Assessment of Learning, for Learning: Operationalizing Milestones Data for Program-Level Improvement

Janae K. Heath, MD, MSCE
Laura Edgar, EdD, CAE
Susan Guralnick, MD, FAAP

Assessment of learning, assessment for learning. This phrase is frequently applied in medical education circles describing assessment strategies for students, residents, and fellows. However, this paradigm can extend beyond learner assessment; in fact, a similar philosophy can be implemented for program-level improvement. In residency and fellowship programs accredited through the Accreditation Council for Graduate Medical Education (ACGME), there is a broad array of assessment data collated and reviewed by the program's clinical competency committee (CCC).[1] Based on these data, CCC discussions lead to assignment of appropriate Milestone levels to their program's trainees and help guide individualized learning plans. Applying such Milestones data in program evaluation can drive important evolution and improvement of training programs.[2]

Milestones data are summarized in the ACGME Accreditation Data System (ADS). ADS provides graphic representations of Milestones trends for each training level for individuals (FIGURE 1A), programs (FIGURE 1B), and nationally (FIGURE 1C). Additionally, ADS highlights historical and current data for each program, showcasing progression of Milestones over time, variation among learners within a single year, and trends across Milestone subcompetencies within a program.

These Milestones data can suggest natural areas for program change yet challenges with data access and interpretation may limit the use of this resource. In fact, according to ACGME Milestones development staff (L. Edgar, oral communication, October 2020), a convenience sampling of more than 15 Milestones specialty development groups demonstrates that, on average, only 7 of 12 members are aware these data are available in ADS. In order to effectively operationalize these data, it is first critical that program leaders understand the scope of the available data and know how to apply the information toward program review.

ADS provides useful visual displays of the Milestones for program-level and nationwide data, using boxplots for each postgraduate year (PGY) of training. These boxplots provide relevant summary statistics for Milestones, including the mean, median, 25th, and 75th quartiles, maximum and minimum values, and outliers for each subcompetency. The national Milestones data also include novel predictive analytics (using predictive probability values [PPVs]), which provide the probability that trainees will not achieve a target Milestone level by the time of graduation along each domain (provided as online supplementary data). These predictive metrics use early Milestone ratings to identify the percentage of learners who do not achieve goals during training, based on data from previously collected national Milestone ratings.[1,3,4] This metric, as well as visual representation of the trajectory and pattern of Milestone-level advancement, can highlight issues with current curricula, assessment methods, Milestone-level assignment, or faculty development needs. This can be performed using program-level box plot data (FIGURE 1B) or by creating PGY-level Milestones charts for each subcompetency (FIGURE 2).

Finally, a newer method available for assessment is the predictive analytics provided in the PPV tables. These PPV tables (provided as online supplementary data), originally intended to identify struggling learners at risk of not achieving a Milestone Level 4 at the time of graduation, may also be used to identify programmatic issues, particularly if several learners in the program have a high probability of not achieving the designated graduation Milestone target. Brainstorming ways to operationalize these data for

DOI: http://dx.doi.org/10.4300/JGME-D-20-00849.1

Editor's Note: The online version of this article contains a sample table of predictive probability values (PPV), which provide estimates that Milestone ratings will fall below Level 4 at the time of graduation; a sample Milestone distribution (by subcompetency and post-graduate year) in multi-year programs; sample Milestone distribution for one-year programs; and approach to analyzing these data.

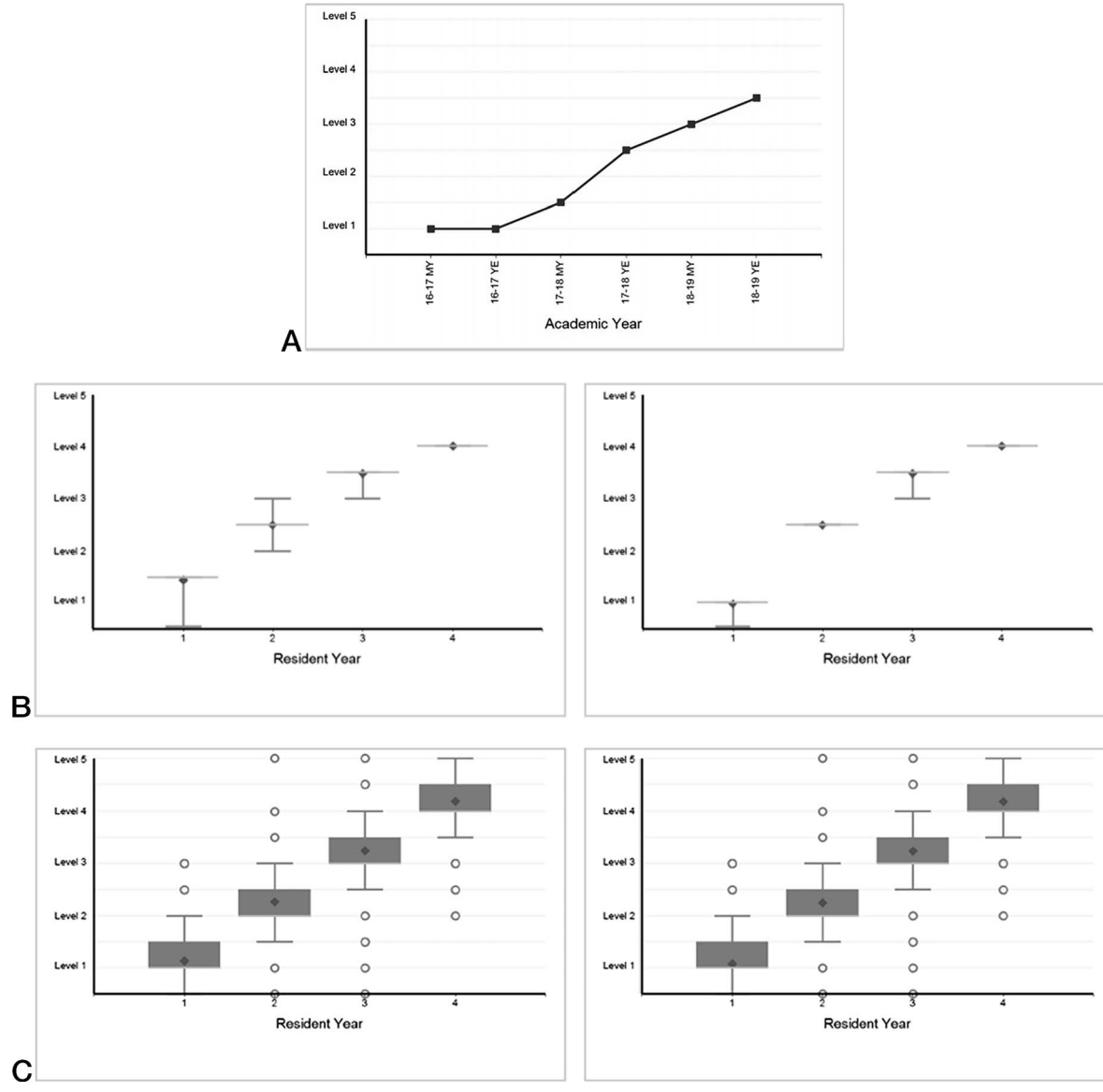

FIGURE 1
Data Examples from the ACGME Accreditation Data System (ADS)

Note: The above figure highlights data representation from ADS, which can be viewed on an individual learner level (1A), program-level (1B), or nationally (1C) for each of the subcompetencies.

iterative review provides the program evaluation committee with a broad array of opportunities for program improvement.

Overall program information can be easily gleaned from review of the ADS database or simple manual extraction for interpretation (see online supplementary data for sample charts and instructions). During the required annual program evaluation, these data can be used to identify strengths, gaps, and areas for improvement. Logistically, however, how can programs approach this extensive data to improve their local processes? Through use of summary statistics, program trends, or predictive analytics, the opportunities for application of Milestones data are vast. Below, we aim to highlight 3 novel future areas to use Milestones data in conjunction with annual program reviews, specifically in the areas of curricular innovation, bias mitigation, and individualization of training pathways.

Using Milestones to Iteratively Drive Curricular Change

An important application to consider in the use of Milestones to achieve program-level change revolves around iterative curricular change. On an individual

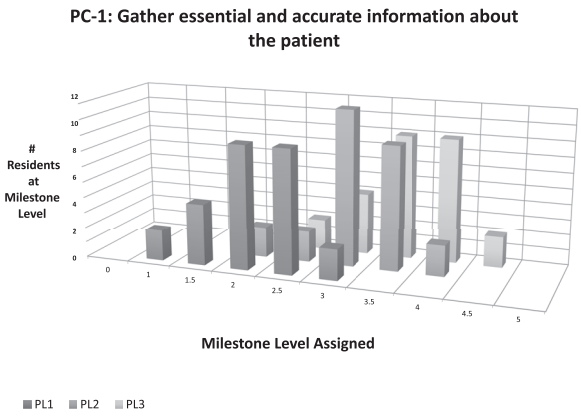

FIGURE 2
Trajectory and Pattern of Program Milestone Level Advancement by Subcompetency

trainee level, failure to progress as expected could signify issues with the curriculum. The absence of expected variation by PGY level and within subcompetencies also raises concern for a problem, either with the curricula, assessment tools, or faculty raters. For example, if a program notes a lack of upward progression of Milestone ratings for systems-based practice 2 ("Coordinate patient care within the health system relevant to their clinical specialty"), a program can implement curricular interventions for patient care coordination, train raters to achieve a shared mental model of expertise, and review/create assessments for such care coordination. Using a quality improvement cycle, a program can use Milestone ratings to determine targets, track progress, and determine success in interventions.

Using Milestones for Program-Level Bias Review

As outlined above, the trajectory and pattern of Milestone-level advancement can highlight issues with current curricula, assessment methods, Milestone-level assignment, or faculty development. Beyond this, it is increasingly recognized that underlying biases impact assessment across graduate medical education (GME), and several studies have highlighted differential Milestone achievement and trajectory based on trainee gender.[5–8] A thorough review of the Milestone trends within a program in conjunction with demographic patterns could promote further understanding of potential underlying program-level biases. The lack of an expected upward trajectory, or alternatively a differential slope between demographic factors, could highlight a need for additional faculty development and program interventions on potential underlying biases affecting assessments. Faculty development using bias reduction workshops, shown in other settings to be effective mitigation strategies of biased decision-making, could be implemented by programs to potentially address these issues.[9–11]

Beyond trends noted within an individual program, one could additionally compare subgroups within their program to national specialty ACGME data. While this may reveal additional program-level disparities prompting intervention, it is important to note that there are caveats when comparing program-level to national specialty ACGME data. The variance between individual program data compared to national data could indicate a need for program adjustment—or could be within expected limits of variation. Regardless, this could highlight a need for added interventions if demographic differences in Milestone trends are observed.

Using Milestones to Create Program Pathway Individualization

Milestone assessments are routinely used for targeted individual learner development. However, beyond individual growth, program Milestones data provide a unique opportunity for novel individualization of training pathways. For example, consistent achievement of early achievement of "readiness for independent practice" could prompt programs to reassess curricula for individuals who have achieved expertise. If data review reveals trainees consistently achieve "ready for independent practice" across certain subcompetencies prior to the final year of training, a program can consider implementing novel roles and rotations to further progress expertise in advanced areas.[12–14]

Review of a program's longitudinal Milestones could also aid in the shift from a dwell-based view of training to a time-variable model,[15] providing opportunities for adjusted training pathways based on achievement of competency. While certainly logistical challenges preclude widespread adoption of this strategy beyond select programs in the transition from undergraduate medical education to GME, this could serve as an area of interesting future scholarship in the use of program-level Milestones data.

Conclusions

The ACGME Milestones serve as an essential component of program evaluation, identifying curricular needs, assessment issues, or faculty development needs. Analyzing these data to identify trends within a program, between programs within an institution, and between institutions, is a critical way to use the available Milestones data for ongoing program improvement. With this knowledge, programs can

begin to operationalize these data in novel ways for program assessment.

References

1. Accreditation Council for Graduate Medical Education. Clinical Competency Committees: A Guidebook for Programs. http://www.acgme.org/portals/0/acgmeclinicalcompetencycommitteeguidebook.pdf. Accessed March 2, 2021.
2. Caretta-Weyer HA, Gisondi MA. Design your clinical workplace to facilitate competency-based education. *West J Emerg Med*. 2019;20(4):651–653. doi:10.5811/westjem.2019.4.43216.
3. Accreditation Council for Graduate Medical Education. Milestones National Report 2019. https://www.acgme.org/Portals/0/PDFs/Milestones/2019MilestonesNationalReportFinal.pdf?ver=2019-09-30-110837-587. Accessed March 2, 2021.
4. 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*. 2020;95(1):97–103. doi:10.1097/ACM.0000000000002899.
5. 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(5):651–657. doi:10.1001/jamainternmed.2016.9616.
6. 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*. 2020;95(2):263–268. doi:10.1097/ACM.0000000000002988.
7. Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender differences in attending physicians' feedback to residents: a qualitative analysis. *J Grad Med Educ*. 2017;9(5):577–585. doi:10.4300/JGME-D-17-00126.1.
8. Klein R, Ufere NN, Rao SR, et al. Association of gender with learner assessment in graduate medical education. *JAMA Netw Open*. 2020;3(7):e2010888. doi:10.1001/jamanetworkopen.2020.10888.
9. Capers Q. How clinicians and educators can mitigate implicit bias in patient care and candidate selection in medical education. *ATS Scholar*. 2020;1(3)211–217. https://doi.org/10.34197/ats-scholar.2020-0024PS.
10. Boscardin CK. Reducing implicit bias through curricular interventions. *J Gen Intern Med*. 2015;30(12):1726–1728. doi:10.1007/s11606-015-3496-y.
11. Milkman KL, Chugh D, Bazerman MH. How can decision making be improved? *Perspect Psychol Sci*. 2009;4(4):379–383. doi:10.1111/j.1745-6924.2009.01142.x.
12. Emanuel EJ, Fuchs VR. Shortening medical training by 30%. *JAMA*. 2012;307(11):1143–1144. doi:10.1001/jama.2012.292.
13. Nyquist JG. Educating physicians: a call for reform of medical school and residency. *J Chiropr Educ*. 2011;25(2):193–195.
14. Aschenbrener CA, Ast C, Kirch DG. Graduate medical education: its role in achieving a true medical education continuum. *Acad Med*. 2015;90(9):1203–1209. doi:10.1097/ACM.0000000000000829.
15. Powell DE, Carraccio C, Aschenbrener CA. Pediatrics redesign project: a pilot implementing competency-based education across the continuum. *Acad Med*. 2011;86(11):e13. doi:10.1097/ACM.0b013e318232d482.

Janae K. Heath, MD, MSCE, is Assistant Professor of Medicine, Division of Pulmonary and Critical Care, University of Pennsylvania; **Laura Edgar, EdD, CAE,** is Vice President, Milestones Development, Accreditation Council for Graduate Medical Education; and **Susan Guralnick, MD, FAAP,** is Associate Dean for Graduate Medical Education, Designated Institutional Official, and Professor of Pediatrics, University of California, Davis.

Corresponding author: Janae K. Heath, MD, MSCE, Hospital of the University of Pennsylvania, janae.heath@pennmedicine.upenn.edu, Twitter @JanaeHeath1