ACADEMIC MEDICINE
Journal of the Association of American Medical Colleges

Author: Smirnova Alina MD, PhD; Sebok-Syer Stefanie S. PhD; Chahine Saad PhD; Kalet Adina L. MD, MPH; Tamblyn Robyn PhD; Lombarts Kiki M.J.M.H. PhD; van der Vleuten Cees P.M. PhD; Schumacher Daniel J. MD, MEd

Title: Defining and Adopting Clinical Performance Measures in Graduate Medical Education: Where Are We Now and Where Are We Going?

DOI: 10.1097/ACM.0000000000002620

Academic Medicine

DOI: 10.1097/ACM.0000000000002620

Defining and Adopting Clinical Performance Measures in Graduate Medical Education: Where Are We Now and Where Are We Going?

Alina Smirnova, MD, PhD, Stefanie S. Sebok-Syer, PhD, Saad Chahine, PhD, Adina L. Kalet, MD, MPH, Robyn Tamblyn, PhD, Kiki M.J.M.H. Lombarts, PhD, Cees P.M. van der Vleuten, PhD, and Daniel J. Schumacher, MD, MEd

A. Smirnova is a PhD researcher, School of Health Professions Education, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands, and the Professional Performance Research Group, Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

S.S. Sebok-Syer is instructor, Department of Emergency Medicine, Stanford University School of Medicine, Palo Alto, California.

S. Chahine is assistant professor and scientist, Centre for Educational Research and Innovation (CERI), Western University, Ontario, Canada.

A.L. Kalet is professor of medicine and surgery, director of Research on Medical Education Outcomes (ROMEO), Unit of the Division of General Internal Medicine and Clinical Innovation, Department of Medicine; and director of Research, Program on Medical Education and Technology, NYU School of Medicine, New York, New York.

R. Tamblyn is professor, Department of Medicine and Department of Epidemiology and Biostatistics, McGill University; medical scientist, McGill University Health Center Research Institute; scientific director, Clinical and Health Informatics Research Group, McGill University; scientific director of Canadian Institutes of Health Research – Institute of Health Services and Policy Research, Montreal, Canada.

K.M.J.M.H. Lombarts is professor and lead investigator, Professional Performance Research Group, Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

C.P.M. van der Vleuten is professor and scientific director, School of Health Professions Education, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands.

D.J. Schumacher is associate professor, Division of Emergency Medicine, and pediatric emergency physician, Cincinnati Children's Hospital Medical Center/University of Cincinnati College of Medicine, Cincinnati, Ohio.

Correspondence should be addressed to Daniel J. Schumacher, Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center/University of Cincinnati College of Medicine, MLC 2008, 3333 Burnet Avenue, Cincinnati, OH 45229; telephone: (513) 803-2639; e-mail: daniel.schumacher@cchmc.org.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Previous presentations: Oral Presentation, Second World Summit on Competency-Based Medical Education, Basel, Switzerland. August 24, 2018.

Abstract

Assessment and evaluation of trainees' clinical performance measures is needed to ensure safe, high-quality patient care. These measures also aid in the development of reflective, high-performing clinicians and hold graduate medical education (GME) accountable to the public. While clinical performance measures hold great potential, challenges of defining, extracting, and measuring clinical performance in this way hinder their use for educational and quality improvement purposes. This article provides a way forward by identifying and articulating how clinical performance measures can be used to enhance GME by linking educational objectives with relevant clinical outcomes. The authors explore four key challenges: defining as well as measuring clinical performance measures, using electronic health record and clinical registry data to capture clinical performance, and bridging silos of medical education and health care quality improvement. The authors also propose solutions to showcase the value of clinical performance measures and conclude with a research and implementation agenda. Developing a common taxonomy of uniform specialty-specific clinical performance measures, linking these measures to large-scale GME databases, and applying both quantitative and qualitative methods to create a rich understanding of how GME affects quality of care and patient outcomes is important, the authors argue. The focus of this article is primarily GME, yet similar challenges and solutions will be applicable to other areas of medical and health professions education as well.

The goal of graduate medical education (GME) is to prepare residents to provide high-quality care for patients.[1] However, evidence suggests that this goal is not always achieved. As an example, Asch and colleagues found that obstetrics and gynecology residency programs can be systematically ranked based on the complication rates of their graduates.[2] That study showed that gaps in clinical performance can vary based on one's training and that those gaps can persist for years following residency. Although shortcomings in clinical performance are often framed as an educational issue, these immediately and directly affect the safety and quality of patient care.[3] Without understanding the impact of GME on care processes, including quality indicators and clinical outcomes, identification of these learning gaps and improvement of patient care is nearly impossible.[4,5] Complicating this issue further, recent evidence suggests 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.[6] To this end, approaches are needed that explicitly align medical education and training with clinical outcomes in ways that are meaningful for residents. This will ensure safe patient care, meaningful performance assessments throughout residency, resident orientation and fluency with the performance improvement process, and data-driven performance improvement indicators for residents. In fact, our understanding of the relationship among medical education, quality of care, and patient outcomes has advanced minimally over the past 40 years. A 1978 article from McGraw and colleagues from 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.[7] Nearly 40 years later, Weinstein's editorial in the *New England Journal of Medicine* bemoaned the same need in the research agenda for medical education.[8]

In this article, we focus on the purposeful and deliberate use of clinical performance measures in GME as a potential solution to the problem of aligning resident education with clinical care quality. Clinical performance measures can include both clinical care process indicators and patient outcomes that have been adapted for use in GME. We first argue for the need to use more clinical performance measures in GME and then focus our discussion on the challenges of their use in practice. Then we identify a way forward by proposing how resident clinical performance measures can be used to promote medical education research and practice improvement by linking educational objectives with clinical outcomes. The final section provides an overview of common issues and practical challenges of accessing and working with the electronic health record (EHR) and concludes by presenting a research and implementation agenda. While clinical performance measures are applicable for any clinical workplace-based learning settings, including undergraduate medical education, continuing medical education, and other health professions education (pharmacists, nursing), we have chosen to focus on GME in order to build on existing research in this field.

The Need for Clinical Performance Measures in GME

The need to connect GME to care quality and clinical outcomes is relevant worldwide; however, this demand is compounded by increasing costs and public accountability regarding how residents are educated.[9] In the United States alone, for example, GME costs taxpayers 15 billion dollars annually.[10] Yet, we know that residency programs differ measurably in the quality of care delivered by their graduates.[2,11] Hence, we need clinical performance measures not only that evaluate individual residents' performance in practice but also that portray characteristics of their learning environments in order to promote safe care delivery.[12] Clinical performance measures may provide a way forward to help residency programs meet the needs of the patient populations they serve.[13] Additionally, clinical performance measures can serve an accountability function for the public investment in training physicians to deliver high-

value, high-quality care by linking GME to health care quality and clinical outcomes. Furthermore, linking medical education with clinical care quality could be used to justify support of education research with clinical revenue. Finally, clinical performance measures can be used for evaluation of programs and the impact of interventions within the context of the clinical health care system.[14]

Since clinical performance measures can reflect both individual and team performance, it is important to define each measure's use. Often clinical performance measures are described at four levels: individual, program, institutional, and national (see Figure 1).[15] At the individual level, clinical performance measures may be used to describe the care attributable to an individual resident. Examples may include both process and outcome measures, such as the extent to which the resident is able to follow protocols or the accuracy of his or her diagnoses compared to a supervisor. Considering individual residents, clinical performance measures should provide meaningful performance data that facilitate insight into their own clinical performance, guide continuing professional development, and enable development of reflective and effective practitioners. If clearly defined, residents' individual-level measures could also be used to track their progress, for example, through personalized dashboards that could be used to facilitate formative feedback discussions between residents and their supervisors.

Similarly, programs often representing residents' aggregate performance can gain insight into trends regarding graduates' spending, prescribing patterns, or even complication rates. This would not only allow residents to choose educational programs based on specific data, but would also support programs as they track their trends and outcomes as a part of the public accountability inherent in GME.[16,17] At higher levels of measurement, the cumulative impact of several programs offered at teaching hospital(s) or institution(s) can be estimated using clinical performance measures that are relevant to the local population(s). Such higher levels

of impact would require clinical performance measures to be prioritized and selected for appropriateness to the population.

Accordingly, clinical performance measures should ultimately provide evidence of validity and reliability for their purpose and describe the most appropriate level of use.[18] Ultimately, to move the field of clinical performance measures in GME forward, a collaboration between several stakeholders in different countries, consisting of educational evaluators, health services researchers, economists, and clinical educators engaged in GME, is needed. An important first step would, therefore, be to create a common taxonomy and framework of clinical performance measures across a range of learning environments, patient populations, and practice settings. Once a common taxonomy is identified, the second challenge is its measurement.

Introducing the Challenges in Measuring Clinical Performance Measures in GME

Most recently, the widespread use and availability of the EHR, as well as clinical registries and administrative, billing, and insurance data has facilitated the interest in the potential of clinical performance measures in GME.[13,17,19] However, available clinical data are currently not being used to their fullest potential.[19] Part of the problem of moving this understanding forward is the challenge of defining and measuring clinical performance in GME using the electronic and health registry data. In the following section, we explore solutions for solving some of the common measurement challenges related to clinical performance measures.

Challenge 1: Defining clinical performance measures in GME

Clinical performance measures in GME need to be clinically relevant, educationally sensitive, and tailored to the stage of the resident's development on the educational continuum. To be clinically relevant, performance measures need to be either related to or derived from measures of health care processes that have an impact on prevention, morbidity or mortality, patient (and patient-reported) outcomes, or patient experiences.[20] To be useful, they must also

demonstrate evidence of validity and reliability in the context in which they are applied, be affordable to implement, and be comparable in multiple settings.[4] Some progress has already been made in defining clinical performance measures specifically for GME.[15,21–24] As an example of defining new measures, one of us has identified specific educationally sensitive patient-related outcomes, including "patient activation," "clinical microsystem activation," and "health literacy," that were not previously defined.[23,24] However, more work needs to be done to adapting existing clinical performance measures for use with residents given that they are not always afforded the same opportunities as attending physicians.[25] In particular, one of us has recommended using process measures and taking the prevalence and treatability of the condition and its population effects into account.[26]

Challenge 2: Measurement issues

As we move toward clearly defining and adapting clinical performance measures for their appropriate level of use, there remain significant measurement barriers (see Table 1). This list, however, is by no means exhaustive or mutually exclusive, as some challenges are applicable at more than one level. The most important challenges include attribution/contribution as well as aggregation and nesting. We address these measurement challenges in more detail below.

Attribution and contribution. Regardless of the approach one takes to clinical performance measures, either defining new measures or adapting existing ones, the fact that patient care and outcomes are multifactorial and complex makes using clinical data as a proxy for residents' clinical performance very challenging in practice. This challenge has been approached in two ways. First, a portion of patient care and/or outcomes from care provided by residents can be considered as attributable to a resident's actions. This approach supposes that one can specify which actual member(s) of the health care team performed the actions that led to particular care quality and outcomes. Attributing care to residents is extremely difficult. 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 other members of the team. However, while 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.[27] Considerable progress has been made in attributing quality measures to individual residents, which is promising. For example, one of us has created a process for defining resident-sensitive quality measures in GME.[22] Beyond GME, Kaplan and colleagues demonstrated the ability to identify individual physician effects among diabetes process and intermediate outcome measures.[28] Others have argued for creating composite measures using standard psychometric techniques (e.g., item-response theory and factor analysis) as a way to capture a wider range of residents' clinical activities.[29–31]

The second approach—termed contribution—addresses these shortcomings by seeing residents as contributing to care processes and outcomes. Contribution analysis builds a case for the relative contribution residents make to quality care through their actions within a system.[32] For example, a contribution analysis might focus on describing the pathways through which a resident capable of performing a defined competency (e.g., communication skills) contributes to proximal and distal clinical outcomes.[32] Contribution is also useful when considering rare events, such as errors. Such events are typically multifactorial, with several contributing individuals or factors.

Both the lens of attribution and that of contribution have their benefits and their drawbacks. Contribution analysis builds a rigorous case for how residents contribute to quality care, but it is a labor-intensive process. Considering attribution may not be as labor intensive because it

requires less effort to make a case for the link to a particular resident, but it fails to consider the complexity of care and the value of collective competence.

While collective competence is ultimately what ensures high quality care,[33] viewing clinical processes and outcomes from both contribution *and* attribution perspectives is one way to account for such interdependence in clinical practice. Meaningful clinical performance measures in GME, therefore, need to consider not only residents, but also their interdependence with other health care professionals. Some of us have characterized the interdependence that exists between individuals within a health care team as "coupling."[27] We suggest that, given the nature of clinical supervision in GME, various "configurations of coupling" exist that ought to be considered when using clinical performance measures with graduate residents. Therefore, attempts to define and develop clinical performance measures in GME should include a variety of individual, coupled, and team-based clinical metrics that are meaningful and relevant for GME and the performance level of interest.

Aggregation and nesting. Institutions tend to describe their program's performance by aggregating their residents' performance. While simple aggregation may allow a single program to track its performance over time, provided the context of the program does not change, aggregation becomes more important when programs are compared to each other. When comparing programs, it is imperative to estimate the minimum number of residents and relevant case-mix variables needed for a reliable measurement. Secondly, residents' individual performance is similarly based on a sample of the patients for whom they have cared. In other words, patients are "nested" within residents (just as residents are "nested" within clinical teaching departments or programs). Multilevel analysis techniques are, therefore, the most appropriate when analyzing such nested data. The use of multilevel analysis techniques may also allow the inclusion of case-mix variables at different levels, such as patient variables, as well as resident and department characteristics (see Table 1).

Challenge 3: Using EHR and clinical registry data to capture clinical performance

This section addresses challenges of sourcing data for clinical performance measures specifically designed for the purpose of assessing residents. With electronic data collection and reporting becoming ubiquitous, EHRs and/or various clinical registries are the most common sources of clinical performance data.[13] We have chosen to focus here on the EHR and/or clinical registries as sources of clinically relevant data since they contain a wealth of already-collected clinical information that can be analyzed. Furthermore, focusing on the EHR does not add an extra burden of data collection and reporting.

While using EHR or clinical registry data may provide access to a vast amount of clinical data, there are a number of issues regarding data access and storage that need to be considered when doing so. Chief among these are issues pertaining to the privacy and confidentiality of patient information, data ownership, and alignment of data storage and access with current national and international laws and regulations.[34] Differences in privacy laws and policies can complicate the collection and exchange of data, thus hindering efforts for comparative research and the creation of a unified framework and taxonomy of clinical performance measures. While individual institutions may only need to make sure they follow local privacy laws, multi-institutional or international initiatives may experience greater obstructions due to differences in privacy laws and even different types of EHR vendors. For longitudinal studies involving large datasets from multiple data sources, a potential solution to overcoming differences in privacy regulations is to apply an institutional review board data repository approach to manage ethical considerations for collecting, pooling, and sharing data.[35]

Aside from accessing EHR or registry data, it can be particularly challenging to link such data with existing educational datasets. Currently, large educational databases are being compiled to amass, analyze, and compare longitudinal data from medical students and residents as well as training programs across the medical educational continuum.[34] These databases are large

repositories of educational (assessment) data on individual residents. If linked to clinical performance data, they have an immense potential for answering many questions provided appropriate "big data" analysis techniques are used and care is taken to minimize potential sources of bias.[17,19,36]

Finally, even if the EHR or clinical registries can be harnessed to provide data about clinical performance, these data do not constitute performance assessment, and assessment in turn does not constitute meaningful feedback required for improvement.[25] If clinical performance indicators become a goal in themselves, they can potentially devalue the focus on improvement and other important, and sometimes unmeasured, aspects of patient care.[37] In order for clinical performance measures to lead to improvements in health care delivery and patient outcomes, processes need to be in place that facilitate their implementation in a way that supports meaningful feedback and continuous improvement activities on all levels. In the following section we lay out the necessary steps to support further implementation of clinical performance measures in practice.

Challenge 4: Bridging the silos of medical education and clinical quality improvement

The complexities involved in defining, measuring, and using clinical performance measures in GME as well as lack of funding have flummoxed progress in this area.[38,39] Overcoming the challenges to the widespread use of clinical performance measures in GME will require a collaborative effort among medical education programs, health services researchers, residency program leaders, quality improvement advisors, hospital administrators, data managers, and legal/privacy officers, all committed to tackling these issues while sharing knowledge and expertise with each other. We believe an international working group should be convened to drive these collaborative efforts forward and develop standards around appropriate use of clinical performance data. Perhaps most important to success in this area is that the efforts of individuals working in this field must bridge the traditional "silos" that exist in academic

medicine and clinical care. For instance, given that clinical performance data are fundamentally important to both groups it would be important to establish the vision and leadership to link medical education and clinical quality improvement.[40] Such efforts need to be accompanied by financial arrangements that facilitate the process because current financing structures in GME and health care delivery perpetuate the existing silos. Thus, alternative or well-aligned sources of funding may be needed to encourage cooperation.

Defining the Path Forward

As previously mentioned, coordinated research efforts with input from many different stakeholders are required to overcome the challenges we have described. Accordingly, we have formed the International Collaborative on Clinical Outcome Research in GME, specifically aimed at tackling these challenges. As a collaborative, we have defined the following research agenda:

- Developing a common taxonomy of clinical performance measures for different specialties as well as a framework for understanding how GME can affect care quality and patient outcomes.
- Defining uniform measures for international, comparative research that supports the reliability and validity of both new and existing clinical performance measures for use across various learning environments and phases of training. This would facilitate large-scale research and maximize the adaptability of the measures.
- Linking clinical performance measures with large-scale GME databases to establish causal links between GME, quality care, and patient outcomes, potentially using advanced methodologies and big data techniques. One way to explore the effects of residency training on quality care and patient outcomes, including the ability to make any causal inferences, is to study cohorts of residents as they go through their clinical training, enter unsupervised practice, and continue their practice as young faculty.

This would allow for measuring the effects of residency training on the individual residents' care quality as well as allow for analysis of other unintended effects of training on patient outcomes.

- Applying both quantitative and qualitative methods to create a rich understanding of how GME affects quality of care and patient outcomes.

As we develop robust clinical performance measures, the focus should remain on:

- Implementing real-time clinical performance measures that align with the work of residents in settings where these data are already available, but not yet commonly used.
- Identifying gaps in existing measures and developing new clinical performance measures in GME.
- Evaluating what measures work, for whom, and under what circumstances.
- Defining the necessary investment, both in monetary and personnel support, to harness and organize clinical performance measures from the EHR and clinical registries.
- Defining a long-term vision, or desired state, for how clinical performance measures should be used in GME.
- Partnering with leaders in fields of patient safety and quality improvement to use existing clinical data for clinical performance measurement in GME.

To Conclude

While clinical performance measures hold great potential, their use is hampered by challenges of defining, measuring, and extracting these measures, as well as the silos of medical education and clinical quality improvement. These challenges can only be overcome through the coordinated collaborative efforts of broad-based stakeholders both nationally and internationally committed to implementing and studying clinical performance measures in GME.

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;376:1923–1958.
2. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes, JAMA. 2009;302:1277–1283.
3. Kogan J, Conforti L, Iobst W, Holmboe E. Reconceptualizing variable rater assessments as both an educational and clinical care problem. Acad Med. 2014;89:721–727.
4. Gruppen L, Frank JR, Lockyer J, et al. Toward a research agenda for competency-based medical education. Med Teach. 2017;39:623–630.
5. Gruppen LD, ten Cate O, Lingard LA, Teunissen PW, Kogan JR. Enhanced requirements for assessment in a competency-based, time-variable medical education system. Acad Med. 2018;93(3Suppl):S17–S21.
6. 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.
7. Magraw RM, Fox DM, Weston JL. Health professions education and public policy: A research agenda. Journal of medical education. 1978;53:539–546.
8. Weinstein DF. Optimizing GME by measuring its outcomes. N Engl J Med. 2017;377:2007–2009.
9. Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. Acad Med. 2004;79:955–960.
10. Iglehart JK. Institute of Medicine report on GME: A call for reform. N Engl J Med. 2015;372:376–381.

11. 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.

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. 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.

14. Dauphinee WD. The role of theory-based outcome frameworks in program evaluation: Considering the case of contribution analysis. Medical Teacher. 2015;37:979–982.

15. 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.

16. Weinstein DF, Thibault GE. Illuminating graduate medical education outcomes in order to improve them. Acad Med. 2018;93:975–978.

17. 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.

18. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane's framework. Med Educ. 2015;49:560–575.

19. Arora VM. Harnessing the power of big data to improve graduate medical education: Big idea or bust? Acad Med. 2018;93:833–834.

20. Lazar EJ, Fleischut P, Regan BK. Quality measurement in healthcare. Annu Rev Med. 2013;64:485–496.

21. Kalet A. The state of medical education research. Virtual Mentor. 2007;9:285–289.

22. 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;1071–1078.

23. 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.

24. 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.

25. Sebok-Syer SS, Goldszmidt MA, Watling CJ, Chahine S, Venance SV, Lingard LA. Using electronic health record data to assess residents' performance in the clinical workplace: The good, the bad, and the unthinkable. Acad. Med. [In press].

26. Tamblyn R. Outcomes in medical education: What is the standard and outcome of care delivered by our graduates? Adv Health Sci Educ Theory Pract. 1999;4:9–25.

27. 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.

28. 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;47:378–387.

29. van Doorn-Klomberg AL, Braspenning JC, Feskens RC, Bouma M, Campbell SM, Reeves D. Precision of individual and composite performance scores: the ideal number of indicators in an indicator set. Med Care. 2013;51:115–121.

30. Chen TT, Lai MS, Lin IC, Chung KP. Exploring and comparing the characteristics of nonlatent and latent composite scores: Implications for pay-for-performance incentive design. Med Decis Making. 2012;32:132–144.

31. Silverman M, Povitz M, Sontrop JM, Shariff SZ. Antibiotic prescribing for nonbacterial acute upper respiratory infections in elderly persons. Annals of Internal Medicine. 2017;167:758–759.

32. 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.

33. Lingard L. Paradoxical truths and persistent myths: Reframing the team competence conversation. Journal of Continuing Education in the Health Professions. 2016;36(Suppl 1:S19–S21.

34. 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;91:690–695.

35. Thayer EK, Rathkey D, Miller MF, et al. Applying the institutional review board data repository approach to manage ethical considerations in evaluating and studying medical education. Medical education online. 2016;21:32021.

36. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: New opportunities, familiar challenges. Clinical epidemiology. 2017;9:245–250.

37. Werner RM, Asch DA. Clinical concerns about clinical performance measurement. Ann Fam Med. 2007;5:159–163.

38. Cook DA, West CP. Perspective: Reconsidering the focus on "outcomes research" in medical education: A cautionary note. Acad Med. 2013;88:162–167.

39. Gebauer S, Steele E. Questions program directors need to answer before using resident clinical performance data. J Grad Med Educ. 2016;8:507–509.

40. Gupta R, Arora VM. Merging the health system and education silos to better educate future physicians. JAMA. 2015;314:2349–2350.

References cited only in the table

41. Weifeng W, Hess BJ, Lynn LA, Holmboe ES, Lipner RS. Measuring physicians' performance in clinical practice: Reliability, classification accuracy, and validity. Evaluation & the Health Professions. 2010;33:302–320.

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

43. Hong CS, Atlas SJ, Chang Y, et al. Relationship between patient panel characteristics and primary care physician clinical performance rankings. JAMA. 2010;304:1107–1113.

44. Martsolf GR, Carle AC, Scanlon DP. Creating unidimensional global measures of physician practice quality based on health insurance claims data. Health Serv Res. 2017;52:1061–1078.

45. Smirnova A, Lombarts KM, Arah OA, van der Vleuten CP. Closing the patient experience chasm: A two-level validation of the Consumer Quality Index Inpatient Hospital Care. Health Expectations. 2017;20:1041–1048.

46. Silkens ME, Smirnova A, Stalmeijer RE, et al. Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. Med Teach. 2016;38:476–481.

47. 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.

48. Bansal N, Simmons KD, Epstein AJ, Morris JB, Kelz RR. Using patient outcomes to evaluate general surgery residency program performance. JAMA Surgery. 2016;151:111–119.

49. Sequist TD, Schneider EC, Li A, Rogers WH, Safran DG. Reliability of medical group and physician performance measurement in the primary care setting. Med Care. 2011;49:126–131.

50. Arah OA. Bias analysis for uncontrolled confounding in the health sciences. Annu Rev Public Health. 2017;38:23–38.

51. Thompson CA, Arah OA. Selection bias modeling using observed data augmented with imputed record-level probabilities. Annals of Epidemiology. 2014;24:747–753.

Figure Legends

Figure 1

Four levels of use for clinical performance measures in graduate medical education

Abbreviation: GME indicates graduate medical education.

Table 1

Common Challenges and Potential Solutions in Using Clinical Performance Measures in GME

Potential uses	Potential challenges	Potential solutions
Assessment of individual residents in the workplace	• Attribution/coupling • Transferability—performance in one domain may not reflect performance in other domains • Lack of variation in the outcome, nesting of patient data within residents • Residents nested within faculty supervisors • Patient case mix differences impeding comparison between residents	• Consensus group methods to develop resident-sensitive quality measures[22] • Use educationally sensitive patient outcome measures[15,23] • Use statistical approached to create composite measures where differences between residents explain a relatively large proportion of variance[28,41,42] • Ensure clinical performance measures are adjusted for patient case mix[31,43] • Apply classic psychometric analysis to establish the validity and reliability of (composite) measures fit for purpose[29,30,44] • Build models to measure interdependence[27]
Evaluation of a training program or a specific program-wide intervention	• Aggregation of patient data across residents and associated decrease in variance of the outcome measure(s) • Attribution of (distant) clinical outcomes • Difference covariates become important on the program level • Unintended consequences of program evaluation • Incomplete data (e.g., multiple institutions) • Lack of consensus on outcomes due to differences between specialties (surgical vs. non-surgical, specialty-specific outcomes) • Data collection and curation given the transient nature of residents in residency programs	• Ensure validity and reliability at the level of aggregation or utilize multilevel statistical techniques[45,46] • Calculate intra-class correlations to ensure sufficient variance in outcome measures • Employ longitudinal study designs to study outcomes delivered by residents after graduation[2,26,47,48] or use program-level outcomes[15] • Contribution analysis[32] • Harmonization of EHR across institutions • Ensure a standardized resident portfolio that includes clinical performance measures across training sites • Ensure minimum number of patients per site are included for reliable evaluation of a program[49]

Impact of several programs within the context of the clinical healthcare system or local population	• Aggregation of patient data across programs and associated decrease in variance of the outcome measure(s) • Different covariates become important on institutional level • Differences between specialties (see above) and hospitals (large vs. small, community vs. academic, number of accredited training programs)	• Ensure validity and reliability at the level of aggregation or utilize multilevel statistical techniques • Calculate intra-class correlations to ensure sufficient variance in outcome measures • Create a composite score based on pre-defined clinical outcome measures from each specialty or use standardized evaluations (patient-experience questionnaires or 360-degree evaluations) or use institutional-level outcomes[15] • Ensure appropriate case mix correction • Consider statistical analyses for potential unmeasured confounding and selection bias[12,50,51]
Collective contribution of GME to the health of whole communities	• Aggregation and storage of large amounts of data • Data ownership • Dataset quality and compatibility • Big data requiring new data management techniques • Prioritization/selection of measures	• Consensus on a national level on which outcome measures must be achieved[15] • National guidelines for data collection, storage and use, data managers with expertise in big data management and analysis[34]

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

Figure 1