The Relationship Between ACGME Duty Hour Requirements and Performance on the American Board of Emergency Medicine Qualifying Examination

Francis L. Counselman, MD
Terry Kowalenko, MD
Catherine A. Marco, MD
Kevin B. Joldersma, PhD
Robert C. Korte, PhD
Earl J. Reisdorff, MD

ABSTRACT

Background In 2003, the Accreditation Council for Graduate Medical Education (ACGME) instituted requirements that limited the number of hours residents could spend on duty, and in 2011, it revised these requirements.

Objective This study explored whether the implementation of the 2003 and 2011 duty hour limits was associated with a change in emergency medicine residents' performance on the American Board of Emergency Medicine (ABEM) Qualifying Examination (QE).

Methods Beginning with the 1999 QE and ending with the 2014 QE, candidates for whom all training occurred without duty hour requirements (Group A), candidates under the first set of duty hour requirements (Group C), and candidates under the second set of duty hour requirements (Group E) were compared. Comparisons included mean scores and pass rates.

Results In Group A, 5690 candidates completed the examination, with a mean score of 82.8 and a 90.2% pass rate. In Group C, 8333 candidates had a mean score of 82.4 and a 90.5% pass rate. In Group E, there were 1269 candidates, with a mean score of 82.5 and an 89.4% pass rate. There was a small but statistically significant decrease in the mean scores (0.04, $P < .001$) after implementation of the first duty hour requirements, but this difference did not occur after implementation of the 2011 standards. There was no difference among pass rates for any of the study groups ($\chi^2 = 1.68$, $P = .43$).

Conclusions We did not identify an association between the 2003 and 2011 ACGME duty hour requirements and performance of test takers on the ABEM QE.

Introduction

On July 1, 2003, the Accreditation Council for Graduate Medical Education (ACGME) instituted new requirements that limited resident duty hours in all ACGME-accredited residency programs in the United States (TABLE 1).

Proponents of these requirements stated they would enhance patient safety and improve the working conditions and education of resident physicians.[1,2] Since their implementation, there have been many reports on the effects of these limits on patient care and resident education.[2–7] A systematic review of the effects of the 2003 duty hour requirement on patient safety, resident well-being, and resident education found no impact on patient care or on resident well-being.[8] The authors did find an unintended negative impact on resident education, including less time spent with attending physicians and decreased attendance at teaching sessions.[8] There have been several studies examining the effect of the ACGME duty hour requirements on board certification examination performances.[4,9–11] The findings have been inconsistent, with improved performance in some specialties, decline in others, and a third group that showed no impact. The relationship between the ACGME duty hour requirements and performance on the American Board of Emergency Medicine (ABEM) Qualifying Examination (QE) has not been examined.

On July 1, 2011, additional ACGME revised duty hour requirements were implemented for all US residency programs (TABLE 1).[12]

Since the implementation of the 2011 ACGME duty hour requirements, there have been limited data published regarding their impact on educational or patient safety outcomes.

Emergency medicine (EM) established specialty-specific duty hour requirements in 1990, which were revised in 1995 and again in 2003.[13] However, these duty hour restrictions only applied to EM rotations, not off-service rotations. Under the current ACGME EM program requirements, up to a maximum of 40%

DOI: http://dx.doi.org/10.4300/JGME-D-15-00591.1

Editor's Note: The online version of this article contains a table of definitions of each group based on the year in which the qualifying examination was administered.

of an EM resident's training may take place off-service (ie, ob-gyn, medical intensive care, pediatric intensive care, etc). It is during these off-service rotations that the 2003 and 2011 ACGME duty hour requirements would have the greatest impact on EM resident education.

This study sought to identify any relationship between ABEM QE performance by graduates of ACGME-accredited EM residency programs and the 2003 and 2011 ACGME duty hour requirements.

What was known and gap
Studies have found varying impacts of duty hour limits on graduates' performance on board certification examinations, a key educational outcomes measure.

What is new
A study explored whether emergency medicine graduates' performance on the American Board of Emergency Medicine qualifying examination changed after implementation of the 2003 and 2011 duty hour standards.

Limitations
This was a single specialty study; other factors may have affected examination performance.

Bottom line
Neither set of duty hour standards resulted in a practical significant change in examination performance.

Methods

This retrospective study used performance data for first-time candidates of the ABEM QE administered from 1999 through 2014. Data were accessible to the investigators only in deidentified, aggregate reports. A total of 5 groups were identified, only 3 of which were studied. Group A were test takers for whom their entire training preceded ACGME duty hour requirements; this involved QE results from 1999 to 2003 (provided as online supplemental material). Group C were test takers who had all of their training under the first set of duty hour requirements: QE results from 2006 to 2011 for EM 1–3 program graduates and 2007 to 2011 for EM 1–4 program graduates. Group E were test takers from EM 1–3 programs who had all of their training under the second set of duty hour requirements (2014). Groups B and D were not studied because of the variable amount of experience with periods of different duty hour standards.

Since the first administration, the ABEM has used a strictly standardized approach in QE test design, item (question) development, standard setting (the establishment of a passing score), administration, and scoring using a criterion reference. Beginning in 2004, the ABEM began to equate the examinations, which stabilizes most differences in year-to-year difficulty in the examination. Prior analysis has shown that the ABEM QE was a psychometrically stable examination prior to 2004, and has remained stable.[14]

Eligible participants included all graduates of ACGME-accredited EM residency programs taking

TABLE 1
2003 and 2011 Accreditation Council for Graduate Medical Education Duty Hour Requirements

Requirement	2003	2011
Maximum hours of work per week	80 h averaged over 4 wk[a] (Individual programs can request an additional 10% increase in the work week if based on a sound educational rationale)[a]	80 h averaged over 4 wk[a] (Individual programs can request an additional 10% increase in the work week if based on a sound educational rationale)[a]
Maximum duty period length	30 h (Can admit patients up to 24 h, then 6 additional hours for transitional and educational activities)	PGY-1s limited to 16 h PGY-2s and above: 28 h (Can admit patients for up to 24 h, plus 4 h for transitional and educational activities)
Maximum in-hospital on-call frequency	Every third night, averaged over a 4-wk period	Every third night, averaged over a 4-wk period
Minimum time off between scheduled duty periods	10 h	All residents should have 10 h; must have 8 h PGY-2s and above: must have 14 h off after 24 h of in-house duty
Maximum frequency of in-hospital night shift/float	Not addressed	6 consecutive nights
Mandatory time off duty	4 d off per month 1 d (24 h) off per week, averaged over 4 wk	4 d off per month 1 d (24 h) off per week, averaged over 4 wk

Abbreviation: PGY, postgraduate year.
[a] Adapted from Desselle B, Dawkins R. Faculty development: overview of the new ACGME duty hour regulations. http://www.medschool.lsuhsc.edu/pediatrics/docs/FD%20-%20Overview%20of%20the%20New%20ACGME%20Duty%20Hour%20Regulations.pdf. Accessed August 24, 2016.

the ABEM QE for the first time from 1999 through 2014. Repeat test takers were excluded to avoid potential practice effect. Only physicians who graduated from programs formatted as postgraduate years 1–3 (PGY-1 to PGY-3) and PGY-1 to PGY-4 were included, because these program types were consistently represented throughout the years studied. Primary outcome measures included the mean QE scores with 95% confidence interval (CI) and passing rates.

This study was granted a waiver for human subject research by the Eastern Virginia Medical School Institutional Review Board.

We performed 1-way analysis of variance (ANOVA) to compare mean scores. To further define any statistical difference among the study groups, Tukey's studentized range test was performed. For passing rates, a 2×3 chi square test was used. A pre hoc level of significance was determined to be $\propto < .01$ for all analyses.

Analyses were performed using R version 3.1.2 (The R Foundation).[15] Specifically, the R packages plyr, reshape, and psych were used in data preparation to transpose the data from one statistical format to another; data analyses were performed using base R and psych.[16–18]

Results

For the period between 1999 and 2014 there were 26 753 test takers. Groups A through E consisted of 20 189 (75.5%) first-time test takers. For the groups studied, the numbers of first-time test takers were: Group A (5690), Group C (8333), and Group E (1269).

Group A had a mean score of 82.8 (95% CI 82.7–83.0), with a 90.2% pass rate; Group C had a mean score of 82.4 (95% CI 82.2–82.5), with a 90.5% pass rate; and Group E had a mean score of 82.5 (95% CI 82.1–82.8), with an 89.4% pass rate.

One-way ANOVA comparing mean scores among the 3 groups was significant ($P \leq .001$). Tukey's studentized ranged test demonstrated a statistically significant difference between groups A and C ($P < .001$). The Tukey comparisons for groups A and E ($P = .10$) and for groups C and E ($P = .83$) were not significantly different. A 2×3 chi square test demonstrated no statistically significant difference in pass rates ($P = .43$).

Discussion

Our study demonstrated no significant difference in pass rates on the ABEM QE for first-time test takers before and after implementation of the 2003 ACGME study hour standards, despite a small but statistically significant difference in mean scores (–0.04) for the pre-2003 and post-2003 cohorts. While statistically significant, there was no practical difference in the context of a passing rate or examination score, since the difference was less than half a point on a 0 to 100 scale. There was no statistically significant difference in pass rates and mean scores on the ABEM QE for the cohorts before and after the 2011 ACGME duty hour requirements. However, the complete impact of the 2011 ACGME duty hour requirements may be unknown until the 2015 QE, the first examination for graduates fully trained under these duty hour requirements in a PGY-1 to PGY-4 program format.

The effect of the ACGME duty hour requirements on board certification performance has been variable across specialties. On the American Board of Internal Medicine board certification examination, there was no significant difference in scores before and after implementation of the 2003 ACGME duty hour standards.[9] This is in contrast to the improving trends on the American Board of Orthopaedic Surgery, Part I (written) Certification Examination.[11] Candidates taking the American Board of Urology QE demonstrated similar improvement,[11] while performance on the American Board of Obstetrics and Gynecology written examination had a downward trend in the pass rate.[11]

Many factors may affect performance on board certification examinations, including the residency curriculum, clinical experience, innate ability, self-study, residency size, and individual motivation. All of the studies discussed above, including our own, were designed to look at associations and could not demonstrate a cause and effect relationship between the implementation of the ACGME duty hour standards and performance on board certification examinations.

This study only examined the results of the ABEM QE. The ABEM QE is a 305 single-best-answer, multiple-choice question examination designed to measure medical knowledge recall (approximately 33% of all items) and clinical synthesis and diagnostic reasoning (approximately 66% of all items).[19,20] Between 10% and 15% of items include pictorial stimuli. The ABEM QE is a criterion-referenced examination; it does not use a performance curve or quota for passing or failing scores. The passing score is determined by ABEM directors who are informed by a modified Angoff standardized setting process. We did not examine performance on the ABEM Oral Certifying Examination in our study because not all boards require an oral examination. One potential reason EM did not see a significant change in performance on the ABEM QE could be specialty-specific duty hour limits that for the specialty predate the ACGME common standards (TABLE 2). Compared with other specialties,

TABLE 2
Self-Imposed, Specialty-Specific Duty Hour Restrictions by the Residency Review Committee for Emergency Medicine (RRC-EM)[a]

Year	RRC-EM Requirement
1990	EM residents rotating in the ED must have an average of 1 day in 7 free from all clinical and educational requirements.
1995	EM residents working in the ED have a maximum 60-h clinical work week.
2003	EM residents working in the ED cannot work more than 72 h/wk, including all didactics and other educational activities.

Abbreviation: ED, emergency department.
[a] From: Wagner MJ, Wolf S, Promes S, et al. Duty hours in emergency medicine: balancing patient safety, resident wellness, and the resident training experience: a consensus response to the 2008 Institute of Medicine resident duty hours recommendations. *Acad Emerg Med*. 2010;17(9):1004–1011.

EM training programs have had a longer experience under duty hour limits, and the most significant impact of the 2003 and 2011 ACGME standards for EM occurred in the off-service rotations. Finally, there could be important effects resulting from the duty hour requirements that are not captured on a multiple-choice question examination.

There are several limitations to this study. First, an analysis of means must be interpreted cautiously because the QE results were not equated prior to 2004. Equating is a psychometric process that accounts for any year-to-year variation in the intrinsic difficulty of an examination. The size of the study cohorts, the adherence to best practices for examination development, and the general physician performance trends over decades of test administration should offer some assurance that the use of non-equated examinations would not have greatly affected study results. Second, the cohort of first-time candidates following the 2011 ACGME duty hour requirements in our study includes only graduates of EM 1–3 programs, not graduates of 4-year programs. A post hoc internal analysis of EM 1–3 program graduates in groups A, C, and E demonstrated similar results. Third, the ABEM QE changed from paper and pencil to an electronic format in 2004. Also, it is not known to what degree individual programs complied with the ACGME duty hour requirements. Finally, other variables may have affected QE scores, such as the number of EM programs, the number of EM residents, changes to the Model of the Clinical Practice of Emergency Medicine, changes in residency program leadership, or other factors related to residency training.

Conclusion

In conclusion, we did not identify an association between the 2003 and 2011 ACGME duty hour requirements and the performance of first-time test takers on the ABEM QE.

References

1. Philibert I, Friedmann P, Williams WT, et al. New requirements for resident duty hours. *JAMA*. 2002;288(9):1112–1114.
2. Fletcher KE, Reed DA, Arora VM. Patient safety, resident education and resident well-being following implementation of the 2003 ACGME duty hour rules. *J Gen Intern Med*. 2011;26(8):907–919.
3. Shonka DC Jr, Ghanem TA, Hubbard MA, et al. Four years of Accreditation Council of Graduate Medical Education duty hour regulations: have they made a difference? *Laryngoscope*. 2009;119(4):635–639.
4. Sneider EB, Larknin AC, Shah SA. Has the 80-hour workweek improved surgical resident education in New England? *J Surg*. 2009;66(3):140–145.
5. Jagannathan J, Vates GE, Pouratian N, et al. Impact of the Accreditation Council for Graduate Medical Education work-hour regulations on neurosurgical resident education and productivity. *J Neurosurg*. 2009;110(5):820–827.
6. Fletcher KE, Underwood W 3rd, David SQ, et al. Effects of work hour reduction on residents' lives: a systematic review. *JAMA*. 2005;294(9):1088–1100.
7. Babu R, Thomas S, Hazzard MA, et al. Worse outcomes for patients undergoing brain tumor and cerebrovascular procedures following the ACGME duty-hour restrictions. *J Neurosurg*. 2014;121(2):262–276.
8. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ*. 2015;7(3):349–363.
9. Silber JH, Romano PS, Itani KM, et al. Assessing the effects of the 2003 resident duty hours reform on internal medicine board scores. *Acad Med*. 2014;89(4):644–651.
10. Falcone JL, Hamad GG. The American Board of Surgery Certifying Examination: a retrospective study of the decreasing pass rates and performance for first-time examinees. *J Surg*. 2012;69(2):231–235.

11. Falcone JL, Feinn RS. The ACGME duty hour standards and board certification examination performance trends in surgical specialties. *J Grad Med Educ*. 2013;5(3):446–457.
12. Romano PS, Volpp K. The ACGME's 2011 changes to residency duty hours: are they an unfunded mandate on teaching hospitals? *J Gen Intern Med*. 2012;27(2):136–138.
13. Wagner MJ, Wolf S, Promes S, et al. Duty hours in emergency medicine: balancing patient safety, resident wellness, and the resident training experience: a consensus response to the 2008 Institute of Medicine resident duty hours recommendations. *Acad Emerg Med*. 2010;17(9):1004–1011.
14. Marco CA, Counselman FL, Korte RC, et al. Emergency physicians maintain performance on the American Board of Emergency Medicine Continuous Certification (ConCert) examination. *Acad Emerg Med*. 2014;21(5):532–537.
15. R Core Team. R: A language and environment for statistical computing. The R Project for Statistical Computing; 2015. http://www.R-project.org. Accessed June 24, 2016.
16. Wickham H. The split-apply-combine strategy for data analysis. *J Statistical Software*. 2011;40(1):1–29. http://www.jstatsoft.org/v40/i01. Accessed June 24, 2016.
17. Wickham H. Reshaping data with the reshape package. *J Statistical Software*. 2007;21(12):1–20. http://www.jstatsoft.org/v21/i12/. Accessed June 24, 2016.
18. Revelle W. Psych: procedures for psychological, psychometric, and personality research. Evanston, IL: Northwestern University; 2014:165. https://cran.r-project.org/web/packages/psych/index.html. Accessed June 28, 2016.
19. American Board of Emergency Medicine. Qualifying examination description and content specifications. https://www.abem.org/public/emergency-medicine-(em)-initial-certification/qualifying-examination/qualifying-examination-description-and-content-specificiations. Accessed June 24, 2016.
20. Marco CA, Counselman FL, Korte RC, et al. Delaying the American Board of Emergency Medicine qualify examination is associated with poorer performance. *Acad Emerg Med*. 2014;21(6):668–693.

Francis L. Counselman, MD, is Professor and Chairman, Department of Emergency Medicine, Eastern Virginia Medical School and Emergency Physicians of Tidewater; **Terry Kowalenko, MD,** is Professor and Chairman, Department of Emergency Medicine, Oakland University William Beaumont School of Medicine; **Catherine A. Marco, MD,** is Professor, Department of Emergency Medicine, Wright State University; **Kevin B. Joldersma, PhD,** is a Psychometrician, American Board of Emergency Medicine; **Robert C. Korte, PhD,** is a Psychometrician, American Board of Emergency Medicine; and **Earl J. Reisdorff, MD,** is Executive Director, American Board of Emergency Medicine.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Francis L. Counselman, MD, Eastern Virginia Medical School, Raleigh Building, Room 304, 600 Gresham Drive, Norfolk, VA 23507, 757.388.3397, counsefl@evms.edu

Received November 5, 2015; revision received March 7, 2016; accepted May 2, 2016.