OPEN

Orthopaedic Surgery Residency Milestones: Initial Formulation and Future Directions

S. Elizabeth Ames, MD
Brent A. Ponce, MD
J. Lawrence Marsh, MD
Stanley J. Hamstra, PhD

From the Department of Orthopaedics and Rehabilitation, College of Medicine, University of Vermont, Burlington, Vermont (Dr. Ames), the Department of Orthopaedic Surgery, University of Alabama, Birmingham, AL (Dr. Ponce), Department of Orthopaedics and Rehabilitation, University of Iowa, Iowa City, IA (Dr. Marsh), Milestones Research and Evaluation, Accreditation Council for Graduate Medical Education, Chicago, IL (Dr. Hamstra), Faculty of Education, University of Ottawa, Ottawa, ON, Canada (Dr. Hamstra), and Department of Medical Education, Feinberg School of Medicine, Northwestern University, Chicago, IL (Dr. Hamstra).

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.jaaos.org).

J Am Acad Orthop Surg 2020;28: e1-e8

DOI: 10.5435/JAAOS-D-18-00786

Abstract

Milestones specific to orthopaedic surgical training document individual resident progress through skill development in multiple dimensions. Residents increasingly interact with and are assessed by surgeons in both academic and private practice environments. Milestones describe the skills that support competence. One of the primary goals of milestones is to provide continuous data for educational quality improvement of residency programs. They provide a dialogue between surgeons who supervise residents or fellows and the program's Clinical Competency Committee throughout a resident's education. The orthopaedic milestones were developed jointly by the Accreditation Council for Graduate Medical Education and the American Board of Orthopaedic Surgery. The working team was designed with broad representation within the specialty. The milestones were introduced to orthopaedic residencies in 2013. Orthopaedics is a 5-year training program; the first comprehensive longitudinal data set is now available for study. This summary provides historical perspective on the development of the milestones, state of current milestone implementation, attempts to establish validity, challenges with the milestones, and the development of next-generation assessment tools.

Orthopaedic surgery has evolved to a highly procedurally based specialty. The optimal development of a surgeon depends on technical skills, clinical care, and professional interactions. Through accreditation, surgical training programs are required to teach and measure milestone achievement of residents in these areas. Surgical training programs vary: private practice environments and highly specialized rotations or electives can all be part of an Accreditation Council for Graduate Medical Education (ACGME)-approved residency. Surgeons educating residents in all of these formats require assessment skills. Surgical practices that receive orthopaedic residents as young attendings will benefit from understanding the current training and assessment processes.

The ACGME launched the Next Accreditation System in 2013 and this new process introduced the milestones.[1] The original intention of Next Accreditation System was for the ACGME to provide milestones as a guiding framework for reporting through the Clinical Competency Committee (CCC) and to allow each program the freedom to continue to use their current and/or develop new assessment tools. Changes and needs for orthopaedics were reviewed in 2015.[2] Orthopaedic faculty and program directors (PDs) have

always exercised their judgment about resident performance; that has not changed. Milestones were intended to supplement that judgment. Milestones were designed as the first opportunity to formally accumulate assessments of skills and behaviors and to have all residents across the specialty evaluated on the same set of competencies. They were not intended to replace existing evaluations but were intended to provide a common framework for reporting resident competence. The program's effectiveness plays a key role in assuring the public that graduating residents possess what is needed for competent orthopaedic practice. Because training programs have grown and evolved over time, the assessment of residents and fellows is no longer limited to academic faculty. The purpose of this article is to describe the rationale for milestones, current challenges to implementation within programs, and ideas for improving usability and validity for all surgeons responsible for educating the next generation of orthopaedic surgeons.

Development of the Orthopaedic Milestones

Milestones document residents' achievement during training. They provide the PD with detailed information about the strengths and weaknesses of the program. Milestones were envisioned as descriptors of resident competence in medical knowledge (MK), technical skills, and attitudes within the larger framework of six core competencies.[3] The competencies that led to the milestones were developed by the ACGME and the American Board of Medical Specialties in 1999: patient care (PC), MK, professionalism (PROF), interpersonal and communication skills, systems-based practice (SBP), and practice-based learning and improvement (PBLI).[3-6]

A workgroup of orthopaedic surgeons was formed in 2011, including experts from each of the major subspecialties. This group spent most of their time developing milestones for PC and MK based on a growth model of learning complex skills.[6] The group was free to structure and form the milestones in the way that they felt would best reflect the needs of the specialty. The overall structure is shown in Supplemental Table 1 (Supplemental Digital Content 1, http://links.lww.com/JAAOS/A376). A full list of workgroup members is shown in Table 1. To help reduce confusion with the organizations involved in resident and fellow education, a list of the participating organizations and their roles has been summarized in Supplemental Table 2 (Supplemental Digital Content 2, http://links.lww.com/JAAOS/A377).

The workgroup decided that each subspecialty should have a pair of representative MK and PC milestones relevant to their area. They chose 16 focus areas thought to represent a spectrum of patient conditions covered by the specialty—think of this like an educational biopsy. Each of the 16 had both a PC and MK milestone, resulting in 32 milestones. This group also worked with the ACGME to develop milestones in the other four competencies, resulting in a total of 41 milestones overall. However, this seemingly rational approach left orthopaedics with the most individual milestones of all specialties.

Final approval for the current milestones set was complete in 2012. The final draft was shared that year at the June 28th meeting of the Council of Orthopaedic Residency Directors/American Orthopaedic Association (CORD/AOA) Annual Meeting in National Harbor, Maryland. This allowed PDs, specialty leaders, and other members of the orthopaedic community to provide comment and feedback. General challenges were discussed but details of the process, relationship to previous evaluation standards, and effects of practical use were difficult to predict. No formal concerns were raised through CORD/AOA or directly to the ACGME working group. The most consistent feedback was based on specific PC and MK milestones without broad consensus. The milestones developed by the committee were then finalized.

The milestones then underwent field testing and a survey requesting PD opinions as to whether the milestones would represent a realistic progression within orthopaedic training. The product was reported in the *Journal of Graduate Medical Education*,[7] and orthopaedic surgery was selected in the group of phase 1 specialties that began collecting data in July 2013. The leaders realized at that time that we would not have a detailed feedback on the feasibility or effectiveness of milestones until several years of experience. The earliest aggregate data were presented at the 130th meeting of CORD/AOA in 2017 in anticipation of developing a working research team to look at June 2018 data and beyond.

Using milestones data for accreditation of programs was initially considered by the ACGME as part of the original Outcomes Project.[6] This direction was changed after subsequent presentation and discussion by the ACGME and the member Boards of the American Board of Medical

None of the following authors or any immediate family member has received anything of value from or has stock or stock options held in a commercial company or institution related directly or indirectly to the subject of this article: Dr. Ames, Dr. Ponce, Dr. Marsh, and Dr. Hamstra.

Specialties in late 2015 and again in 2016. Feedback was obtained from meetings with most core specialty PDs' associations, listening sessions at the ACGME annual education conference, and other venues. Challenges with data collection, some issues with content of milestones,[8,9] and the issue of using milestones data for accreditation were discussed.

The group reached consensus that milestones data should not be used for accreditation decisions until/unless sufficient research demonstrated defensible levels of validity for such a purpose. The focus was to report the annual data to the ACGME and for programs to use their data and the national data for continuous quality improvement to enhance educational outcomes. This conclusion has been supported in subsequent analyses and recommendations.[10-12] Current ACGME policy is that review committees do not use individual resident milestones data in making program accreditation decisions (Table 2).

The current statement of intended use is to describe the educational trajectory of a resident through their program and use data describing all residents to explore variation in educational outcomes.[13,14] Benefits of using milestones data in this continuous quality improvement mode have been reported recently and include the ability to identify struggling residents earlier, identify gaps in the curriculum, and focus on using the information in discussions with a resident to create an individualized improvement plan.[8,15]

Table 1

Orthopaedic Surgery Milestone Development Groups (Association at the Time of Development, Not during Publication)

Working group	
Peter J. Stern, MD	University of Cincinnati College of Medicine
Stephen Albanese, MD	SUNY Upstate Medical University
Mathias Bostrom, MD	Hospital for Special Surgery
Charles S. Day, MD	Harvard Medical School/Beth Israel Deaconess Medical Center
Steven L. Frick, MD	University of Central Florida
William Hopkinson, MD	Loyola University Stritch School of Medicine
Shepard Hurwitz, MD	University of North Carolina
Keith Kenter, MD	University of Cincinnati College of Medicine
John S. Kirkpatrick, MD	University of Florida College of Medicine
J. L. Marsh, MD	University of Iowa
Anand M. Murthi, MD	MedStar Union Memorial Hospital
Terrence Peabody, MD	Northwestern University
Lisa A. Taitsman, MD	University of Washington
Brian C. Toolan, MD	University of Chicago Medicine
Kristy Weber, MD	Johns Hopkins School of Medicine
Rick W. Wright, MD	Washington University School of Medicine
Pamela Derstine, PhD, MHPE	Executive Director, Orthopaedics Review Committee
Laura Edgar, EdD, CAE	Senior Associate Director of Outcome Assessment, ACGME
Advisory group	
Stephen Albanese, MD	SUNY Upstate Medical University
Timothy Brigham, MDiv, PhD	Chief of Staff, Senior VP: Education
Marybeth Ezaki, MD	University of Texas Southwestern
Richard H. Gelberman, MD	Washington University School of Medicine
Christopher Harner, MD	University of Texas Houston
Shepard R. Hurwitz, MD	University of North Carolina
Joseph D. Zuckerman, MD	New York University

Current State of Milestones

Milestones are a report of programmatic judgment of a resident's level of competence based on multiple assessments. Under the current system assessment, data are collected from individual faculty raters and collated by the CCC. Every 6 months, the CCC deliberates and generates reporting milestones for each resident. The milestone ratings sent to the ACGME represent the resident's competence at that point in time. In this way, the milestones have become a summative assessment. Milestones in aggregate should also be used for program improvement by the PD, CCC, and the Program Evaluation Committee (PEC).

The PEC should look at overall milestone reports within the program and is charged with biannual review of data. The PEC comprises faculty members, residents, and others expert in the program's overall design. It is ultimately responsible for the design and implementation of quality improvement efforts regarding the educational program. This is one way in which the milestones and other resident assessments can be

Table 2

Milestones Not Intended for Accreditation of Programs

How will the Milestones be used by the ACGME?

> Resident/fellow performance on the Milestones will become a source of specialty-specific data for the specialty Review Committees to use for continuous quality improvement in assessing programs and for facilitating improvements to program curricula and resident assessment. In the early phase, the Milestone data will be used as formative assessment of the quality of residency/fellowship programs. Review Committees will not judge a program based on the level assessed for each resident/fellow. . . [emphasis added]

Copyright Accreditation Council for Graduate Medical Education. Used with permission. Milestones FAQ. http://www.acgme.org/Portals/0/MilestonesFAQ.pdf. Accessed October 24, 2018.

Statement at the beginning of each Milestones set (p.iv):

> Level 4 is designed as the graduation target but does not represent a graduation requirement. Making decisions about readiness for graduation is the purview of the residency program director. Study of milestone performance data will be required before the ACGME and its partners will be able to determine whether milestones in the first four levels appropriately represent the developmental framework, and whether milestone data are of sufficient quality to be used for high-stakes decisions. . . [emphasis added]

Copyright Accreditation Council for Graduate Medical Education and American Board of Orthopedic Surgery. Used with permission. Milestones FAQ. https://www.acgme.org/Portals/0/PDFs/Milestones/OrthopaedicSurgeryMilestones.pdf?ver=2015-11-06-120524-887. Accessed October 24, 2018.

ACGME = Accreditation Council for Graduate Medical Education

used to provide feedback to the program and for the PD to adjust overall program design.

This process provides continuous data for educational quality improvement of residency programs. Milestones track individual progress through each area. The incorporation of the milestones is designed to guide residents toward understanding what is expected of them in key areas. The specialty of orthopaedics is a 5-year training program; ACGME specifically avoided an association between postgraduate year and level. Milestone level 1 corresponds with an entry-level resident with some background competency in general medicine. Levels 2 and 3 represent an evolution from a resident requiring substantial supervision to one who can be expected to complete most of the basic tasks associated with clinical care and technical skills. Level 4 represents a proficient learner able to demonstrate independence, and level 5 represents a very high level of achievement more likely to reflect advanced training such as fellowship. This reflects skill development from beginner to proficient, following a model of skill acquisition through instruction and practice that has been adapted for the context of competency-based medical education.[6]

Milestone rating sequences will vary because individual residency programs have the freedom to design their own curricular structures. A system that recognizes that individual residency programs vary allows programmatic and individual innovation for format, timing, and curriculum. Optimal measurement is not prescriptive but should be informative. This ensures that individual residency programs can measure where they are relative to general expectations and to the overall standard encoded within the milestones themselves.

Milestones Challenges

The orthopaedic milestones have presented several challenges to orthopaedic programs and PDs. The intended use of milestones requires a thoughtful data-driven process for every resident every 6 months. The goal is laudable, but the implementation has been challenging, partly because of the perception—by some—that there are too many milestones. This opinion is based on stakeholder opinion collected by the ACGME milestones staff during the period of initial rollout.[8-12] There appears to be a general perception that the large number of milestones in orthopaedics can impede the work of the CCC. If programs were expected to add a set of independent assessment tools for faculty to complete on each resident, the consensus is that it would be untenable. Thus, in practice, programs tend to ask individual faculty to use the milestones reporting forms for workplace-based assessments, and the CCC then collates data across all available forms for reporting to the ACGME. This can result in the perception that it is more of a box-checking exercise than a thoughtful evaluation. With the current set of 41 milestones in orthopaedic surgery, a CCC of a 30-resident program must make 1,230 (30×41) separate milestone decisions at each CCC meeting, not to mention the challenges in asking all faculty raters to adapt to a new system of assessment. PDs had little guidance when first tasked with this job. Some shortcuts were possible, but the task felt overwhelming which has led to concerns about the quality of the data. A survey of PDs on

current practices would help to further identify specific challenges and inform improvements to the process.

The complication of assessing nontechnical skills has been challenging in all specialties. PROF and communication skills are difficult to measure, but there is now a fledgling literature on this.[16-20] PBLI and SBP competencies may relate to clinical performance measures already collected by the healthcare institution.[21,22] Some resources have been developed and provided in response to these concerns, but there is still much work to be done. Studies looking at CCC processes have recently been published, which may help identify best practices regardless of specialty.[8,9,23] Reported options include breaking down the CCC to evaluate different groups of residents, such as one for senior and another for junior residents. Others include presenting the CCC with averaged data from each clinical service rather than milestones from every faculty member. Some programs have developed detailed faculty development resources and systems for mapping assessment tools onto milestones.[24,25] One task for the CORD could be to identify best practices among orthopaedic surgery PDs or programs of different sizes and disseminate them to the community at large.

Another milestones challenge has been the perceived lack of normative data for a program to assess its results compared with other programs. These reports have been available on the ACGME website since 2016[26] but have not been broadly disseminated within the orthopaedic surgery community (Figure 1). Similarly, advice from orthopaedic educators to make these data useful for specific residency program improvement has been lacking. In this situation, it is hard for PDs and faculty to see the value proposition of the milestones data and the return on investment in time and resources used to collect the data. Early information has been provided through CORD at annual meetings, but there is much PD education still to do.

Milestones data also have the potential to help identify individual residents who are having difficulty earlier than previously available.[15] This is one target for specific research, but this potential has not been fully realized to date in orthopaedic surgery. It has been difficult for PDs to use milestones data to assess individual resident performance curves for comparative analysis, and there are reasons for this. Orthopaedic milestones data were first collected from programs in December 2013, which means that complete longitudinal cohort data has only been available since June 2018. Longitudinal analyses from other specialties with shorter training programs are just now being published and may offer clues for similar analysis with the orthopaedic milestones. Studies are currently underway with the orthopaedic data set to provide guidance for PDs.

There need to be more widely available tools and resources to assist programs in collecting, analyzing, and interpreting the milestones data for their intended purpose, guiding curricular design[2] rather than individual assessment. PD organizations such as CORD and the Milestones Department of the ACGME are the frontline groups for working together to create tools in orthopaedics. This is the only way to tackle the lingering sense of frustration that these data have not yet been used this way.

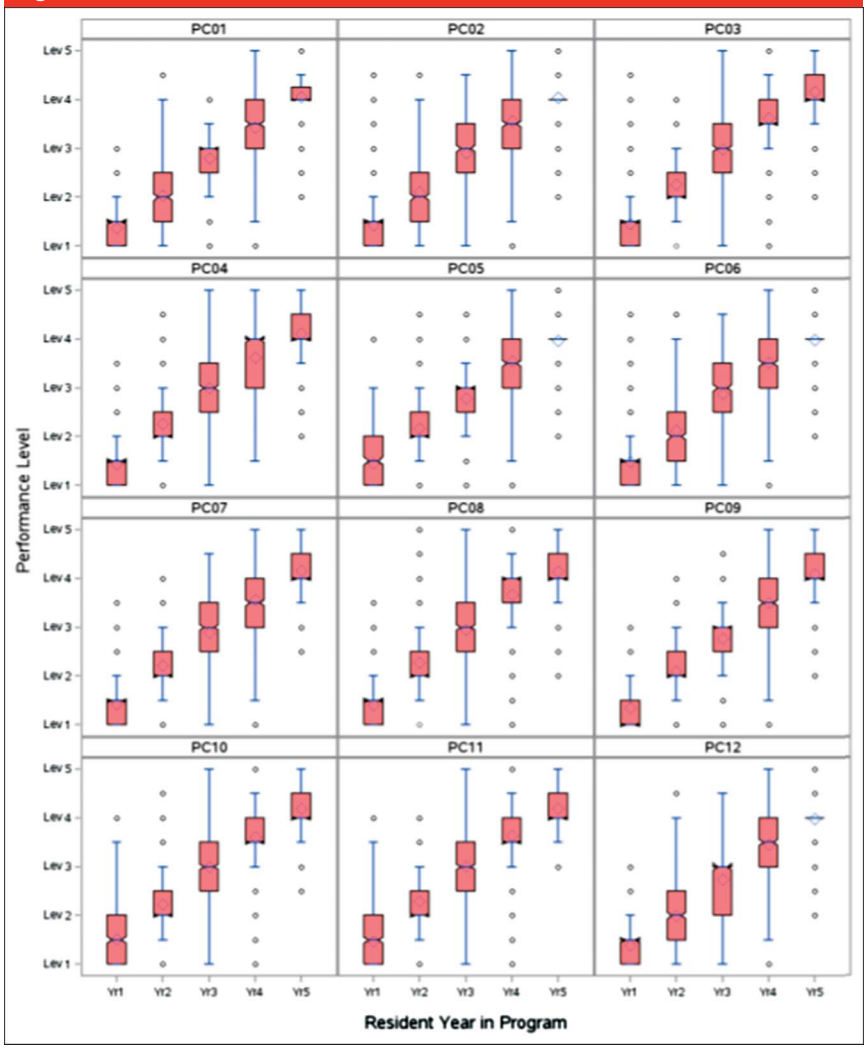
The above challenges have made it difficult for orthopaedic PDs to make good use of the milestones to promote faculty development and facilitate communication with residents. Programs manage the information locally. Some programs have incorporated the milestones into their rotation end evaluations. Other programs have held onto former evaluation systems and use the resulting data to inform the milestones, which can be daunting due to a lack of meaningful assessment tools that link directly to the milestones. There are examples in other specialties but these have not been developed yet in orthopaedics.[24,27] Establishment of national best practices shared through PD education would be useful.

The choice of clinical entities (rotator cuff, supracondylar humerus, etc.) for PC and MK milestones was made to assess at least one area relevant to each orthopaedic subspecialty. Not all of these are relevant at each point in residency training, so data are requested when no updated information is directly available (eg, pediatric septic hip). This has caused one of the major frustrations that could be addressed with redesign.[2] Other specialties chose a different approach focusing more on general tasks that can be assessed in any clinical area or diagnosis—examples include history-taking, managing perioperative care, and general procedural skills—as one broad approach and then creating specific milestones where a highly specialized skill set is required. The specializations are often population based (eg, children versus adults) but could be anatomic in the musculoskeletal system (eg, shoulder or ankle).

Next Steps to Establish the Validity of Milestones

Orthopaedic programs graduated the first class of residents with the full set of PGY-1 to PGY-5 Milestone data in June 2018. Efforts are underway to evaluate the national specialty data with respect to overall performance, performance at the level of each subcompetency, and programmatic traits that may influence the data. Subcompetencies with

Figure 1



Excerpt from the 2016 Milestones Annual Report. This figure shows the wide range of attainment of milestones by subcompetency in orthopaedic surgery. Gaps in curriculum and widespread availability of effective assessment tools are thought to be notable contributors to the observed variation in these early findings, although true variation in resident competence cannot be ruled out. These early data represent useful information for making improvements in graduate medical education, in accordance with Deming's proposition for the value of variance in CQI.[14] Data are shown for only the first 12 patient care subcompetencies. Copyright 2016 Accreditation Council for Graduate Medical Education. Used with permission. See https://www.acgme.org/What-We-Do/Accreditation/Milestones/Resources. CQI = continuous quality improvement

high variability may represent poor design or poor subject choices. The data will provide the first look at whether each milestone represents a progression of skills, knowledge, and behaviors based on the variance seen by program, by graduate year, and within the program by looking at each individual resident.

Milestones 2.0, which represents a major revision of design and content for all specialties, is currently being phased in across all specialties, with work in orthopaedic surgery beginning in the spring of 2019. A working group is now assembled of PDs and education leaders within orthopedics. A group of PDs is currently analyzing the 5-year data, and this information will be available to the taskforce. Future orthopaedic milestones should be well aligned with other specialties because many of these competencies cross the medical specialties. PC and MK both need substantial revising because orthopaedics initially created a broader and more detailed assessment of topics than other specialties. An equally important task will be to redesign the other core competencies (PROF, interpersonal and communication skill, SBP, PBLI) so they fit within the context of orthopaedic training and practice. A draft of revisions in these areas has been completed and will be available to the working group for orthopaedic surgery during the revision process. These efforts essentially address two aspects of validity: content and relations with other variables.[28,29]

Other aspects of validity will be important to assess. There are two current assessments in orthopaedic residencies that offer opportunities to compare with milestones data—the Orthopaedic In-training Examination (American Academy of Orthopaedic Surgeons) and the American Board of Orthopaedic Surgery certification examinations. Both are monitored for performance, and validity research should be conducted to examine correlations of these with milestone ratings. It will be important to explore correlations with independently obtained measurements of clinical performance such as patient outcomes residing in clinical registry databases and assessment of both surgical and clinical skills.

Response process validity can be investigated by examining what goes into milestones ratings.[30] For example, if faculty raters or CCCs are overwhelmed by the process of generating milestones data for reporting purposes, there may be a tendency to

rush through the process. This may create tendencies for bias[31,32] or overgeneralization, which might show up as unwarranted stringency, leniency (halo effect), or straight-lining, where residents are assigned the same milestone rating for all 41 subcompetencies. Straight-lining exists in most specialties,[33] but it is generally at a lower rate in orthopaedic surgery, ranging from 6% to 13% (versus over 30% in some specialties) across PGY levels 1 to 5 in the latest milestones data set from 2017 to 2018.[34] This can be detected and, if problematic, can be addressed by making the rating process easier and simpler. Efforts to create simpler assessment tools using hand-held devices are one way to improve response process validity.[35,36] Simple and effective methods for faculty development can also help.[37]

Future Directions

The goal with Milestones 2.0 is to adjust the structure and/or content as necessary based on 5 years of experience and the data analysis. In our opinion, the resulting data from evaluation of Milestones 1.0 and the acknowledgement of the issues that these milestones have created will lead to a need for a substantial change in Orthopaedic Milestones 2.0, and not just a small shift and improvement of language. The original design focused on key disease entities; the redesign may be better served with a more general approach. Many PDs believe the number should be dramatically decreased and reassembled around broad areas of competency rather than orthopaedic subspecialties. Examples of technical skill milestones might be open major procedures, open minor procedures, and arthroscopic procedures. Additional PC milestones might be surgical indications and nonsurgical care. Medical knowledge might consist of fewer milestones based on region or patient age.

These are just examples and the actual direction will need to be decided by the community, including CORD and the recently formed Milestones 2.0 workgroup. A survey of orthopaedic PDs sent in early 2019 was designed to obtain feedback on the current milestones and the process of collecting milestone data. The other critical requirement of the workgroup will be to communicate the findings from the analysis of the first 5 years of data with PDs; these will also be available in 2019. The workgroup will also be key in communicating the changes recommended in Milestones 2.0. A role for CORD could be to identify best practices among programs for CCC processes or faculty support for collecting milestones data—for example, how to best handle data in programs of different sizes, and then to share these ideas with the community at large.

Summary

Orthopaedic Milestones 1.0 has been an important step forward by providing a uniform framework for tracking residents' development of competence that can be consistently applied across all programs in the country. Soon we will have national data on how the various milestones have performed. These data will be published and shared. It is now time to take a serious look at a substantial revision that will correlate with other initiatives in resident assessment in orthopaedic surgery. The future holds promise for markedly improved resident assessment that will improve resident knowledge of their progress, identify gaps in programs, better assure competency, and help to develop a more uniform curriculum of orthopaedic surgery training.

Acknowledgments

The authors thank the following colleagues for their helpful comments and discussion in drafting this article: Eugene Jang, Eric Holmboe, Laura Edgar, John Potts, Tom Nasca.

References

References printed in **bold type** are those published within the past 5 years.

1. Nasca TJ, Philibert I, Brigham T, Flynn TC: The next GME accreditation system—rationale and benefits. *N Engl J Med* 2012; 366:1051-1056.
2. **Van Heest AE, Dougherty PJ: CORR® curriculum—orthopaedic education: Operative assessment and the ACGME milestones: Time for change. *Clin Orthop Relat Res* 2015;473:775-778.**
3. **Holmboe ES: The journey to competency-based medical education—implementing milestones, *Marshall J Med* 2017;3(1):2.**
4. **Goldman RH, Tuomala RE, Bengtson JM, Stagg AR: How effective are new milestones assessments at demonstrating resident growth? 1 year of data. *J Surg Educ* 2017; 74:68-73.**
5. Stern PJ, Albanese S, Bostrom M, et al: Orthopaedic surgery milestones. *J Grad Med Educ* 2013;5(1 suppl 1):36-58.
6. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S: General competencies and accreditation in graduate medical education. *Health Aff (Millwood)* 2002;21: 103-111.
7. Swing SR, Beeson MS, Carraccio C, et al: Educational milestone development in the first 7 specialties to enter the next accreditation system. *J Grad Med Educ* 2013;5:98-106.
8. **Conforti LN, Yaghmour NA, Hamstra SJ, et al: The effect and use of milestones in the assessment of neurological surgery residents and residency programs. *J Surg Educ* 2018;75:147-155.**
9. **Edgar L, Roberts S, Yaghmour NA, et al: Competency crosswalk: A multispecialty review of the Accreditation Council for Graduate Medical Education milestones across four competency domains. *Acad Med* 2018;93:1035-1041.**
10. **Holmboe ES: Realizing the promise of competency-based medical education. *Acad Med* 2015;90:411-413.**
11. **Hauer KE, Clauser J, Lipner RS, et al: The internal medicine reporting milestones:**

Cross-sectional description of initial implementation in U.S. residency programs. *Ann Intern Med* 2016;165:356-362.

12. **Holmboe ES, Call S, Ficalora RD: Viewpoint: Milestones and competency-based medical education in internal medicine. *JAMA Intern Med* 2016;176:1601-1602.**

13. **Holmboe ES, Yamazaki K, Edgar L, et al: Reflections on the first 2 years of milestone implementation. *J Grad Med Educ* 2015;7: 506-511.**

14. Deming WE: *The New Economics for Industry, Government, Education*, ed 2. Cambridge, MA, MIT Press, 2000.

15. **Holmboe ES, Yamazaki K, Nasca TJ, Hamstra SJ. Using longitudinal milestones data and learning analytics to facilitate the professional development of residents: early lessons from three specialties. *Acad Med* 2019; July 23 [Epub ahead of print].**

16. Gauger PG, Gruppen LD, Minter RM, Colletti LM, Stern DT: Initial use of a novel instrument to measure professionalism in surgical residents. *Am J Surg* 2005;189:479-487.

17. **Irby DM, Hamstra SJ: Parting the clouds: Three professionalism frameworks in medical education. *Acad Med* 2016;91: 1606-1611.**

18. Hodges BD, Ginsburg S, Cruess R, et al: Assessment of professionalism: Recommendations from the Ottawa 2010 conference. *Med Teach* 2011;33:354-363.

19. **Moreau KA, Eady K, Tang K, et al: The development of the PARENTS: A tool for parents to assess residents' non-technical skills in pediatric emergency departments. *BMC Med Educ* 2017;17:210.**

20. **Moreau KA, Eady K, Frank JR, et al: A qualitative exploration of which resident skills parents in pediatric emergency departments can assess. *Med Teach* 2016; 38:1118-1124.**

21. **Lagoo J, Berry WR, Miller K, et al: Multisource evaluation of surgeon behavior is associated with malpractice claims. *Ann Surg* 2019;270:84-90.**

22. Johnston DL, Rowan-Legg A, Hamstra SJ: Examining the educational value of a CanMEDS roles framework in pediatric morbidity and mortality rounds. *BMC Med Educ* 2014;14:262.

23. **Holmboe ES, Batalden P: Achieving the desired transformation: Thoughts on next steps for outcomes-based medical education. *Acad Med* 2015;90:1215-1223.**

24. Warm EJ, Mathis BR, Held JD, et al: Entrustment and mapping of observable practice activities for resident assessment. *J Gen Intern Med* 2014;29:1177-1182.

25. **Warm EJ, Held JD, Hellmann M, et al: Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med* 2016;91:1398-1405.**

26. Hamstra SJ, Edgar L, Yamazaki K, Holmboe ES: Milestones Annual Report 2016. Accreditation Council for Graduate Medical Education. https://www.acgme.org/What-We-Do/Accreditation/Milestones/Resources. Accessed January 14, 2019.

27. **Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ: Entrustability scales: Outlining their usefulness for competency-based clinical assessment. *Acad Med* 2016;91: 186-190.**

28. Cook DA, Beckman TJ: Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med* 2006;119:166.e7-166.e16.

29. Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education: *Standards for Educational and Psychological Testing*, Washington, DC, American Educational Research Association, 2014.

30. **Hamstra SJ, Yamazaki K, Barton MA, Santen SA, Beeson MS, Holmboe ES: A national study of longitudinal consistency in ACGME milestone ratings by clinical competency committees: Exploring an aspect of validity in the assessment of residents' competence. *Acad Med* 2019;94: 1522-1531.**

31. **Dayal A, O'Connor DM, Qadri U, Arora VM: Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med* 2017;177:651-657.**

32. **Santen SA, Yamazaki K, Holmboe ES, Yarris LM, Hamstra SJ: Comparison of male and female resident milestone assessments during emergency medicine residency training: A national study. *Acad Med* 2019; September 10 [Epub ahead of print].**

33. **Beeson MS, Hamstra SJ, Barton MA, et al: Straight line scoring by clinical competency committees using emergency medicine milestones. *J Grad Med Educ* 2017;9: 716-720.**

34. Hamstra SJ, Yamazaki K, Edgar L, Sangha S, Holmboe ES: Milestones National Report 2018. Accreditation Council for Graduate Medical Education. https://www.acgme.org/What-We-Do/Accreditation/Milestones/Resources, p.24. Accessed January 14, 2019.

35. George BC, Teitelbaum EN, Meyerson SL, et al: Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ* 2014;71:e90-e96.

36. **Connolly A, Goepfert A, Blanchard A, et al: myTIPreport and training for independent practice: A tool for real-time workplace feedback for milestones and procedural skills. *J Grad Med Educ* 2018;10:70-77.**

37. 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.