent accessed
34015	Care Plan reviewed
41115	Universal Charge Line details viewed
17213	Pharmacy for encounter modified
24224	Arrival Date and Time subcomponent accessed
24225	Means of Arrival subcomponent accessed
17136	Episodes activity accessed
17171	Document viewed
24204	Admission Comments subcomponent accessed
17041	Allergies modified
49063	Infusion stopped in Narrator
49060	MyNote opened
15172	Medicare as Secondary Payor Part I form accessed

15173 Medicare as Secondary Payor Part II form accessed
 15174 Medicare as Secondary Payor Part III form accessed
 15037 Coverage Summary form accessed
 60008 View FastNote activity
 17319 Alternative selection form aborted
 17218 Home Medications modified
 17622 Communication printed
 41035 Coding form selected
 17784 Results acknowledged in IB
 18181 Results reviewed in IB
 15175 Medicare as Secondary Payor Part IV form accessed
 34124 Inpatient Data Validate activity accessed
 15653 MSPQ opened
 60021 PEND ATTESTATION TO CLINICAL NOTE
 17117 SmartSets activity selected
 60002 Route from inpatient notes activity
 17320 Alternative orders not taken and original order placed
 34145 MAR administration timeout
 34172 Patient Education assessment filed
 15012 Registration Coverage HTML member table accessed
 17237 Order printed from Chart Review
 24571 L&D Status for Non-L&D Workflows accessed in EPT workflow
 41014 Hospital Account History accessed
 60016 Edit attestation on clinical note
 15403 Guarantor Account Add'l Items HTML display subcomponent accessed
 1103 Patient Sidebar report accessed
 15043 Account Summary form accessed
 17210 Child order released for a procedure order
 49035 Events accessed in Narrator
 38205 Coding Query viewed
 17128 Patient Instructions navigator section accessed
 24240 Admission Source and Level of Care subcomponent accessed
 15013 Edit Claim Information window accessed
 17257 Results Review printed
 41039 Hospital Billing Tx Inquiry accessed
 17139 Clinical References activity accessed
 49014 ED Visit Navigator Arrival section accepted
 17039 Chief Complaint modified
 41114 Hospital Billing report accessed
 17252 Problem List reviewed
 34708 Flowsheet data copied forward
 49062 Order completed in Narrator
 49013 Orders accessed in Single Screen Disposition
 17208 Medication taking status modified
 15143 New Coverage (CVG) created
 15144 New hospital account created
 34023 Deleted Sign Out Report Comment
 49052 ED Triage Activity viewed

Supplemental Appendix 1: continued

17133	Contacts navigator section accessed
34128	Patient device data accessed
34702	LDA removed
24554	ADT Transfer Acuity and LOC subcomponent accessed
24556	ADT Transfer Bed Manager Info subcomponent accessed
24568	ADT Transfer Room & Bed Type subcomponent accessed
94023	Report viewed
34025	Marked Discharge Instructions as Reviewed
49075	One-Step Medication filed
49036	Staff events accessed in Narrator
24150	Encounter Events Summary accessed
41101	Hospital Account information accessed from Charge Review Workqueue
26026	Contacts navigator section viewed
38007	Deficiency Completion activity accessed
17277	Specialty comments modified
49065	Result edited in Narrator
34052	OB and Lactation Status accepted
48008	Label History Activity accessed
17103	Addendum opened
17302	Active Guidelines web integration record accessed
17330	Immunization record saved
49066	Staff updated in Narrator
15177	Medicare as Secondary Payor Part V form accessed
17312	Active Guidelines is launched
17619	Letters activity accessed
17620	Letters activity exited
17905	SmartForm saved
34158	Patient Education point added
17516	Letter status changed
41045	Hospital Billing Unit Charge Entry accessed
94005	Integration link opened
15146	View Only MSP Questionnaire accessed
17904	SmartForm viewed
34703	Existing LDA added
49040	Clinical References viewed
41092	Account Status Summary accessed
41037	Hospital Billing Liability Bucket Information accessed
94240	Patient photo taken from Haiku
17515	Letter printed
17134	Vitals navigator section accessed
17200	Research Studies activity accessed
34163	Patient Education resolved
49059	Event accepted in Narrator
56010	Procedure Event Log accessed
24247	Discharge Diagnoses subcomponent accessed
15102	Coverage Termination Questionnaire accessed

49041 Clinical References accepted
 24503 Event Management activity accessed
 34040 Education Sidebar accessed
 15142 New guarantor account created
 17207 Medication note saved
 34144 MAR printed
 49070 Patient Events Log viewed
 15654 MSPQ view only
 17032 Printing from Comm Mgt section
 17174 Document created
 17223 Image edited
 17275 Document printed
 55006 Treatment plan accessed
 15132 MSP Pull EGHP Info window accessed
 15182 Medicare as Secondary Payor Part V form page 3 accessed
 17035 Benefit plan changed
 34168 Patient Education documentation deleted
 41040 Account accessed within Combine Accounts form
 49006 ED Patient Events Log activity is accessed
 55005 Treatment plan modified
 15652 Registration subcomponent viewed
 17331 Immunization record deleted
 49021 ED Visit Navigator Disposition section accepted
 89040 LDA created via procedure documentation
 17457 Outside historical medications filed
 34125 Inpatient BestPractice Alerts navigator accessed
 15181 Medicare as Secondary Payor Part II form page 2 accessed
 17042 Allergies timed out
 17322 External link in an alternative form clicked
 17618 Letter viewed
 26030 Contacts section saved
 41094 Claim Information form accessed
 60022 Order Canceled from Note
 15176 Medicare as Secondary Payor Part IV form page 2 accessed
 17236 Encounter printed from Chart Review
 17627 FYI flag changed
 24210 Previous Admissions subcomponent accessed
 24230 ED Complaint subcomponent accessed
 34165 Patient Education marked not applicable
 34166 Patient Education marked applicable
 41095 Account Summary accessed from Enterprise Guarantor Summary
 49019 ED Visit Navigator Clinical Impressions section accepted
 49074 ED Wet Read report viewed
 15001 Add Account window opened in patient registration
 17146 Image imported
 17150 Result note filed
 17344 Immunization registry data requested
 20076 Document Information form accessed

Supplemental Appendix 1: continued

15014	Edit Claim Information window for new claim accessed
15153	Registration Hospital Account questionnaire accessed
15520	Coverage FC form HTML Summary subcomponent accessed
15521	Coverage FC form Plan and All Group Info subcomponent accessed
15522	Coverage FC form Signature and Accept Assignment subcomponent accessed
15523	Coverage FC form Payor and Sequence Number subcomponent accessed
15524	Coverage FC form Non-Managed Care Effective Dates and Coverage grid subcomponent accessed
17119	Letters activity accessed
41024	Hospital Billing Tx Details accessed
41034	Code Integration accessed
41107	Hospital Billing Bucket Information accessed
41120	Claim source accessed
53018	Patient information accessed through Edit Claim Vals

Supplemental Appendix 2: Resident Prioritized Electronic Health Record Click Types for Attribution of Primary Interns on Internal Medicine Inpatient Teams

Metric ID Number	Metric Name	Resident Circles	Resident Stars
14200	Printing occurred	9	2
17015	Problem List activity accessed	7	1
17102	Progress Notes navigator section accessed	10	5
17104	Patient Encounter opened	4	1
17106	Medications activity accessed	10	4
17108	Order list changed	8	1
17126	Allergies activity accessed	5	1
17128	Patient Instructions navigator section accessed	11	1
17134	Vitals navigator section accessed	12	4
17227	Note viewed in Chart Review	11	6
17228	Encounter viewed in Chart Review	7	1
17234	Order viewed in Chart Review	10	4
17239	Note printed from Chart Review	6	1
17280	Chart searched	5	1
20040	SmartText used	4	1
20045	SmartPhrase used	6	
24247	Discharge Diagnoses subcomponent accessed	6	
24248	Discharge Disposition and Destination subcomponent accessed	5	
24249	Discharge Comments subcomponent accessed	3	
34005	Notes viewed	10	4
34010	Notes activity and navigators accessed	4	2
34101	Inpatient MAR activity accessed	11	7
34110	Inpatient Synopsis activity accessed	3	1
34111	Flowsheet activity accessed	10	1
34122	Inpatient Sign Out Report activity accessed	13	9
34123	Inpatient Intake/Output activity selected	8	2
34127	Inpatient Orders section accessed	13	7
48001	Order History Activity accessed	5	
49005	Viewed Discharge Instructions	7	2
72002	Care Everywhere Request Outside Records activity accessed	8	1

Supplemental Appendix 3: Author Prioritized Electronic Health Record Click Types for Attribution of Primary Interns on Internal Medicine Inpatient Teams

Metric ID Number	Metric Name
14200	Printing occurred
17015	Problem List activity accessed
17040	Diagnosis association updated
17102	Progress Notes navigator section accessed
17104	Patient Encounter opened
17106	Medications activity accessed
17108	Order list changed
17126	Allergies activity accessed
17128	Patient Instructions navigator section accessed
17134	Vitals navigator section accessed
17227	Note viewed in Chart Review
17228	Encounter viewed in Chart Review
17233	Media viewed in Chart Review
17234	Order viewed in Chart Review
17239	Note printed from Chart Review
17280	Chart searched
20001	In Basket message viewed
20040	SmartText used
20045	SmartPhrase used
24210	Previous Admissions subcomponent accessed
24230	ED Complaint subcomponent accessed
24246	Discharge Date and Time subcomponent accessed
24247	Discharge Diagnoses subcomponent accessed
24248	Discharge Disposition and Destination subcomponent accessed
24249	Discharge Comments subcomponent accessed
24253	Discharge Code and Means of Departure subcomponent accessed
34005	Notes viewed
34010	Notes activity and navigators accessed
34101	Inpatient MAR activity accessed
34110	Inpatient Synopsis activity accessed
34111	Flowsheet activity accessed
34122	Inpatient Sign Out Report activity accessed
34123	Inpatient Intake/Output activity selected
34127	Inpatient Orders section accessed
34150	Order Reconciliation Section accessed

34997	Time for Inpatient Medication adjusted
48001	Order History Activity accessed
48007	Prescription history activity accessed
49001	ED Visit Navigator Arrival section accessed
49002	ED Visit Navigator Disposition section accessed
49003	Visit diagnoses modified
49005	Viewed Discharge Instructions
49008	Visit Navigator Flowsheet section is accessed
49030	ED Narrator Note accessed
49032	Orders accessed from Narrator
49033	Flowsheets accessed from Narrator
49034	Medications accessed in Narrator
49035	Events accessed in Narrator
60000	Print notes from inpatient notes activity
72002	Care Everywhere Request Outside Records activity accessed

Supplemental Appendix 4: Attribution of Primary Interns on Internal Medicine Inpatient Teams Decision Points in Model 2

Path (decision point)	Validation Data		
	Count	No	Yes
0 Root Node	642	91.1%	8.9%
1 Progress note author = Yes	72	50.0%	50.0%
2 Progress note author = No	570	96.3%	3.7%
3 PGY = 2,3	22	100.0%	0.0%
4 PGY = 1	50	28.0%	72.0%
5 PGY = 3	173	100.0%	0.0%
6 PGY = 2	192	100.0%	0.0%
7 PGY = 1	205	89.8%	10.2%
8 Resident placed orders on patient for specified date < 0.01	43	30.2%	69.8%
9 Resident placed orders on patient for specified date >= 0.01	7	14.3%	85.7%
A Click rank among residents < 1.98	50	70.0%	30.0%
B Click rank among residents >= 1.98 AND < 3.94	117	95.7%	4.3%
C Click rank among residents >= 3.94	38	97.4%	2.6%
D Click rank among residents < 1.98	37	24.3%	75.7%
E Click rank among residents >= 1.98 AND < 2.96	5	60.0%	40.0%
F Click rank among residents >= 2.96	1	100.0%	0.0%
G Discharge summary author = 1	15	33.3%	66.7%
H Discharge summary author = 0	35	85.7%	14.3%
I Number of clicks deemed to have medium importance among residents < 9.12	95	97.9%	2.1%
J Number of clicks deemed to have medium importance among residents >= 9.12 AND < 21.28	10	100.0%	0.0%
K Number of clicks deemed to have medium importance among residents >= 21.28	12	75.0%	25.0%
L Number of clicks deemed to be of high importance by physician authors < 7.3	8	12.5%	87.5%
M Number of clicks deemed to be of high importance by physician authors >= 7.3 AND < 26.28	27	22.2%	77.8%
N Number of clicks deemed to be of high importance by physician authors >= 26.28	2	100.0%	0.0%
O Number of clicks deemed to be of high importance by physician authors < 17.52	5	60.0%	40.0%
P Number of clicks deemed to be of high importance by physician authors >= 17.52 AND < 18.98	0	–	–
Q Number of clicks deemed to be of high importance by physician authors >= 18.98	0	–	–
R Number of clicks deemed to have high importance among residents < 0.29	1	100.0%	0.0%
S Number of clicks deemed to have high importance among residents >= 0.29	14	28.6%	71.4%
T Number of clicks deemed to be of high importance by physician authors < 32.12	32	87.5%	12.5%
U Number of clicks deemed to be of high importance by physician authors >= 32.12 AND < 75.92	1	100.0%	0.0%
V Number of clicks deemed to be of high importance by physician authors >= 75.92	2	50.0%	50.0%
W Number of clicks deemed to be of high importance by physician authors < 20.44	9	100.0%	0.0%
X Number of clicks deemed to be of high importance by physician authors >= 20.44 AND < 35.04	1	100.0%	0.0%
Y Number of clicks deemed to be of high importance by physician authors >= 35.04	0	–	–



Chapter 10

Discussion



I envision a system in which we promise those who depend on us total access to the help they need, in the form they need, when they need it. Our system will promise freedom from the tyranny of individual visits with overburdened professionals as the only way to find a healing relationship; will promise excellence as the standard, valuing such excellence over ill-considered autonomy; will promise safety; and will be capable of nourishing interactions in which information is central, quality is individually defined, control resides with patients, and trust blooms in an open environment. It is a new system and a lot of the old tools won't work anymore. Those who cling to their old tools and allow our organization to disintegrate will find little sense either in the burning present or in the challenging future. For them, sensemaking will have failed, and the panic of isolation will drive them up a slope that is too far and too steep for them to make it. For the rest, the possibility of invention and opportunity to make sense — new sense — will open not just routes of escape but vistas of achievement that the old order could have never imagined.¹

Donald Berwick, MD, 11th Annual National Forum on Quality Improvement in Health Care, New Orleans, Louisiana, December 7, 1999

The foundation of competency-based medical education (CBME) is ensuring the outcomes of training prepare graduates to meet the needs of populations of patients and even provide *better* care.^{2,3} These goals make aligning the outcomes of learners with the outcomes of patients a fundamental objective in CBME. However, understanding of the relationships among medical education, quality of care, and patient outcomes has advanced minimally over the past 40 years.^{4,5} Medical educators must embrace this gap, and this dissertation seeks to help fill the void in this area. A fundamental reason this gap continues to exist may be that medical education and health care systems function as separate silos. Moving forward, both health care delivery systems and medical education must align their purposes and goals in order to prepare trainees for the primary aim of meeting societal needs.⁶⁻⁸ Sklar and colleagues have proposed a bold goal to better align medical education and health care delivery: medical education should adopt the Triple Aim.⁷ Coined by Berwick and colleagues, the Triple Aim claims the overall goals of a health care system are threefold: 1) improve the individual experience of care, 2) improve the health of a population, and 3) reduce per capita costs of care.⁹ While the final goal is perhaps most germane to the high-cost American context that the Triple Aim arises from, containing health care costs is a reasonable goal for any health care system. Thus, the Triple Aim, perhaps more simply referred to as “high value care,” applies broadly.¹⁰ To achieve high value care, we need a new way of sensemaking in health care that leads to “routes of escape” from the current state that lead to “vistas of achievement” that have as yet not been imagined.¹ Medical education can, and should, lead this work. As chapters 1 and 2 note, a concentration on new, patient-focused means of performance assessment is a path forward in these efforts. Chapters

4 through 8 take steps in this direction through detailing the development, implementation, and initial validity evidence for resident sensitive quality measures (RSQMs). However, chapters 3 and 9 present logistical considerations in this work, including the importance of distinguishing attribution from contribution and harnessing the electronic health record for the purposes of extracting practice data.

This chapter details a vision for the role of medical education in achieving high value health care. After proposing that medical education should lead this charge, we turn our attention to the role that patient-focused assessment plays in this goal and discuss major findings from the studies presented in this dissertation in light of this ambition and current literature. Finally, we discuss implications for future research and practice as well as strengths and limitations of the approach taken in this dissertation.

A Vision for Medical Education: Embracing Adaptive Challenges and Igniting an Escape Fire

Achieving high value health care is an adaptive challenge, not a technical problem.^{11,12} A technical problem is one in which the problem and solution are clear, the work will be managed by an expert, and the focus is on expediting a fix for the problem. A child presenting to the emergency department with a head laceration is an example of a technical problem. The problem is a laceration, and the solution is suturing the laceration. This will be done by an emergency medicine physician, who does this work routinely, and the focus will be on doing this as quickly as possible on the day the child sustained the injury. An adaptive challenge, on the other hand, does not present a clear problem or solution. In fact, adaptive challenges can be difficult to identify (and easy to deny). Solving adaptive challenges requires experimentation and making new discoveries. This work cannot be done by the edict of experts but instead requires stakeholders' investment in addressing the challenge. This work requires changes in beliefs, roles, relationships, and approaches.

Achieving high value health care is an adaptive challenge. Berwick has suggested that health care needs an *escape fire* to make sense from its current state and provide a path forward.¹ Traditionally, an escape fire is an area of grass or forest that is purposefully set on fire to provide a safe area of protection from an approaching wildfire. This concept has been extended beyond firefighting, where an escape fire is an improvisational solution to a problem that cannot be solved through traditional methods, i.e., a solution to an adaptive challenge.¹

The traditional escape fire approach was first used by a man named Wag Dodge in the Mann Gulch (Montana, United States) fire of 1949. It is worth understanding the basics of this fire to

understand its application to health care and performance assessment in medical education in particular. Berwick recounts the story as follows:

Twenty miles north of Helena, Montana, the Missouri River flowing north cuts into the eastern slope of the Rocky Mountains.... Lewis and Clark passed through this spectacular formation...and named it "Gates of the Mountains." Two miles downriver from the Gates, a small, two-mile-long canyon runs down to the Missouri from the northeast. This is Mann Gulch. It is the site of a tragedy: the Mann Gulch Fire.... On the afternoon of August 4, 1949, a lightning storm started a small fire near the top of the southeast ridge of Mann Gulch.... The fire was spotted the next day.... Fifteen [smokejumpers, who fight wildfires]...parachuted to the head of the gulch at 4:10 P.M.... They were joined on the ground by a fireguard, who had spotted the fire. Otherwise, the smokejumpers were isolated from the outside world. The smokejumpers were a new organization... reinventing the approach to forest fire containment — aggressive, highly tactical, and coordinated. To them, the Mann Gulch fire...appeared routine. They were wrong. The first reconnaissance team headed down the south side of the gulch. The foreman, Wag Dodge, became worried that the group could get trapped on that side. He ordered them to come back and cross with the rest of the men to the north side of the gulch, opposite the fire, and head down the hill so that the river, an escape route, would be at their backs.... Dodge was the first to spot the impending disaster — the fire had jumped the gulch from the south side to the north. It had ignited the grass only two hundred yards ahead of the lead smokejumpers, blocking their route to the river. Now a race began. Dodge knew that the grassfire would cut off the route to the river and would head swiftly up the north slope toward the firefighters. He ordered the group to reverse course immediately and head back up the slope toward the ridge crest, hoping to get over it before the fire did. The north slope of Mann Gulch is steep — a 76 percent slope on the average.... It is hard even to walk up such a slope, but these young men were trying to run up it. Add air one hundred degrees at the start and superheated by the rushing fire, add the poor visibility from smoke and airborne debris, add the weight of the packs and tools these men were taught never to drop, and add their inexperience with the pace and heat of grassfires — far hotter and moving a lot faster than fires in forests [they were trained for]. At 5:45 P.M...the fire was traveling toward them at...1.4 miles an hour. Ten minutes later...it was traveling at...7 miles an hour. Wag Dodge knew they would lose the race to the top. With the fire barely two hundred yards behind him, he did a strange and marvelous thing. He invented a solution. On the spot. His crew must have thought he had gone crazy as he took some matches out of his pocket, bent down, lit a match, and set fire to the grass directly in front of him. The fire spread quickly uphill ahead of him and he stepped into the middle of the newly burnt area. He called to his crew to join him as he lay down in the middle of the burnt ground. Dodge had invented what is now called an

“escape fire,” and soon after Mann Gulch it became a standard part of the training of all Forest Service firefighters. But on August 5, 1949, no one followed Wag Dodge.... The fire raged past Wag Dodge and overtook the crew. Only three made it to the top of the ridge.¹

In the Mann Gulch fire, one man, placed in a situation of chaos and crisis, improvised a non-traditional solution that provided the way out of his perilous situation. The answer was among them the whole time, but it took his ingenuity to see it. Those not in his escape fire were challenged to make it to the top of the ridge not only because the slope was too far and too steep but because they were also being weighed down by the tools they had been taught to never drop. Berwick argues that we need an escape fire to achieve high value care. The slope to reach this goal is too far and too steep given our current approach. Furthermore, we are weighed down by clinging to our current tools. We need an escape fire that allows a new way of sensemaking to emerge.

Can medical education provide the escape fire for health care?

Should medical education provide the escape fire for health care?

Answering these questions requires a consideration of what drives health care. In 2019, this is not medical education. But, there is no reason that it cannot be. The pillars of medical education, research, and patient care exist in every modern health care system. An escape fire for health care will likely come from one of them. However, because improving health care is not a technical problem but an adaptive challenge, solving the current issues will require problem solving that engages stakeholders from all three pillars for transformational changes. However, one pillar will need to move first and bring the others on board. In this work, the definitions of what these entities entail will need to change. Patient care cannot mean “profit source” as it does in some health care systems. Rather, it will need to be defined as providing care that is safe, effective, patient-centered, timely, efficient and equitable and engaging the patient in defining what that care looks like.¹³⁻¹⁵ Research cannot be defined as advancing science for the sake of science. Rather, it will need to be defined as advancing knowledge for the sake of populations of patients. Finally, medical education cannot be defined as advancing theory and defining new ways of training. Rather, it will need to be defined as continually preparing providers to meet the needs of patients. ***In short, a focus on the patient will be foundational to starting the escape fire.***

While solving the adaptive challenge of achieving high value health care will require collaborative efforts of key stakeholders, a single entity will need to light the match for the escape fire. Who will be the “Wag Dodge” to step up? Can this be the medical education community?

Patient-Focused Performance Assessment: Kindling Medical Education's Escape Fire

Medical education can take several measures to place focus on the patient. Stated simply, everything that medical education does should have the patient in mind. This has not been true of medical education's past. Medical education has placed extensive focus on the advancement of theory, with calls to increase, rather than decrease, this focus in recent years.¹⁶⁻¹⁸ This focus is rightly placed. Without theory, medical education has no skeleton enabling it to stand. However, in the future, without a focus on the patient, medical education will become superfluous and irrelevant. It risks being engulfed by the wildfire of modern health care.

Research efforts in medical education are increasingly placing focus on patients. Any one of these efforts could serve as the match to light the escape fire. Given the focus of this dissertation, we will detail how patient-focused assessment can serve this role. This is an apt focus because medical education has a tendency to cling to existing assessment approaches, sometimes holding them in higher esteem than the constructs they represent.^{19,20} Moreover, traditional assessment often places more focus on the learner's ability to do something that is deemed important to the learner or to the assessor and often not to the patient.³ Patient-centered assessment approaches, such as RSQMs and entrustment that were studied in this dissertation, can shift this focus. But, this will require a willingness to move from traditional approaches of assessment to include new ones that are more sensitive to the goal of achieving performance assessment that matters for patients. In this way, we are presented with the same challenges as the smokejumpers on Mann Gulch.

Why, 20 years after his call for an escape fire, has no one taken Berwick up on this charge? Like the smokejumpers, we may not want to, or cannot conceive being able to, drop our tools that may be holding us back from reaching our goals. Since the Mann Gulch fire, multiple firefighters have been ordered to drop their tools, refused, and been killed by fire with their tools beside them, not far from safe zones they could have reached had they been lighter and faster without their tools.¹ Why would they not drop their tools, and what can we learn for assessment in medical education from their inaction? Some of the reasons may be unique to firefighters. As Weick notes in his commentary on Berwick's escape fire speech, fires are loud and firefighters may not hear orders to drop their tools; fighting fire is mentally and physically taxing and they may be exhausted and not thinking clearly; and sometimes survival comes by using tools to clear the underbrush and set up a fire shelter, underscoring the importance of retaining one's tools.¹

Weick also details reasons why smokejumpers may not drop their tools that bear an eerie resemblance to health care and assessment in medical education. First, firefighters may not drop their tools because they have little familiarity with alternatives, creating a perceived risk.

Second, dropping one's tools may be seen as admitting failure. These are common reasons traditional assessment tools are retained in medical education as well. They become what we know and depend on. Complicating this, there is perceived power in psychometrics.²¹ How else can we continue to ignore the mounting evidence for qualitative assessment data?²²⁻²⁵ Third, social dynamics may lead to retaining tools. If no other firefighters are dropping their tools, no one wants to be the firefighter without tools once they return to their base. In medical education performance assessment, if everyone is using verbatim milestones to assess residents on the front line even against advice not to do so, there is likely social pressure to also use verbatim milestones.²⁶ Fourth, firefighters may not know how to drop their tools. In testimony following the Mann Gulch fire, a surviving smokejumper told of a colleague running for his life...with a shovel in hand. The survivor providing the testimony grabbed the shovel...and found a tree to gently lean it against. We sometimes treat medical education assessment tools in the same manner. We become wedded to them. Finally, Weick notes

It is perhaps oddest of all to imagine that the firefighters didn't drop their tools because they didn't have any. But that's what I suspect happened. And that's what I suspect Berwick fears is happening in medical circles. Fire suppression and medical work call for capabilities that involve lives, people, even perhaps a calling. Firefighting tools...are named after famous firefighters, they are designed solely for firefighting, and their skillful use is the mark of a seasoned firefighter and central to that person's identity. The fusion of tools with identities means that under conditions of threat it makes no more sense to drop one's tools than it does to drop one's pride or one's sense of self.¹

This phenomenon of identity leading to not moving to new ways to avoid identity loss has been described by Scharmer in the change management literature.²⁷ These reasons for not abandoning one's tools, whether as a smokejumper or as a medical educator assessing learners' performance, are deeply rooted in culture, familiarity, and identity. It is logical that abandoning them may feel very uncomfortable, even blasphemous. However, the slope to high value health care is too far and too steep, and traditional assessment tools may be weighing us down. ***We need to lighten our load and focus on the patient. Patient-focused assessment approaches, including RSQMs, provide this focus. However, they face undeniable threats if we cling to our traditional assessment approaches.***

Why RSQMs May Be Part of the Escape Fire and Not Consumed in the Surrounding Blaze

While assessment in medical education has long been entrenched in psychometrics, we are entering a post-psychometric era.²²⁻²⁵ As the "subjective and collective" find a home, and

possibly become the head of household, in assessment, new ways of envisioning assessment can emerge. Given the simultaneous calls to focus assessment on what matters for patients,²⁸⁻³¹ RSQMs may find a ready home.

Our main research aim in this dissertation was to determine the characteristics of developing and implementing RSQMs, quality measures that are largely attributable to residents and important to patient care. Chapters 4 and 5 detail the development of RSQMs through engaging consensus group methodology to first brainstorm candidate measures and then prioritize and select measures from that longer list for inclusion as final measures. In chapter 5, we present data about how the resident and supervisor stakeholder groups we engaged in this work each had blind spots that were seen by the other group. This underscores the importance of engaging multiple stakeholder groups in developing RSQMs. Chapters 6 and 7 then detail the implementation of RSQMs in the actual clinical learning environment. There were three noteworthy findings when RSQMs were implemented. First, they largely applied, but exceptions where residents were not given the chance to meet them were found. Common exceptions included the following: 1) some asthma measures did not apply when patients presented through the trauma/medical resuscitation suite, 2) discharge measures for all three illnesses only applied when patients were not admitted, 3) some medications for asthma were not given in all encounters (e.g., ipratropium), and 4) patients arriving from an outside hospital sometimes already had a work-up that obviated RSQMs (e.g., head imaging in closed head injury). Second, RSQMs displayed a wide range of completion across residents and encounters. Fewer than 10 measures (of 65 total) were met in most or all encounters, and several measures were met in less than half of the encounters included. Finally, addressing our secondary research aim considering the association between RSQMs and entrustment decisions made about residents, only RSQMs for asthma had a positive association with entrustment decisions. However, RSQM composite scores were significantly higher for residents entrusted with unsupervised practice or the ability to supervise others than for residents entrusted to lower supervision levels for all conditions (asthma, closed head injury, and all conditions combined) other than bronchiolitis. This offers validity evidence for RSQM composite scores; however, the absolute difference is modest.

As detailed in chapter 8, clinical competency committee members used RSQMs to some extent in the entrustment decision they made about the resident they reviewed. This is promising for two reasons. First, it suggests that faculty will use RSQMs. Thus, RSQMs may help fulfill a vision set forth by Wong and Holmboe, who have argued that faculty can serve a key role in better aligning educational environments with clinical environments to improve educational and clinical outcomes.³² To do this, they argue that both educational and clinical outcomes must focus on the patient, as we propose in this dissertation. In their examples comparing traditional, learner-centered faculty perspectives with their proposed, patient-centered faculty perspective,

they highlighted faculty completing end-of-rotation assessment forms based on their individual interactions with residents as an example of the former. Their proposed patient-centered model would have faculty complete such forms based on individual interactions with the resident but also direct observation of interactions between the resident and patients/family members, feedback from other members of the interprofessional team, and *patient outcome data*—such as RSQMs. They also gave the example of faculty traditionally receiving quality data that are not discussed with residents and the need to transform this practice so that faculty *and residents* receive such data *and review and discuss these data regularly as a team*. RSQMs are an ideal candidate for this purpose.

The second promising reason for suggesting that clinical competency committee members will use RSQMs in making entrustment decisions is one of the findings we present in chapter 7—RSQMs may be beneficial in informing better entrustment decisions. With traditional assessment efforts presenting variability and fallibility in performance judgments between assessors as well as within individual assessor's work,^{25,33-37} some have wondered if entrustment decisions are just another imperfect assessment framework.³⁸ While expert judgment is important, informing this expert judgment with objective data, as Wong and Holmboe call for, would likely be beneficial.³² Based on our findings, it seems that RSQMs may provide such objective data.

RSQMs may also provide a spark because quality metrics for attending physicians have become commonplace,^{39,40} and payment models are starting to move in the direction of reimbursing for care provided based on these metrics.⁴¹ Thus, there is a need for more, and better, quality measures in health care. As Lindenauer and Shojania noted in 2008, “the urgent need to improve now applies not only to health care quality but to the quality measures themselves.”⁴² In addition to being measures of resident performance, RSQMs are also measures of patient care quality that are desired. The advantage of patient-focused performance assessment approaches, such as RSQMs, are that they provide just as much information about a learner's performance as they provide about the quality of a patient's care.

RSQMs will also be needed if, as Warm and Mathis have suggested, future iterations of accreditation standards for residency training include measuring and improving patient care outcomes.⁴³

Given the need to bridge the silos of medical education and quality improvement to link educational outcomes in graduate medical education to relevant clinical outcomes,^{44,45} RSQMs can also enhance meaningful involvement of residents in quality improvement work and provide insights into the quality of care provided by residents for use in their personal improvement efforts.⁴⁶ To this end, Gupta and Arora detail a three-pronged approach to better align the silos of health systems focused on quality and education: 1) support leaders who bridge these silos

because such individuals are scarce, 2) engage residents in the actual work of promoting value-based care because they provide much direct patient care, and 3) engage residents in projects that are in their scope, align with health system goals, and improve patient care.⁴⁷ It is reasonable to posit that RSQMs could be included in the final two parts of this vision.

Finally, RSQMs may not only survive but thrive for the reason stated in the title of chapter 2—the future of high quality care depends on better physician performance assessment, namely, one focused on the patient. Given the literature presented in this chapter and throughout the dissertation, this may no longer be a belief we need to promote. It may be accepted as a truth to get us to the future state of health care being called for over the past few decades.

Development and Implementation of RSQMs: Future Research and Practice Implications

Scale Considerations

This dissertation embarks on uncharted territory in medical education assessment. RSQMs are a novel patient-focused assessment strategy. Thus, considerable work remains to scale this innovation. First, RSQM development to date has been limited to the specialty of pediatrics and specifically pediatric emergency medicine. Moving forward, we need to use what has been learned to expand to other settings within pediatrics as well as to other specialties. Next steps within pediatrics as well as considerations for how RSQMs may look similar and different in those settings were explored in chapter 5. The study presented there suggests that the outpatient general pediatrics clinic as well as hospital medicine are the next settings to move to within pediatrics. Beyond this, moving to other specialties, including other medical specialties (e.g., internal medicine, neurology), surgical specialties (e.g., general surgery, orthopedics), and hospital-based specialties (e.g., anesthesiology, radiology), will be important to determine how RSQMs are similar and different in those specialties.

Relevant to efforts to scale, while the RSQM development process described in chapters 4 and 5 is not unwieldy, it does present a modest level of effort over a few months to develop each new set of RSQMs for a specialty area. It also requires engaging stakeholders in this process who may or may not be interested in participating based on other demands on their time. Therefore, future expansion will need to consider how RSQM development efforts can effectively engage stakeholders, which will need to also include nurses, patients, and family members, in other settings and specialties on both a local and national level without unintentionally making the process too arduous to complete.

Association With Entrustment Considerations

This dissertation begins to explore the association between RSQMs and entrustment decisions. Given the international adoption of entrustable professional activities (EPAs) over the past decade,⁴⁸⁻⁵⁸ more work should be done to explore this area. This includes establishing further validity and reliability evidence for entrustment decisions, which is widely lacking even beyond this dissertation.^{56,58-60} Our preliminary findings in this area, presented in chapter 7, are provocative. We found a statistically significant, but modest, positive relationship between entrustment decisions and RSQM composite scores for asthma but not for bronchiolitis or closed head injury. However, asthma RSQM composite scores only increased 0.03 (out of 1) for every 1 (out of 7) point increase in entrustment decision scores. While statistically significant, this is likely not educationally or clinically significant. We would expect a positive relationship between RSQM composite scores and entrustment decisions across all entrustment decisions and conditions considered. This discrepancy raises the question of whether RSQMs may be better suited than entrustment decisions to assess resident performance in a manner that matters for patients. At the very least, it begs the question of whether a combination of RSQMs as well as entrustment, or other assessment, decisions around aspects of care not measured by RSQMs may be needed.

While we believe the entrustment framework is a long-needed patient-focused assessment framework, it may not be optimized yet. In fact, assessment efforts using an entrustment framework have been shown to suffer from the same “frame of reference” problems that other assessment efforts face. Namely, supervisors use self as the frame of reference in determining what level of supervision the learner needs rather than what the patient needs to receive safe, effective, patient-centered care.⁶¹ Furthermore, the “self” that is used for comparison could be the current clinical skills of the assessor, what the assessor was capable of at the same level of training as the learner being assessed, or some other reference.³⁴ Realizing the spectrum of clinical skills of supervisors alone gives pause to the hope of ever calibrating raters.

If entrustment decisions do not align with the delivery of expert-informed optimal care, which the findings of chapter 7 call into question, other measures to determine whether patients receive such care must be sought. With the noted caveat that RSQMs may not be sufficient to stand to assess resident performance, the range of RSQM composite scores suggests their potential promise for discerning performance better than traditional assessment efforts that suffer from more range restriction, making RSQMs a ready candidate. Furthermore, given their primary focus on the provision of optimal care to patients, RSQMs could perhaps be a large, or even primary, form of assessment for residents to inform entrustment decisions. This could lead to entrustment decisions that are associated with measures of quality care, which is unquestionably the desired outcome.

Construct Considerations

EPAs provide a lens of supervision and the opportunity to assess learners using a construct-aligned scale,⁶² reflecting the priorities and expertise of assessors who supervise trainees in the clinical

learning environment. These types of scales have been shown to be more reliable than other global scales in work-based assessment efforts.^{63,64} Are RSQMs another patient-focused assessment framework that may be construct aligned? Future research should investigate this area.

Attribution Considerations

The focus of this dissertation was intentionally on quality measures that could be largely *attributed* to residents. However, as chapter 3 details, quality measures span a range from individual attribution to the opposite extreme where countless individuals and entities contribute to the outcome of interest. An example from this opposite extreme might be early mortality of a patient with diabetes mellitus. This outcome will be determined by a plethora of variables arising from the patient, multiple members of the health care team, aspects of the health care system, and even the neighborhood community of the patient.⁶⁵ The portions of the continuum between these extremes will be important to learn more about. Sebok-Syer and colleagues are currently beginning to explore this area, seeking to determine the spectrum of interdependence in emergency medicine performance metrics for residents.⁶⁶ For example, they have found that the metric of “turnaround time between sign-up [for a patient] and first orders” is felt to be 94% attributable to the resident, ordering of lab tests and medications are at or near 50%, time to fluids is only 16% attributable to the resident, and length of stay is 0% attributable to the resident. Expanding this work in the future will be important not only for further understanding the interdependence of care, an interest of Sebok-Syer and colleagues,⁶⁷ but also for further understanding and developing RSQMs to determine which quality metrics are, and are not, largely attributable to individual residents.

Utility Considerations

In chapter 8, we found that faculty serving on the clinical competency committee largely use RSQMs to some extent in making entrustment decisions, and we also heard that they believe RSQMs would provide meaningful feedback to residents on their performance. Feedback measures based on care provided for the purposes of continuous quality improvement are greatly needed in medical training. Residents report suboptimal training in and understanding of quality improvement as well as the perception that they do not contribute meaningfully to quality improvement efforts.^{68,69} Furthermore, many traditional quality measures are not aligned with the work residents complete, and few efforts have tried to address this gap.⁷⁰⁻⁷⁴ We believe these circumstances send a strong message that we must embrace any efforts to provide residents with quality metric feedback about their performance and then subsequently determine if such feedback leads to improved RSQM performance. Feedback can be hard to come by in training, even in the advent of competency-based assessment that places focus on the importance of direct observation and real-time feedback.⁷⁵⁻⁷⁷ RSQMs capture whether residents are completing components of their work that have been deemed important to patient care by fellow residents and supervisors. Moreover, these measures demonstrate a wide range of performance

(composite measure scores of 0.47 to 1 [out of 1] for asthma, 0.35 to 0.91 for bronchiolitis, and 0.44 to 0.89 for closed head injury), as detailed in chapter 6. Almost no residents in this study attained all possible RSQMs in their patient encounters and thus RSQMs inform residents with feedback about care components they are not providing that have been deemed important to patient care. Future work should explore how RSQMs can be optimally used in this fashion.

Quality metrics for attending physicians have become commonplace, and payment models are starting to move in the direction of reimbursing for care provided based on these metrics.³⁹⁻⁴¹ The effects of such audit and feedback have been shown to be small to moderate but can be improved through practical approaches, including providing feedback regularly, addressing barriers to feedback use, addressing credibility of the data provided, preventing defensive reactions to the data provided, and constructing feedback through social interaction.^{78,79} Taking these considerations into account will be important when implementing RSQM feedback to residents. However, there is a large potential unintended consequence of this practice to be considered as well. Namely, residents and their supervisors may focus only on what can be measured and not on those things that are equally, or more, important but not as easily measured.⁸⁰ As Deming noted: “It is wrong to suppose that if you can’t measure it, you can’t manage it — a costly myth.”⁸¹ Measuring performance in some areas may diminish intrinsic motivation to perform well in other areas.^{40,82} This loss of motivation could become an unintended consequence among residents, making it important to monitor and address this as RSQMs are implemented.

Funding Considerations

Finally, as chapter 2 notes, with only 2% of global health expenditure dedicated to health education for all professions combined,³ the further development, implementation, and study of RSQMs should receive funding and resource prioritization from public and private sectors.

Strengths and Limitations

Perhaps the greatest strength of RSQMs is that they integrate patient care quality with the educational process through assessment, providing a novel direction for resident performance assessment that places focus on the patient and could be a spark that combines with others to start an escape fire for improving health care. Their focus on individuals is both a strength and a limitation. Health care is provided by teams, presenting a possible limitation. However, we currently graduate, certify, and credential individuals. This is not likely to change in the foreseeable future, making the focus on individuals a strength as well. Furthermore, focusing on individuals may be the most effective way to drive their personal improvement,^{83,84} and health care systems are currently considering and utilizing managed care payment models based on

demonstrated quality of care provided by individuals.^{40,41} As we note in chapter 3, medical education and health care need to place focus on individuals as well as teams, programs, and systems for these reasons. The focus on individuals presents another perceived limitation in the eyes of some. Namely, RSQMs are by definition process measures when quality improvement often advocates for outcomes measures, or at least intermediate measures between these two extremes. However, despite this desire, process measures are largely what exist in the quality improvement community at this time as well, and process measures, compared to outcome measures, have the strength of being actionable for providers.^{39,85} As noted previously, the development of RSQMs to date has not engaged all stakeholders, such as nurses, patients, and families. Future efforts will need to do this. Patients and families, in particular, will be important to engage.⁸⁶ Finally, further validity evidence is needed both for RSQMs as well as entrustment decisions, and future work should prioritize gathering of evidence for both RSQMs and entrustment.

Medical Education Is at a Fundamental Crossroads

If we hope to realize our quest for high value health care, an escape fire is inevitable. Medical education needs to provide the spark by placing a clear, undeniable, and sustained focus on the patient in everything it does. Patient-focused assessment strategies, such as RSQMs, can kindle this flame.

However, threats exist. Some do not recognize the far and steep slope to high value health care, clinging to old tools with a reluctance to adopt new ones, and calling for the need to decouple clinical education from clinical care.^{87,88} These individuals might be best served by leaning into the words of Lao Tzu: "In pursuit of knowledge, every day something is acquired; in pursuit of enlightenment, every day something is dropped (quoted in Wieck⁸⁹)."

Despite the noise of challengers, others in the medical education community are making a clear call for medical education to lean into its responsibility to place the patient first and align its goals with those of health care at large—high value care.⁷

Which vision for the future will prevail? We believe that patient-focused assessment will be realized. When it is, as Berwick notes

A lot of the old tools won't work anymore. Those who cling to their old tools...will find little sense either in the burning present or in the challenging future. For them, sensemaking will have failed, and the panic of isolation will drive them up a slope that is too far and too steep for them to make it. For the rest, the possibility of invention and opportunity to make sense — new sense — will open not just routes of escape but vistas of achievement that the old order could have never imagined.¹

References

1. Berwick D. *Escape Fire: Lessons for the Future of Health Care*. New York, NY: The Commonwealth Fund; 2002.
2. Hall JN, Lloyd T. Defining the target: Competency-based medical education for better patient care. *Med Teach*. 2018;1.
3. 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.
4. Magraw RM, Fox DM, Weston JL. Health professions education and public policy: A research agenda. *J Med Educ*. 1978;53(7):539–546.
5. Weinstein DF. Optimizing GME by measuring its outcomes. *N Engl J Med*. 2017;377(21):2007–2009.
6. Roemer BM, Azevedo T, Blumberg B. Looking at graduate medical education through a different lens: A health care system's perspective. *Acad Med*. 2015;90(9):1231–1235.
7. Sklar DP, Hemmer PA, Durning SJ. Medical education and health care delivery: A call to better align goals and purposes. *Acad Med*. 2018;93(3):384–390.
8. Busari JO. Comparative analysis of quality assurance in health care delivery and higher medical education. *Adv Med Educ Pract*. 2012;3:121–127.
9. Berwick DM, Nolan TM, Whittington J. The Triple Aim: Care, health, and cost. *Health Affairs*. 2008;27(3):759–769.
10. Porter ME. What is value in health care? *N Engl J Med*. 2010;363(26):2477–2481.
11. Heifetz R, Linsky M. *Leadership on the Line, With a New Preface: Staying Alive Through the Dangers of Change*. Boston, MA: Harvard Business Review Press; 2017.
12. Heifetz RA, Laurie DL. The work of leadership. *Harv Bus Rev*. 1997;75(1):124–134.
13. Baker A. *Crossing The Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
14. Batalden M, Batalden P, Margolis P, et al. Coproduction of healthcare service. *BMJ Qual Saf*. 2016;25(7):509–517.
15. Health Affairs. Health Policy Brief: Patient Engagement. https://http://www.healthaffairs.org/doi/10.1377/hpb20130214.898775/full/healthpolicybrief_86.pdf. Published 2013. Accessed September 19, 2019.
16. Prideaux D, Spencer J. On theory in medical education. *Med Educ*. 2000;34:888–889.
17. Rees CE, Monrouxe LV. Theory in medical education research: How do we get there? *Med Educ*. 2010;44:334–339.
18. Laksov KB, Dornan T, Teunissen PW. Making theory explicit—An analysis of how medical education research(ers) describe how they connect to theory. *BMC Med Educ*. 2017;17:18.
19. Hanson JL, Rosenberg AA, Lane JL. Narrative descriptions should replace grades and numerical ratings for clinical performance in medical education in the United States. *Frontiers in Psychology*. 2013;4:668.
20. Whitehead CR, A K. Faith-based medical education. *Adv in Health Sci Educ*. 2017;22:1–3.
21. Schuwirth LW, van der Vleuten CP. Changing education, changing assessment, changing research? *Med Educ*. 2004;38(8):805–812.

22. Hodges B. Assessment in the post-psychometric era: Learning to love the subjective and collective. *Medical teacher*. 2013;35(7):564–568.
23. Ten Cate O. The false dichotomy of quality and quantity in the discourse around assessment in competency-based education. *Adv Health Sci Educ Theory Pract*. 2015;20(3):835–838.
24. Ginsburg S, van der Vleuten CP, Eva KW. The hidden value of narrative comments for assessment: A quantitative reliability analysis of qualitative data. *Acad Med*. 2017;92(11):1617–1621.
25. Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract*. 2007;12(2):239–260.
26. Holmboe ES, Edgar L, Hamstra S. The Milestones Guidebook. Accreditation Council for Graduate Medical Education. <https://www.acgme.org/Portals/0/MilestonesGuidebook.pdf>. Published 2016. Accessed September 19, 2019.
27. Scharmer CO. *Theory U: Leading From the Future as It Emerges*. San Francisco, CA: Berrett-Koehler Publishers, Inc; 2009.
28. ten Cate O. Entrustment decisions: Bringing the patient into the assessment equation. *Acad Med*. 2017;92(6):736–738.
29. Norcini JJ. The role of assessment in supporting the movement toward patient-centred care. *Perspect Med Educ*. 2017;6(1):5–6.
30. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018:1–8.
31. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med*. 2014;89(5):721–727.
32. 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.
33. Govaerts M, Schuwirth LWT, van der vleuten CPM, Muijtjens AMM. Workplace-based assessment: Effects of rater expertise. *Adv in Health Sci Educ*. 2011;16:151–165.
34. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: A conceptual model. *Med Educ*. 2011;45(10):1048–1060.
35. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: Rethinking the etiology of rater errors. *Acad Med*. 2011;86(10):S1–S7.
36. Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the ‘black box’ differently: Assessor cognition from three research perspectives. *Med Educ*. 2014;48(11):1055–1068.
37. Eva KW. On the generality of specificity. *Med Educ*. 2003;37(7):587–588.
38. Hatala R, Ginsburg S, Hauer KE, Gingerich A. Entrustment ratings in internal medicine training: Capturing meaningful supervision decisions or just another rating? *J Gen Intern Med*. 2019;34(5):740–743.
39. Berenson R, Pronovost P, Krumholz H. Achieving the potential of health care performance measures: Timely analysis of immediate health policy issues. Robert Wood Johnson Foundation. <https://www.rwjf.org/en/library/research/2013/05/achieving-the-potential-of-health-care-performance-measures.html>. Published 2013. Accessed August 23, 2016.

40. Berenson R, Kaye D. Grading a physician's value—The misapplication of performance measurement. *N Engl J Med*. 2013;369:2079–2081.
41. Berenson R. 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.
42. Lindenauer PK, Shojania KG. The era of big performance measurement: Here at last? *Jt Comm J Qual Patient Saf*. 2008;34(6):307–308.
43. Warm EJ, Mathis BR. Ambulatory education: Time to move from process to outcome. *J Grad Med Educ*. 2019;11(2):143–145.
44. 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;92(12):1740–1748.
45. Sargeant J, Wong BM, Campbell CM. CPD of the future: A partnership between quality improvement and competency-based education. *Med Educ*. 2018;52(1):125–135.
46. Smirnova A, Sebok-Syer SS, Chahine S, et al. Defining and adopting clinical performance measures in graduate medical education: Where are we now and where are we going? *Acad Med*. 2019;94(5):671–677.
47. Gupta R, Arora VM. Merging the health system and education silos to better educate future physicians. *JAMA*. 2015;314(22):2349–2350.
48. ten Cate O, Scheele F. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542–547.
49. ten Cate O, Snell L, Carraccio C. Medical competence: The interplay between individual ability and the health care environment. *Med Teach*. 2010;32:669–675.
50. Carraccio C, Englander R, Gilhooly J, et al. Building a framework of entrustable professional activities, supported by competencies and milestones, to bridge the educational continuum. *Acad Med*. 2017;92(3):324–330.
51. Caverzagie KJ, Cooney TG, Hemmer PA, Berkowitz L. The development of entrustable professional activities for internal medicine residency training: A report from the Education Redesign Committee of the Alliance for Academic Internal Medicine. *Acad Med*. 2015;90(4):479–484.
52. Chen HC, van den Broek WS, ten Cate O. The case for use of entrustable professional activities in undergraduate medical education. *Acad Med*. 2015;90(4):431–436.
53. Englander R, Flynn T, Call S, et al. Toward defining the foundation of the MD degree: Core entrustable professional activities for entering residency. *Acad Med*. 2016;91(10):1352–1358.
54. Hart D, Franzen D, Beeson M, et al. Integration of entrustable professional activities with the milestones for emergency medicine residents. *West J Emerg Med*. 2019;20(1):35–42.
55. Hauer KE, Kohlwes J, Cornett P, et al. Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ*. 2013;5(1):54–59.
56. Meyer EG, Chen HC, Uijtdehaage S, Durning SJ, Maggio LA. Scoping review of entrustable professional activities in undergraduate medical education. *Acad Med*. 2019;94(7):1040–1049.

57. O'Dowd E, Lydon S, O'Connor P, Madden C, Byrne D. A systematic review of 7 years of research on entrustable professional activities in graduate medical education, 2011–2018. *Med Educ.* 2019;53(3):234–249.
58. Shorey S, Lau TC, Lau ST, Ang E. Entrustable professional activities in health care education: A scoping review. *Med Educ.* 2019;53(8):766–777.
59. Hauer KE. Seeking trust in entrustment: Shifting from the planning of entrustable professional activities to implementation. *Med Educ.* 2019;53(8):752–754.
60. Mink RB, Schwartz A, Herman BE, et al. Validity of level of supervision scales for assessing pediatric fellows on the common pediatric subspecialty entrustable professional activities. *Acad Med.* 2018;93(2):283–291.
61. Apramian T, Cristancho S, Sener A, Lingard L. How do thresholds of principle and preference influence surgeon assessments of learner performance? *Ann Surg.* 2018;268(2):385–390.
62. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: Outlining their usefulness for competency-based clinical assessment. *Acad Med.* 2016;91(2):186–190.
63. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: Construct alignment improves the performance of workplace-based assessment scales. *Med Educ.* 2011;45(6):560–569.
64. Crossley J, Jolly B. Making sense of work-based assessment: Ask the right questions, in the right way, about the right things, of the right people. *Med Educ.* 2012;46(1):28–37.
65. Beck AF, Anderson KL, Rich K, et al. Cooling the hot spots where child hospitalization rates are high: A neighborhood approach to population health. *Health Aff.* 2019;38(9):1433–1441.
66. Sebok-Syer S, Shepherd L, McConnell A, et al. “EMERGING” electronic health record data metrics: Insights and implications for assessing residents’ clinical performance in emergency medicine. [Unpublished, 2019].
67. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: implications for assessment and entrustment. *Med Educ.* 2018;52(Suppl. 1):23–24.
68. Butler JM, Anderson KA, Supiano MA, Weir CR. “It feels like a lot of extra work”: Resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med.* 2017;92(7):984–990.
69. Schumacher DJ, Leslie LK, Van KD, Freed GL. Pediatric residents’ reports of quality improvement training and experiences: Time for an improvement cycle? *Acad Pediatr.* 2019;19(4):399–403.
70. Swing SR, Schneider S, Bizovi K, et al. Using patient care quality measures to assess educational outcomes. *Acad Emerg Med.* 2007;14(5):463–473.
71. Haan CK, Edwards FH, Poole B, Godley M, Genuardi FJ, Zenni EA. A model to begin to use clinical outcomes in medical education. *Acad Med.* 2008;83(6):574–580.
72. 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.
73. Yin HS, Jay M, Maness L, Zabar S, Kalet A. Health literacy: An educationally sensitive patient outcome. *J Gen Intern Med.* 2015;30(9):1363–1368.

74. Cook DA, West CP. Reconsidering the focus on "outcomes research" in medical education: A cautionary note. *Acad Med.* 2013;88:162–167.
75. Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. *Med Teach.* 2013;35:569–574.
76. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: From Flexner to competencies. *Acad Med.* 2002;77(5):361–367.
77. Holmboe E. Realizing the promise of competency-based medical education. *Acad Med.* 2015;90(4):411–413.
78. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: Effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 2006;19(2):CD000259.
79. Brehaut JC, Colquhoun HL, Eva KW, et al. Practice feedback interventions: 15 suggestions for optimizing effectiveness. *Ann Intern Med.* 2016;164(6):435–441.
80. Berenson R. If you can't measure performance, can you improve it? *JAMA.* 2016;315(7):645–646.
81. Deming WE. *The New Economics: For Industry, Government, Education.* 2nd ed. Cambridge, MA: MIT Press; 2000.
82. Cassel CK, Jain SH. Assessing individual physician performance: Does measurement suppress motivation? *JAMA.* 2012;307(24):2595–2596.
83. Sandars J CT. Self-regulation theory: Applications to medical education: AMEE Guide No. 58. *Med Teach.* 2011;33:875–886.
84. Schoen DA. *The Reflective Practitioner: How Professionals Think in Action.* New York, NY: Basic Books, Inc; 1983.
85. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med.* 2016;374(6):504–506.
86. Nelson EC, Dixon-Woods M, Batalden PB, et al. Patient focused registries can improve health, care, and science. *BMJ.* 2016;354:i3319.
87. Norman G, Norcini J, Bordage G. Competency-based education: Milestones or millstones? *J Grad Med Educ.* 2014;6(1):1–6.
88. Norman G. Medical education: Past, present and future. *Perspect Med Educ.* 2012;1(1):6–14.
89. Weick KE. *Making Sense of the Organization, Volume 2: The Impermanent Organization.* Chichester, UK: John Wiley & Sons, Ltd; 2009.



Summary

Chapter 1

Chapter 1 describes recent calls to link educational outcomes and patient care outcomes in graduate medical education. Achieving this goal is the foundation of competency-based medical education, but current gaps in quality improvement training and resident performance assessment practices hinder efforts to achieve it. Concerning gaps in assessment, there is both unwanted variation in assessments as well as failure to align assessment with patient care. In this chapter, we propose that patient-centered performance assessment, specifically entrustable professional activities (EPAs) and quality measures that are relevant to the work that residents complete (which we term resident-sensitive quality measures, or RSQMs), is a way to overcome these barriers. By closing these gaps, we will begin to provide the ability to better link educational outcomes with patient care outcomes. However, EPAs currently lack sufficient empirical evidence despite their international popularity over the past decade, and there are also insufficient quality measures for the work that residents complete. This situation leads to the main research questions for this dissertation:

What are the characteristics of developing and implementing quality measures, which we term resident-sensitive quality measures (RSQMs), that are largely attributable to residents and important to patient care?

What is the association between entrustment decisions made about residents and their performance on RSQMs?

The studies in this dissertation took place in the U.S. pediatric residency context. Specifically, we focused in the pediatric emergency department for our initial efforts in this novel research direction.

Chapter 2

Chapter 2 extends the call to focus on new approaches of performance assessment to better link educational and patient care outcomes, arguing that the future of high-quality care depends on better physician performance assessment. This chapter proposes a three-pronged research agenda in this area. First, EPAs should be implemented and studied broadly. Second, the association between entrustment decisions made about residents and quality measures for the care they provide should be investigated. Finally, with only 2% of global health expenditure dedicated to health education for all professions combined, these efforts should receive funding and resource prioritization from public and private sectors.

Chapter 3

The research program presented in this dissertation focuses on RSQM performance and entrustment decisions for individuals. However, given the recent focus on health care provided by teams, it is important to justify the focus on individuals. Chapter 3 argues the importance of focusing on individuals (by considering *attribution* of care) as well as teams, programs, and systems (by considering *contribution* to care). Contribution focuses on the various degrees different entities contribute to an outcome, and it plays an important role in analyzing program- and system-level outcomes that inform program evaluation and program-level improvements for the future. Equally important in health care are the role of the individual and aspects of care that can be attributed to an individual team member. Providing individuals with data related to their performance on the team has value in driving their personal improvement. Furthermore, we graduate, certify, and credential individual physicians, not teams, for medical practice. Additionally, incentive-based payment models focus on the quality of care provided by individuals, making it important to be able to determine outcomes of performance attributed to individuals. In this chapter, we explore how attribution and contribution analyses can be used in a complimentary fashion to discern what outcomes can and should be attributed to individuals as well as to teams and programs.

Chapter 4

Chapter 4 describes efforts to begin closing the training and assessment gap with respect to quality measures available for use with residents through the development of RSQMs for use in the pediatric emergency department setting. In this study, we used a nominal group technique (NGT) and Delphi process with faculty and fellow supervisors at Cincinnati Children's Hospital Medical Center to develop (through the NGT) and prioritize (through the Delphi) possible RSQMs for acute asthma exacerbation, bronchiolitis, and closed head injury encounters. Groups were asked to generate (in the NGT) and rate (in the Delphi) possible RSQMs based on two criteria: 1) importance of the measure to the illness of interest and 2) likelihood that a resident, and not another member of the team or the team collectively, completes the measure. The NGT produced 67 measures for asthma, 46 for bronchiolitis, and 48 for closed head injury. These were used to populate the first round of the Delphi process. After two rounds, 18 measures for asthma, 21 for bronchiolitis, and 21 for closed head injury met automatic inclusion criteria. In round three, participants categorized the potential final measures by their top 10 and next 5. Prioritized measures often fell into one of three categories: 1) appropriate medication use, 2) documentation, and 3) information provided at patient discharge. In summary, this study provides a template for identifying and developing RSQMs that may help to promote high-quality care delivery during and following training.

Chapter 5

Chapter 5 extends the work of chapter 4 in two ways: 1) replicating the process of developing RSQMs for asthma, bronchiolitis, and closed head injury in the pediatric emergency department setting with pediatric residents, and 2) engaging resident and supervisor stakeholders to develop and inform next steps in creating RSQMs. Like faculty and fellow supervisors, residents in the NGT and Delphi groups placed considerable focus across the three illnesses of interest on measures that pertained to 1) appropriate medication use, 2) documentation, and 3) information provided at patient discharge. Subsequent focus groups with residents as well as faculty and fellow supervisors highlighted hospital medicine and general pediatrics as priority areas for developing future RSQMs but also noted contextual variables that influence the application of similar measures in different settings (e.g., team structure of general pediatrics clinic may limit the utility of some measures). Residents and supervisors had both similar and unique insights into developing RSQMs. For example, the resident NGT produced more measures than the supervisor NGT, and both focus groups felt this may be because residents work on the front line and know more details of care provided there. However, supervisors also noted that residents may not be able to prioritize well or do not know resources available to help complete their work, and residents noted that supervisors may not know details that go into actually providing frontline care. In summary, this chapter builds on the work of chapter 4 to continue to pave the path forward in developing future RSQMs by exploring specific settings, measures, and stakeholders to consider when undertaking this work.

Chapter 6

Chapter 6 presents the range of individual and collective performance for the RSQMs introduced in chapters 4 and 5 when these RSQMs are used to assess pediatric residents working in the pediatric emergency department. During the 2017–2018 academic year, pediatric residents in the Cincinnati Children’s Hospital Medical Center pediatric emergency department were assessed using RSQMs for patient encounters focused on acute asthma exacerbation (21 RSQMs), bronchiolitis (23), and closed head injury (19). Of included encounters, 83 residents cared for 110 patients with asthma, 112 with bronchiolitis, and 77 with closed head injury. Residents had the opportunity to meet most RSQMs, but exceptions existed (e.g., patients with asthma exacerbation presenting to the medical resuscitation area had initial orders placed by the supervising physician, per protocol, precluding residents from meeting some RSQMs that pertained to initial orders). In encounters presenting the opportunity, there was a wide range of frequency for how often residents performed RSQMs. One closed head injury measure was met in all encounters. Across illnesses, another 9 RSQMs were met in almost all encounters, but most RSQMs were not met, with a frequency ranging from several times to the majority of encounters. To determine the proportion of individual RSQMs correctly performed out of total possible RSQMs that could have been performed, we generated a composite score for each encounter. RSQM composite scores demonstrated notable range and variation: asthma mean,

0.81 (SD: 0.11); bronchiolitis mean, 0.62 (SD: 0.12); closed head injury mean, 0.63 (SD: 0.10). In summary, chapter 6 demonstrates that individual and composite RSQMs capture and discern resident performance across patient encounters, opening the door for their use in other contexts and by other specialties.

Chapter 7

Chapter 7 explores the association between RSQM composite scores for individual pediatric emergency department encounters, supervisor entrustment decisions (as measured by supervision level assigned) for those encounters, and patient acuity and complexity for those encounters. The 83 residents and 299 patient encounters for asthma, bronchiolitis, and closed head injury presented in chapter 6 were used for the study presented in this chapter. To measure the association of RSQM composite scores with the other variables of interest, we used mixed models to account for observations nested within residents. Entrustment decisions were positively associated with asthma RSQM composite scores (beta-coefficient 0.03; $P = .0004$), but there was no significant association between RSQM composite scores and entrustment decisions for bronchiolitis or closed head injury. While statistically significant, the positive linear relationship for asthma is likely not educationally or clinically significant because RSQM composite scores only increased 0.03 (out of 1) for every 1 (out of 7) point increase in entrustment score. A positive relationship would be expected between entrustment decisions and RSQM composite scores. Given that RSQMs are a more objective measure of performance, the findings of this study should prompt consideration of whether RSQM performance data should help inform entrustment decisions. RSQM composite scores were significantly higher when acuity was also higher and significantly lower when acuity was also lower for both asthma ($P = .0009$) and bronchiolitis ($P = .01$). However, RSQM composite scores were almost identical for the different levels of acuity for closed head injury ($P = .94$). There was no association between RSQM composite scores and patient complexity.

Chapter 8

Having established initial implementation evidence for using RSQMs in the pediatric emergency department, chapter 8 sought to consider how RSQM data might be used by those who make assessment decisions outside the pediatric emergency department. The study presented in this chapter achieved this goal by exploring how individual clinical competency committee (CCC) members interpret, use, and prioritize RSQM data inserted into their usual review processes. In this constructivist grounded theory study, we purposively and theoretically sampled 19 pediatric residency CCC members from Cincinnati Children's Hospital Medical Center. Participants were provided a resident assessment portfolio consisting of performance ratings and narrative comments for five rotations as well as RSQM data for one of these rotations. They were asked to review with the purpose of making a decision about the residents' ability to care for patients presenting with common, acute problems (a general pediatric entrustable

professional activity). Data collection consisted of two phases: 1) observation and think aloud in which participants reviewed the resident portfolio, and 2) semistructured interviews to probe participant reviews. Analysis moved from close reading to coding and theme development followed by the development of a model illustrating how themes interact with one another. We identified five dimensions for how participants view and use RSQMs: 1) ability to orient to RSQMs: confusing to self-explanatory; 2) propensity to use RSQMs: reluctant to enthusiastic; 3) RSQM interpretation: requires contextualization to self-evident; 4) RSQMs for assessment decisions: not sticky to sticky; and 5) expectations for residents: potentially unfair to fair use of RSQMs. These dimensions interact with one another to generate three profiles regarding how participants used RSQM data: eager incorporation, willing incorporation, and disinclined incorporation. The former two profiles were most common, with most participants using RSQMs to some extent, supporting the inclusion of RSQMs as resident assessment data for CCC review.

Chapter 9

As RSQMs continue to be developed, it will be important to optimize ways to extract them from the electronic health record (EHR) and to automatically identify which residents they should be attributed to in settings where multiple residents provide care to a given patient, such as the inpatient environment. Chapter 9 sought a method for attributing care of individual patients to individual interns based on “footprints” in the EHR. Primary interns caring for patients on an internal medicine inpatient service were recorded daily by five attending physicians of record at University of Cincinnati Medical Center in August 2017 and January 2018. These records were considered gold standard identification of primary interns. The following EHR variables were explored to determine representation of primary intern involvement in care: postgraduate year, progress note author, discharge summary author, physician order placement, and logging clicks in the patient record. These variables were turned into categorical attributes (e.g., progress note author: yes/no), and informative attributes were selected and modeled using a decision tree algorithm. A total of 1,511 access records were generated; 116 were marked as having a primary intern assigned. All variables except discharge summary author displayed at least some level of importance in the models. The best model achieved 78.95% sensitivity, 97.61% specificity, and an area under the receiver-operator curve of approximately 91%. Thus, this study successfully predicted primary interns caring for patients on inpatient teams using EHR data with excellent model performance. These results provide a foundation for attributing patients to primary interns for the purposes of determining profiles of patient diagnoses by trainee, supporting continuous quality improvement efforts in graduate medical education, and ultimately attributing RSQMs to residents.

Chapter 10

Chapter 10 details calls for ensuring high value health care and posits that medical education can and should drive progress toward this goal. We argue that getting medical education to

the point where it can do this work will require a clear, sustained focus on the patient, which is not traditionally a notable focus of medical education research or practice. Patient-focused assessment efforts, such as RSQMs, can help achieve this goal. The summary of our research is that RSQMs developed through consensus group methodology largely do reflect the performance of residents and that performance varies widely across residents and encounters. Future development of RSQMs will need to consider engaging other stakeholders and expanding to other settings and specialties. Our research suggests future settings for pediatrics, including hospital medicine and primary care outpatient pediatrics, but these efforts should also be undertaken in other medical specialties as well as surgical and hospital-based specialties. Completing this work should include engaging with stakeholders we have not yet involved in our work to date, including nurses and patients/families.

Our findings for the association between RSQMs and entrustment decisions made about residents varied. RSQM composite scores were significantly higher for residents entrusted with unsupervised practice or the ability to supervise others than for residents entrusted to lower supervision levels for all conditions (asthma, closed head injury, and all conditions combined) other than bronchiolitis. This finding offers validity evidence for RSQM composite scores; however, the absolute difference was modest. Additionally, when considering each supervision level and not combining them into two categories (at least entrusted with unsupervised practice or not entrusted to that level), a significant positive linear relationship between RSQM composite scores and entrustment decisions was only seen for asthma. However, that observation for asthma is likely not educationally or clinically significant. We believe this finding may be more problematic for the consideration of entrustment as an assessment framework than for using RSQMs to inform resident assessment. Specifically, we posit that RSQMs may help to improve the veracity of entrustment decisions.

Despite the need for further understanding and validity evidence, we believe patient-focused assessment efforts, such as RSQMs and entrustment, should be embraced given the promise to ensure high value care embodied in their attention to the patient. However, hazards to their use exist. There can be reluctance to adopt new ways of approaching assessment. Nonetheless, recent conversations in medical education about moving away from psychometrics and toward welcoming the subjective in performance assessment suggest change is possible. With calls for ensuring high value health care as well as uniting the missions of medical education and health care delivery, medical education may be at a crossroads. Efforts such as RSQMs can help the community take a clear step toward focusing on the patient in our assessment efforts.



Samenvatting

Hoofdstuk 1

Hoofdstuk 1 beschrijft de recente vraag naar een verbeterde link tussen de onderwijsresultaten uit de medische vervolgopleiding en de kwaliteit van de patiëntenzorg. Hoewel het bereiken van dit doel aan de basis ligt van competentiegericht onderwijs, belemmeren huidige hiaten in toetspraktijken en trainingen gericht op kwaliteitsverbetering pogingen om dit te bereiken. Met hiaten in toetspraktijken refereren we aan de ongewenste variatie in toetsen alsook aan de slechte afstemming van deze toetsing op de patiëntenzorg. In dit hoofdstuk stellen we dat patiëntgerichte toetsing van prestaties, en in het bijzonder *entrustable professional activities* (EPA's) en kwaliteitsmaten die betrekking hebben op het werk dat aiossen verrichten (welke we in het navolgende "aios-specifieke kwaliteitsmaten", oftewel ASKM's zullen noemen), een manier is om deze belemmeringen weg te nemen. Met het nemen van deze hordes zetten we een eerste stap richting een betere afstemming van onderwijsresultaten op de patiëntenzorg. Hoewel EPA's de afgelopen 10 jaar veel internationale populariteit hebben genoten, is het empirische bewijs daarvoor momenteel onvoldoende; daarnaast zijn er onvoldoende kwaliteitsmaten voor het werk dat aiossen verrichten. Dit bracht ons tot de volgende centrale onderzoeksvragen voor dit proefschrift:

Welke aspecten komen er kijken bij de ontwikkeling en toepassing van kwaliteitsmaten, welke we "aios-specifieke kwaliteitsmaten" (ASKM's) noemen, die grotendeels aan aiossen zijn toe te schrijven en die van belang zijn voor de patiëntenzorg?

Wat is het verband tussen *entrustment decisions*, oftewel besluiten om een kritische activiteit aan aiossen toe te vertrouwen, en hun prestaties op de ASKM's?

De studies in dit proefschrift vinden plaats in de context van een medische vervolgopleiding Kindergeneeskunde in de VS. Bij onze eerste stappen in deze baanbrekende onderzoeksrichting richtten we ons meer specifiek op de afdeling spoedeisende hulp (SEH) van dit specialisme.

Hoofdstuk 2

In Hoofdstuk 2 wordt de vraag naar nieuwe benaderingen van prestatietoetsing teneinde onderwijsresultaten en de kwaliteit van de patiëntenzorg beter op elkaar te laten aansluiten, verder aangevuld door te stellen dat de toekomst van kwalitatief hoogwaardige zorg afhangt van een verbeterde toetsing van artsprestaties. Er wordt een drieledige onderzoeksagenda op dit gebied voorgesteld. Ten eerste moeten EPA's ingevoerd en uitgebreid onderzocht worden. Ten tweede moet het verband tussen *entrustment decisions* over aiossen en kwaliteitsmaten voor de zorg die zij verlenen onderzocht worden. Ten slotte zouden de publieke en private sector deze inspanningen moeten financieren en daar voorrang aan moeten geven, aangezien slechts

2% van de totale uitgaven voor de gezondheidszorg naar gezondheidszorgonderwijs voor alle beroepen samen gaat.

Hoofdstuk 3

Het in dit proefschrift voorgestelde onderzoeksprogramma richt zich op individuele ASKM-prestaties en op besluiten om een kritische activiteit aan individuen toe te vertrouwen. Gezien de recente focus op zorgverlening door teams is het echter van belang deze toespitsing op individuen te verantwoorden. Hoofdstuk 3 licht toe waarom het belangrijk is te focussen op zowel individuen (door de *toeschrijving* van zorg aan een individu te overwegen) als teams, programma's en systemen (door de *bijdrage* aan zorg te overwegen). 'Bijdrage' richt zich op de diverse maten waarin verschillende eenheden aan een resultaat bijdragen. Daarbij speelt het een belangrijke rol bij het analyseren van resultaten op programma- en systeemniveau die als input dienen voor programma-evaluatie en -verbeteringen voor de toekomst. In de zorg is de rol van een individu en de aspecten van zorg die aan een individueel teamlid kunnen worden toegeschreven echter net zo belangrijk. Het is waardevol om individuen informatie te verschaffen over hun functioneren in het team, omdat dit hen kan aanzetten tot persoonlijke verbetering. Verder is het zo dat we individuele artsen, niet teams, laten afstuderen, hen diplomeren en hen kwalificaties toekennen voor de artsenpraktijk. Bovendien zijn prestatiebekostigingsmodellen ook gericht op de kwaliteit van de door individuen verleende zorg, waardoor het belangrijk is om de resultaten van individuele prestaties te kunnen meten. In dit hoofdstuk onderzoeken we hoe analyses van 'toeschrijving' en 'bijdrage' aanvullend kunnen worden toegepast om te achterhalen welke resultaten aan individuen en welke aan teams en programma's kunnen en moeten worden toegeschreven.

Hoofdstuk 4

Hoofdstuk 4 beschrijft pogingen om de inhaalslag met betrekking tot kwaliteitsmaten voor gebruik onder aiossen te maken, door ASKM's te ontwikkelen voor gebruik op de afdeling kinder-SEH. In deze studie gebruikten we een *nominal group technique* (NGT) en Delphi-methode met stafleden en collega-supervisors van het Cincinnati Children's Hospital Medical Center (CCHMC) om mogelijke ASKM's te ontwikkelen (d.m.v. de NGT) en te prioriteren (d.m.v. de Delphi) voor contacten met patiënten met acute verergering van astma, bronchiolitis en licht schedeltrauma. We vroegen de groepen om mogelijke ASKM's te bedenken (in de NGT) en deze te scoren (in de Delphi) op basis van de volgende twee criteria: 1) het belang van de maat voor de betreffende ziekte; en 2) de kans dat een aios, en geen ander teamlid of het team gezamenlijk, de handeling waarop de maat van toepassing is, verricht. De NGT leverde 67 maten op voor astma, 46 voor bronchiolitis en 48 voor licht schedeltrauma. Deze vormden de input voor de eerste ronde van de Delphi-methode. Na twee rondes voldeden 18 maten voor astma, 21 voor bronchiolitis en 21 voor licht schedeltrauma aan de criteria voor automatische inclusie. Uit deze potentiële eindmaten stelden de deelnemers in ronde drie hun top 10 en de

volgende 5 op. De maten die geprioriteerd werden, vielen vaak in een van de volgende drie categorieën: 1) passend medicatiegebruik; 2) documentatie; en 3) informatieverstrekking bij ontslag van de patiënt. Samenvattend biedt deze studie een richtsnoer voor het identificeren en ontwikkelen van ASKM's die kwalitatief hoogwaardige zorgverlening tijdens en na de opleiding kunnen helpen bevorderen.

Hoofdstuk 5

Hoofdstuk 5 vult het werk van hoofdstuk 4 op twee manieren aan: 1) het herhaalt het proces voor het ontwikkelen van ASKM's voor astma, bronchiolitis en licht schedeltrauma op de kinder-SEH met aiossen Kindergeneeskunde; en 2) het betreft belanghebbenden als aiossen en supervisors ten behoeve van de ontwikkeling van ASKM's en vervolgstappen daarin. Net als de stafleden en collega-supervisors richtten aiossen in de NGT en Delphi-rondes zich bij de drie betreffende ziektes voornamelijk op maten die betrekking hadden op: 1) passend medicatiegebruik; 2) documentatie; en 3) informatieverstrekking bij ontslag van de patiënt. De daaropvolgende focusgroepen met zowel aiossen als stafleden en collega-supervisors merkten ziekenhuisgeneeskunde en Algemene Kindergeneeskunde aan als prioritaire gebieden voor het ontwikkelen van toekomstige ASKM's, maar wezen ook op omgevingsvariabelen die de toepassing van vergelijkbare maten in andere settings beïnvloeden (de teamsamenstelling van de afdeling Algemene Kindergeneeskunde kan bijvoorbeeld de bruikbaarheid van enkele maten beperken). Aiossen en supervisors hadden zowel vergelijkbare als unieke inzichten in het ontwikkelen van ASKM's. Zo leverde de NGT met aiossen meer maten op dan de NGT met supervisors en beide focusgroepen meenden dit te kunnen toeschrijven aan het feit dat aiossen volop in de praktijk bezig waren en daardoor meer details kenden van de zorg die daar verleend werd. Supervisors merkten echter op dat aiossen mogelijk niet voldoende in staat zijn de juiste prioriteiten te stellen of niet weten welke middelen tot hun beschikking staan om hun werk goed te kunnen verrichten. Aiossen daarentegen, gaven aan dat supervisors mogelijk zelfs geen weet hebben van de details die bij het verlenen van zorg aan de frontlinie komen kijken. Samenvattend bouwt dit hoofdstuk voort op hoofdstuk 4 om de weg vooruit te blijven effenen voor de ontwikkeling van toekomstige ASKM's, door specifieke settings, maten en belanghebbenden waarmee rekening gehouden moet worden bij het ondernemen van dit werk te onderzoeken.

Hoofdstuk 6

Hoofdstuk 6 presenteert de verschillende individuele en collectieve resultaten voor de in hoofdstuk 4 en 5 besproken ASKM's, nadat deze waren gebruikt om aiossen Kindergeneeskunde werkzaam op de kinder-SEH te beoordelen. Tijdens het academisch jaar 2017-2018 werden aiossen Kindergeneeskunde op de afdeling kinder-SEH van het CCHMC beoordeeld aan de hand van ASKM's voor acute verergering van astma (21 ASKM's), bronchiolitis (23) en licht schedeltrauma (19). Van de geïncludeerde patiëntcontacten zorgden 83 aiossen voor 110 patiënten met astma,

112 met bronchiolitis en 77 met licht schedeltrauma. Op enkele uitzonderingen na (bijvoorbeeld wanneer patiënten met verergering van astma zich meldden bij de medische reanimatie-afdeling, waardoor het zorgproces, overeenkomstig het protocol, in gang gezet werd door de toeziend arts, konden de aiossen niet voldoen aan enkele ASKM's die op deze handelingen betrekking hadden), waren aiossen in de gelegenheid om de meeste ASKM-relevante activiteiten te verrichten. De frequentie waarmee ASKM-gerelateerde handelingen zich voordeden was hoog en dit bood aiossen ruim de kans om aan de kwaliteitsmaten te voldoen. Aan één kwaliteitsmaat voor licht schedeltrauma werd in alle patiëntcontacten voldaan. Over de drie ziekten werd in bijna alle patiëntcontacten nog eens aan 9 ASKM's voldaan, hoewel in een aantal tot het merendeel van de patiëntcontacten niet werd voldaan aan de meeste ASKM's. Om vast te stellen hoeveel individuele ASKM's van het totale aantal ASKM's dat had kunnen worden uitgevoerd correct werden uitgevoerd, werd er voor elk patiëntcontact een samengestelde score uitgerekend. De samengestelde scores voor de ASKM's liepen opvallend uiteen en vertoonden opmerkelijk veel variatie: gemiddelde voor astma: 0,84 (SD: 0,11); gemiddelde bronchiolitis: 0,62 (SD: 0,12); licht schedeltrauma: 0,63 (SD: 0,10). Samenvattend laat hoofdstuk 6 zien dat individuele en samengestelde ASKM's de prestaties van aiossen over patiëntcontacten heen kunnen vangen en onderscheiden, wat de deur opent voor het gebruik ervan in andere contexten en door andere specialismen.

Hoofdstuk 7

Hoofdstuk 7 onderzoekt het verband tussen samengestelde ASKM-scores voor individuele patiëntcontacten op de afdeling kinder-SEH, entrustment decisions die supervisors nemen voor die patiëntcontacten (zoals blijkt uit de mate van toegewezen supervisie) en de ernst* en complexiteit van deze patiëntcontacten. Voor de in dit hoofdstuk gerapporteerde studie gebruikten we de 83 aiossen en 299 contacten met astma-, bronchiolitis- en licht-schedeltraumapatiënten die in hoofdstuk 6 werden gepresenteerd. Om het verband tussen de samengestelde ASKM-scores en de andere relevante variabelen te meten, werden mixed models toegepast om de binnen aiossen geneste observaties te verklaren. Er bestond een positieve correlatie tussen entrustment decisions en de samengestelde ASKM-scores voor astma (bèta-coëfficiënt van 0,03; $p = 0,0004$), maar er werd geen significante correlatie gevonden tussen samengestelde ASKM-scores en entrustment decisions voor bronchiolitis en licht schedeltrauma. Hoewel statistisch significant, heeft dit positieve lineaire verband voor astma waarschijnlijk geen onderwijskundige noch klinische relevantie, omdat de samengestelde ASKM-scores bij elke toename van 1 punt (op 7) in entrustment-score enkel met 0,03 (op 1) toenamen. Er wordt verondersteld dat er een positief verband bestaat tussen entrustment decisions en de samengestelde ASKM-scores. Aangezien ASKM's een meer objectieve maatstaf voor prestaties vormen, zouden we ons naar aanleiding van de bevindingen van deze studie moeten afvragen of entrustment decisions (al dan niet deels) gebaseerd moeten worden op uit ASKM's verkregen prestatiegegevens. De samengestelde ASKM-scores waren significant hoger naarmate de ernst toenam en significant

lager naarmate deze afnam voor zowel astma ($p=0,0009$) als bronchiolitis ($p=0,01$). De samengestelde ASKM-scores voor licht schedeltrauma bleven echter vrijwel identiek bij wisselende maten van ernst ($p=0,94$). Er werd geen verband gevonden tussen samengestelde ASKM-scores en patiëntcomplexiteit.

Hoofdstuk 8

Toen het eerste bewijs van invoering van ASKM's op de afdeling kinder-SEH er eenmaal lag, trachtte hoofdstuk 8 na te gaan hoe ASKM-gegevens zouden kunnen worden gebruikt door mensen van buiten de afdeling kinder-SEH verantwoordelijk voor het nemen van beoordelingsbesluiten. Om dit doel te bereiken onderzocht de studie in dit hoofdstuk hoe individuele leden van de Klinische Competentiecommissie (KCC) de ASKM-data interpreterten, gebruikten en prioriteren wanneer deze in hun gebruikelijke beoordelingsprocedure worden opgenomen. In deze constructivistische-gefundeerde-theoriestudie namen we een doelgerichte en theoretische steekproef van 19 KCC-leden van de vervolgopleiding Kindergeneeskunde van het CCHMC. Participanten ontvingen een beoordelingsportfolio van een aios voorzien van beoordelingsscores en narratief commentaar over vijf klinische stages, alsook ASKM-gegevens over een van deze stages. Vervolgens werd hen gevraagd om deze te beoordelen en daarbij tot een besluit te komen over de mate waarin zij de aios in staat achtten om de zorg voor patiënten met veelvoorkomende acute problemen op zich te nemen (een entrustable professional activity voor Algemene Kindergeneeskunde). De dataverzameling besloeg twee fasen: 1) het observeren van en hardop denken door participanten tijdens het beoordelen van het portfolio van de aios; en 2) semigestructureerde interviews om participanten verder te bevragen over hun beoordelingen. De analyse ging van zorgvuldig lezen over op het coderen en ontwikkelen van thema's, wat uitmondde in een model dat in kaart bracht hoe de verschillende thema's zich onderling tot elkaar verhielden. We onderscheidden vijf dimensies voor de manier waarop participanten naar ASKM's keken en deze gebruikten: 1) Het zich kunnen oriënteren op de ASKM's: variërend van 'verwarrend' tot 'dat wijst zich vanzelf'; 2) De geneigdheid om de ASKM's daadwerkelijk te gebruiken: variërend van 'met terughoudendheid' tot 'met enthousiasme'; 3) De interpretatie van ASKM's: variërend van 'behoeft plaatsing in de juiste context' tot 'is duidelijk'; 4) ASKM's voor beoordelingsbesluiten: variërend van 'is niet lastig' tot 'is lastig'; en 5) Verwachtingen voor aiossen: variërend van 'het gebruik van ASKM's is mogelijk oneerlijk' tot 'is eerlijk'. Het samenspel van deze dimensies leidde tot drie profielen die weergaven hoe participanten gebruik maakten van de ASKM-gegevens, te weten: enthousiaste toepassing, gewillige toepassing en terughoudende toepassing. De eerste twee profielen kwamen het vaakst voor: de meeste participanten maakten in zekere mate gebruik van de ASKM's, wat spreekt vóór de opname van ASKM's ten behoeve van de beoordeling van aiossen door de KCC.

Hoofdstuk 9

Naarmate ASKM's verder worden ontwikkeld, wordt het belangrijk dat we een optimale manier vinden om deze uit het elektronisch patiëntendossier (EPD) te halen, alsook om automatisch vast te stellen aan welke aiossen zij moeten worden toegeschreven in settings waar meerdere aiossen gemoeid zijn met de zorg voor een bepaalde patiënt, zoals de kliniek. Hoofdstuk 9 ging op zoek naar een methode waarmee de zorg voor individuele patiënten met behulp van "voetafdrukken" in het EPD aan individuele aiossen kon worden toegeschreven. In augustus 2017 en januari 2018 registreerden vijf dienstdoende specialisten van het University of Cincinnati Medical Center dagelijks de activiteiten van aiossen die zorgdroegen voor patiënten op een ziekenhuisafdeling Interne Geneeskunde. Deze gegevens beschouwden we als ultiem ijkpunt voor het identificeren van aiossen. Om te kunnen vaststellen in welke mate de aios bij de zorg betrokken was, keken we naar de volgende variabelen: vervolgopleidingsjaar, voortgangsnotitie van auteur, ontslagsamenvatting van auteur, door arts aangevraagde onderzoeken en activiteitenlogs uit het patiëntendossier. We zetten deze variabelen om in categorale kenmerken (bijv. voortgangsnotitie auteur: ja/nee), selecteerden informatieve kenmerken en rangschikten deze met behulp van een beslisboomalgoritme. Er werden in totaal 1.511 activiteiten geregistreerd, waarvan er 116 werden toegeschreven aan aiossen. Uit de resultaten bleek dat alle variabelen, behalve de 'ontslagsamenvatting auteur', ten minste enige mate van belangrijkheid vertoonden in de modellen. Het beste model bereikte 78,95% sensitiviteit, 97,61% specificiteit en een gebied onder de ROC-curve van circa 91%. Kortom, deze studie voorspelde met succes, door gebruik te maken van gegevens uit het EPD, de mate waarin aiossen binnen een ziekenhuisteam voor patiënten zorgdroegen. Daarbij kwam een uitstekend model naar voren. Hiermee wordt de basis gelegd voor het toeschrijven van patiënten aan aiossen zodat voor elke aios in kaart kan worden gebracht met welke patiëntdiagnoses deze effectief te maken heeft gehad. Deze kennis kan worden ingezet ten bate van voortdurende kwaliteitsverbeteringsdoeleinden in de medische vervolgopleiding en uiteindelijk helpen om ASKM's aan aiossen toe te schrijven.

Hoofdstuk 10

Hoofdstuk 10 beschrijft opnieuw de herhaalde vraag naar een betere waarborging van kwalitatief hoogwaardige zorg en stelt dat het medisch onderwijs een belangrijke bijdrage kan en moet leveren aan het bereiken van dit doel. Wij zijn van mening dat we ons hiervoor ten volle en voortdurend moeten richten op de patiënt, iets dat voor onderzoek van medisch onderwijs of de medische-onderwijspraktijk van oudsher nooit een grote focus is geweest. Patiëntgerichte toetsmethoden zoals ASKM's kunnen helpen om dit doel te bereiken. De kernboodschap van ons onderzoek is dat ASKM's die met behulp van consensusgroepsmethoden zijn ontwikkeld grotendeels de handelingen van aiossen weerspiegelen en dat de verrichte handelingen sterk van aios tot aios en van patiëntcontact tot patiëntcontact verschillen. Bij de toekomstige ontwikkeling van ASKM's moet worden overwogen om andere belanghebbenden in het proces te betrekken en dit uit te breiden naar andere settings en specialismen. Ons onderzoek

reikt al enkele mogelijke toekomstige settings op het gebied van Kindergeneeskunde aan, waaronder ziekenhuisgeneeskunde en de poli Algemene Kindergeneeskunde, maar het is van belang dat er ook onderzoek wordt verricht binnen andere medische specialismen, alsmede chirurgische en ziekenhuisspecialismen. Om dit werk te voltooien zouden belanghebbenden die we tot nog toe niet in ons werk hebben opgenomen, zoals verplegers, patiënten en hun families, betrokken moeten worden. Onze bevindingen met betrekking tot het verband tussen de ASKM's en entrustment decisions over aiossen liepen uiteen. Voor alle aandoeningen (astma, licht schedeltrauma en alle andere aandoeningen samen), behalve bronchiolitis, gold dat de samengestelde ASKM-scores significant hoger waren wanneer aiossen hun verantwoordelijkheden zonder supervisie mochten uitvoeren of wanneer zij belast waren met de supervisie van anderen dan wanneer zij onder een grotere mate van supervisie stonden. Dit levert validiteitsbewijs voor de samengestelde ASKM-scores, hoewel het absolute verschil gering was. Voorts zagen we dat, wanneer de supervisieniveau's afzonderlijk werden beschouwd in plaats van deze in twee categorieën op te delen (1: mag op zijn minst taken zonder supervisie uitvoeren, en 2: heeft ongeacht het niveau nog supervisie nodig), er alleen voor astma een significant positief lineair verband bestond tussen de samengestelde ASKM-scores en entrustment decisions. Waarschijnlijk heeft deze observatie voor astma echter geen onderwijskundige noch klinische relevantie. Naar onze mening vormt deze bevinding eerder een te overwegen bedreiging voor het gebruik van entrustment als beoordelingskader dan voor het gebruik van ASKM's ten behoeve van de beoordeling van aiossen. Anders gezegd, we stellen dat ASKM's kunnen helpen de nauwkeurigheid van entrustment decisions te verbeteren. Ondanks de behoefte aan verder inzicht en validiteitsbewijs, vinden wij dat patiëntgerichte toetsmethoden zoals ASKM's en entrustment moeten worden omarmd aangezien zij, door hun aandacht voor de patiënt, kwalitatief hoogwaardige zorg beloven te borgen. Het gebruik van deze toetsmethoden is echter niet zonder risico. De overgang op nieuwe benaderingen van toetsing kan met enige weerstand gepaard gaan. Dit gezegd te hebben, beloven recente discussies waarin ervoor gepleit wordt de psychometrie achter te laten en in plaats daarvan het subjectieve in prestatiebeoordelingen te omarmen, dat verandering wél mogelijk is. Dankzij de herhaalde vraag naar een betere waarborging van kwalitatief hoogwaardige zorg alsmede een betere afstemming tussen de missies van het medisch onderwijs en de zorg, staat het medisch onderwijs mogelijk aan de vooravond van een keerpunt. Vernieuwingen zoals ASKM's kunnen de gemeenschap helpen om een duidelijke stap te zetten richting het einddoel om de patiënt in onze beoordelingsmethoden centraal te stellen.



Valorization

Relevance

The relevance of the development and implementation of resident-sensitive quality measures (RSQMs) is argued in the introduction (chapter 1) and discussion (chapter 10) of this thesis. In brief, the foundation of competency-based medical education is ensuring that the outcomes of training prepare graduates to meet the needs of populations of patients.¹ For more than 40 years, a medical education research agenda that focuses on the relationship between training and patient outcomes has been advocated.^{2,3} During this time, little progress has been made in implementing such an agenda. However, research articles and perspectives pieces from the past few years are hastening work in this area.³⁻¹⁰ The studies presented in this thesis make important contributions to this body of work, advancing our understanding of how we may link learner outcomes and patient outcomes through developing a patient-focused approach to assessment—RSQMs. RSQMs not only provide feedback to residents about the quality of the care they provide, addressing a substantial gap in graduate medical education training,¹¹⁻¹⁴ but can also serve to assess their performance at the individual level as well as at the program level (when aggregating data for several residents). To this end, RSQMs are objective measures of performance that can help balance some of the unwanted variability in performance assessment that currently exists.¹⁵⁻²⁰ In particular, the findings presented in chapter 7 raise the question of whether RSQM data can, and should, inform assessment decisions that are made using the current prevailing approach to assessment, entrustable professional activities (EPAs). Other than RSQMs, EPAs are the only other patient-focused means of assessment described in the literature to date.

Medical education having a primary focus on the patient is not optional, as noted in chapter 10. Rather, it is the foundation of relevant medical education. This truth is moving toward center stage as health care systems internationally continue to strive for higher quality care for patients. However, graduate medical education has not traditionally placed focus on the patient but rather on the abilities of learners. Moving forward, education and training must be considered in the context of, and aligned with, health care delivery systems if future care is to meet societal needs.^{21,22} With this in mind, advances such as RSQMs will make important and necessary contributions to the future of graduate medical education and health care more broadly.

Innovation

While recent perspective papers have noted the importance of using quality measure data to determine educational outcomes, little research has been done in this area.^{5,6} The work of individuals who have pursued investigations in this area has tended to focus on dyads (e.g., resident and supervisor) and systems of care.^{23,24} However, RSQMs focus on the individual providing care. While care is ultimately provided by teams and systems, we graduate, certify,

and credential individuals. Furthermore, individualized feedback is important to driving personal improvement. Therefore, a focus on the individual is not only innovative but also critically important.

Target Groups

RSQMs will likely be viewed as most useful by residency program leaders who are required to assess resident performance. In the United States, these program leaders are also required by the Accreditation Council for Graduate Medical Education to provide practice feedback to residents for their use in personal quality improvement efforts. As chapter 8 illustrates, RSQMs may also have notable resonance with clinical competency committee members as a type of assessment data to inform the summative assessment decisions they make about residents.

If the traditional silos of medical education and health care delivery indeed align their foci, as chapters 1 and 10 argue is paramount, RSQMs will also find applicability among health systems leaders; quality officers in care delivery systems; quality improvement scientists; and accrediting, certifying, and credentialing bodies.

While we engaged residents in the development of RSQMs, we do not yet know what their experience is with receiving feedback using RSQMs. Hopefully, residents will find RSQMs to be beneficial to their development and improvement efforts, but their reactions to RSQMs remain a key area for future research.

Finally, patients are increasingly attentive to outcomes of care and the quality of care they receive from providers and institutions. This evolution opens the door for patients and families to take interest in RSQMs in the future.

Schedule and Implementation

We have successfully studied RSQMs in a local context (i.e., single institution) and have automated the reporting of a subset of the asthma RSQMs to residents in this local setting as well. An important next step is determining how residents view and interpret RSQM performance feedback when it is provided to them.

We also need to determine whether the RSQMs developed at Cincinnati Children's Hospital Medical Center can be applied in other institutions. To explore the generalizability of RSQMs, we are currently engaging in a multisite, multicountry study funded by the National Board of Medical Examiners' Stemmler Fund.

As noted in chapter 5, hospital medicine and general pediatrics are the most important settings to focus on for the continued development of RSQMs. We have begun developing RSQMs for

both settings. With an eye toward generalizability beyond a single institution, the development of these measures is engaging residents and faculty from across the United States.

In addition to developing RSQMs for other settings within pediatrics, we are currently developing measures for the internal medicine general medicine inpatient wards. This work will allow us to explore differences between the types and nature of RSQMs appropriate for internal medicine and for pediatrics. Looking at a second specialty will also position us to continue expanding the development of RSQMs in other specialties.

Finally, the American Board of Pediatrics (ABP) is considering a substantially increased use of EPAs to determine residents' ability to sit for the initial certification examinations the ABP offers in general pediatrics and all pediatric subspecialties. Given the findings of chapter 7, which suggest RSQM data may be beneficial to making entrustment decisions, once further validity evidence is available, RSQMs could be useful in the ABP's expanded use of EPAs.

Activities and Products

As noted in the previous sections, several activities are currently expanding the products of the work detailed in this thesis. The work in this thesis has also been disseminated in grand rounds presentations at Cincinnati Children's Hospital Medical Center and other institutions. Additionally, this work was the focus of a webinar hosted by the Association for Medical Education in Europe (AMEE) in the fall of 2019 and was shared in a plenary presentation at the Association of American Medical Colleges' Integrating Quality Conference in 2019. Furthermore, research presentations of studies contained in this thesis have been presented at the Royal College of Physicians and Surgeons of Canada's International Conference on Residency Education in 2019, the Second World Summit on Competency-Based Medical Education at AMEE 2016, at the Pediatric Academic Societies Meeting in 2019, and at the Rogano Conference in 2016 and 2019. Presentation at several national and international conferences is also planned for 2020.

The early dissemination of RSQM work has garnered attention that has led to some of the invited presentations noted above. This work has also been discussed a few times in the past year on the KeyLIME (Key Literature in Medical Education) podcast that has a large international listenership.

References

1. 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.
2. Magraw RM, Fox DM, Weston JL. Health professions education and public policy: A research agenda. *J Med Educ*. 1978;53(7):539–546.
3. Weinstein DF. Optimizing GME by measuring its outcomes. *N Engl J Med*. 2017;377(21):2007–2009.
4. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. *Acad Med*. 2018;93(7):975–978.
5. Chahine S, Kulasegaram K, Wright S, et al. A call to investigate the relationship between education and health outcomes using big data. *Acad Med*. 2018;93(6):829–832.
6. Arora VM. Harnessing the power of big data to improve graduate medical education: Big idea or bust? *Acad Med*. 2018;93(6):833–834.
7. Gillespie C, Zabar S, Altshuler L, et al. The Research on Medical Education Outcomes (ROME) Registry: Addressing ethical and practical challenges of using “bigger,” longitudinal educational data. *Acad Med*. 2016;91(5):690–695.
8. Triola MM, Hawkins RE, Skochelak SE. The time is now: Using graduates’ practice data to drive medical education reform. *Acad Med*. 2018;93(6):826–828.
9. Caverzagie KJ, Lane SW, Sharma N, et al. Proposed performance-based metrics for the future funding of graduate medical education: Starting the conversation. *Acad Med*. 2018;93(7):1002–1013.
10. Smirnova A, Sebok-Syer SS, Chahine S, et al. Defining and adopting clinical performance measures in graduate medical education: Where are we now and where are we going? *Acad Med*. 2019; 94(5):671–677.
11. Nasca TJ, Wiess KB, Bagian JP. Improving clinical learning environments for tomorrow’s physicians. *N Engl J Med*. 2014;370(11):991–992.
12. Wagner R, Weiss KB, Passiment ML, Nasca TJ. Pursuing excellence in clinical learning environments. *J Grad Med Educ*. 2016;8:124–127.
13. Weiss KB, Co JPT, Bagian JP, CLER Evaluation Committee. Challenges and opportunities in the 6 focus areas: CLER National Report of Findings 2018. *J Grad Med Educ*. 2018;10(4 Suppl):25–48.
14. Butler JM, Anderson KA, Supiano MA, Weir CR. “It feels like a lot of extra work”: Resident attitudes about quality improvement and implications for an effective learning health care system. *Acad Med*. 2017;92(7):984–990.
15. Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract*. 2007;12(2):239–260.
16. Govaerts M, Schuwirth LWT, van der Vleuten CPM, Muijtjens AMM. Workplace-based assessment: Effects of rater expertise. *Adv Health Sci Educ*. 2011;16:151–165.
17. Kogan JR, Conforti L, Bernabeo E, Iobst W, Holmboe E. Opening the black box of clinical skills assessment via observation: A conceptual model. *Med Educ*. 2011;45(10):1048–1060.

18. Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: Rethinking the etiology of rater errors. *Acad Med.* 2011;86(10):S1–S7.
19. Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: Assessor cognition from three research perspectives. *Med Educ.* 2014;48(11):1055–1068.
20. Eva KW. On the generality of specificity. *Med Educ.* 2003;37(7):587–588.
21. Roemer BM, Azevedo T, Blumberg B. Looking at graduate medical education through a different lens: A health care system's perspective. *Acad Med.* 2015; 90(9):1231–1235.
22. Sklar DP, Hemmer PA, Durning SJ. Medical education and health care delivery: A call to better align goals and purposes. *Acad Med.* 2017;93(3):384–390.
23. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: Implications for assessment and entrustment. *Med Educ.* 2018;52(Suppl. 1):23–24.
24. 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.



Acknowledgements

My hope was to comprise a dream team of supervisors for my PhD work, and my team has far exceeded my expectations. Cees, the opportunity to have a legend in medical education as my primary supervisor has been the opportunity of a lifetime for me. You take the more challenging aspects of research and make them understandable. You are incredibly giving of your time, and I am still not sure how you answer e-mails and return edits so quickly. Most importantly, you are down to earth and caring.

Eric H., thank you for agreeing to go on this PhD journey with me. I know you are busy and that we did not know each other very well before I started this research. Your willingness to join nevertheless had an undeniable effect on the products of my PhD work. You consistently elevated the rigor and impact of my work with your sage insights.

Jamiu, having another pediatrician who is passionate about the intersection of medical education and patient care quality on my team was a gift. Although we share interests, you brought perspectives that I would not have otherwise seen or considered.

Carol, although you are a member of my supervisor team, you are so much more to me. We met when I was a chief resident. When I joined the Milestone Project Working Group, everyone in the room, except for me, was a highly experienced educator. I expected to learn a lot and offer very little. However, that is not how you work. You gave me the opportunity to be an equal member of the team. Beyond that, you paired me with you for our writing dyads. That early work quickly blossomed: You have been my closest mentor for more than a decade. You are more than a mentor, though. You are a sponsor. You have changed the trajectory of my career twice. Before you retire, we should aim for a trifecta. I can never repay you for the ways that you have shaped my thinking and my career. Please know that I am always looking for ways to pay it forward to extend your legacy.

Lorelei, I can never thank you enough for joining my PhD team for my final study. The opportunity to work with you and extend that work beyond this study is a professional dream. I wanted to add someone with qualitative research prowess to the only qualitative study in my PhD research. You have unparalleled qualifications for that role. You are incredibly giving of your time and talents and have a way of teaching that makes learning easy.

Eric W., Ben, Matt, Dana, and Dan, you have the best program of assessment in the world, and collaborating with you has expanded my thinking and my work in ways I could never have anticipated. I look forward to continuing to foster our collective work in the coming years.

Steve and Rich, thank you for investing time and resources in me to make this PhD journey possible. Without the support of the division and belief in my talents and abilities, this work would not have been possible. I am also grateful for the institutional support in terms of internal grant funding from Cincinnati Children's Hospital Medical Center. This funding has enabled me to redirect some of the money and resources you both committed to me to extending my PhD work and other research beyond what is included in this thesis.

Over the past several years, the American Board of Pediatrics has supported my career in several ways. The leadership of this organization has chosen me to be a member of the Task Force on Subspecialty Clinical Training and Certification, an invited participant in the Future of Testing Conference (where I first met Cees in person and discussed my intentions to do a PhD at Maastricht University), and most recently a member of the Education and Training Committee. They have also provided external funding for my research. Beyond this, they have supported Carol in closely mentoring me.

I have had countless other mentors and colleagues who have influenced my work over the past several years. It is not possible to thank each of you, but I hope you know how grateful I am for your friendship, collaboration, and the opportunity to work with you.

Jayna, thank you for letting me pursue this PhD. Without your love and support, this accomplishment would not have been possible. I am grateful for your willingness to expand your solo parenting to make my increased travel possible despite your being a busy physician as well. Mom, thank you for giving me a love of learning and asking questions from an early age. The best research comes from the best questions, inquisitive minds, and people that are willing to challenge convention and think in different ways. You instilled all of these things in me. Rory, I love spending time with just you and me. There have been times in this journey that I have been able to do less of that, but they have not been common. Please know that these times served to increase my joy and anticipation for our next day or trip together. Adalina, your sweet disposition and love for life is infectious. Thank you for giving me energy and inspiration.



Curriculum Vitae

Daniel J. Schumacher is an associate professor of pediatrics at Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine in Cincinnati, Ohio, United States. He graduated magna cum laude from Luther College (Decorah, Iowa, United States) with a bachelor of arts in biology. He then completed medical school, a three-year pediatric residency, and pediatric chief residency year at the University of Wisconsin before completing a three-year pediatric emergency medicine fellowship at Cincinnati Children's Hospital Medical Center.

Clinically, he works as a pediatric emergency medicine physician in the pediatric emergency department at Cincinnati Children's Hospital Medical Center. Academically, he is the co-director of the Cincinnati Children's Education Research Unit and Education Research Scholars Program, and he is currently externally funded by the Josiah Macy Jr. Foundation, American Board of Pediatrics Foundation, National Board of Medical Examiners, and Association of American Medical Colleges.

His career in medical education began in graduate medical education administration at the Boston Combined Residency Program in Pediatrics (Boston Children's Hospital/Harvard Medical School/Boston Medical Center/Boston University School of Medicine), where he served as an associate program director, clinical competency committee chair, and co-director of the Academy of Medical Education. In 2015, he changed focus from program administration to medical education research. In his administration and research roles, he has established himself as a recognized leader in the area of competency-based assessment at the local, national, and international levels. He has presented and published widely in this area, has led multisite national research studies with this focus, and has participated in efforts in this arena with several organizations, including the Accreditation Council for Graduate Medical Education (ACGME), Association of American Medical Colleges, Royal College of Physicians and Surgeons of Canada, and Josiah Macy Jr. Foundation. Highlights of this work include serving as one of eight members of the Pediatrics Milestone Project Working Group that authored the pediatrics milestones now being used for resident and fellow assessment in all pediatric residency and fellowship programs accredited by the ACGME in the United States.

He has conducted several local and national medical education research studies and published approximately 60 peer-reviewed manuscripts over the last decade, with first author publications in the *Journal of the American Medical Association (JAMA)*, *JAMA Pediatrics*, *Academic Medicine*, *Medical Teacher*, *Pediatrics*, *Academic Pediatrics*, and the *Journal of Graduate Medical Education*.

He is the father to Rory and Adalina and husband to Jayna.



SHE Dissertation 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.

Kikukawa, M. (December 17, 2019) *The Situated Nature of Validity: Exploring the Cultural Dependency of Evaluating Clinical Teachers in Japan*

Kelly, M. (December 10, 2019) *Body of Knowledge: An Interpretive Inquiry Into Touch in Medical Education*

Klein, D. (November 6, 2019) *The Performance of Medical Record Review as an Instrument for Measuring and Improving Patient Safety*

Bollen, J. (November 1, 2019) *Organ Donation After Euthanasia: Medical, Legal and Ethical Considerations*

Wagner-Menghin, M. (September 25, 2019) *Self-Regulated Learning of History-Taking: Looking for Predictive Cues*

Wilby, K. (July 2, 2019) *When Numbers Become Words: Assessors' Processing of Performance Data Within OSCEs*

Szulewski, A. (June 20, 2019) *Through the Eyes of the Physician: Expertise Development in Resuscitation Medicine*

McGill, D. (May 29, 2019) *Supervisor Competence as an Assessor of Medical Trainees: Evaluating the Validity and Quality of Supervisor Assessments*

van Rossum, T. (February 28, 2019) *Walking the Tightrope of Training and Clinical Service: The Implementation of Time Variable Medical Training*

Amalba, A. (December 20, 2018) *Influences of Problem-Based Learning Combined With Community-Based Education and Service as an Integral Part of the Undergraduate Curriculum on Specialty and Rural Workplace Choices*

Melo, B. (December 12, 2018) *Simulation Design Matters: Improving Obstetrics Training Outcomes*

Olmos-Vega, F. (December 7, 2018) Workplace Learning Through Interaction: Using Socio-Cultural Theory to Study Residency Training

Chew, K. (December 6, 2018) Evaluation of a Metacognitive Mnemonic to Mitigate Cognitive Errors

Sukhera, J. (November 29, 2018) Bias in the Mirror: Exploring Implicit Bias in Health Professions Education

Mogre, V. (November 07, 2018) Nutrition Care and its Education: Medical Students' and Doctors' Perspectives

Ramani, S. (October 31, 2018) Swinging the Pendulum From Recipes to Relationships: Enhancing Impact of Feedback Through Transformation of Institutional Culture

Winslade N. (October 23, 2018) Community Pharmacists' Quality-of-Care Metrics: A Prescription for Improvement

Eppich, W. (October 10, 2018) Learning Through Talk: The Role of Discourse in Medical Education

Wenrich, M. (September 12, 2018) Guided Bedside Teaching for Early Learners: Benefits and Impact for Students and Clinical Teachers

Marei, H. (September 7, 2018) Application of Virtual Patients in Undergraduate Dental Education

Waterval, D. (April 26, 2018) Copy but Not Paste, an Exploration of Crossborder Medical Curriculum Partnerships

Smirnova, A. (April 4, 2018) Unpacking Quality in Residency Training and Health Care Delivery

