

# Fourth-Year Medical School Course Load and Success as a Medical Intern

Christopher J. Richards, MD  
 Kenneth J. Mukamal, MD, MPH  
 Nikki DeMelo, MS  
 C. Christopher Smith, MD, FACP

## ABSTRACT

**Background** The fourth year of medical school has come under recent scrutiny for its lack of structure, cost- and time-effectiveness, and quality of education it provides. Some have advocated for increasing clinical burden in the fourth year, while others have suggested it be abolished.

**Objective** To assess the relationship between fourth-year course load and success during internship.

**Methods** We reviewed transcripts of 78 internal medicine interns from 2011–2013 and compared the number of intensive courses (defined as subinternships, intensive care, surgical clerkships, and emergency medicine rotations) with multi-source performance evaluations from the internship. We assessed relative risk (RR) and 95% confidence interval (CI) of achieving excellent scores according to the number of intensive courses taken, using generalized estimating equations, adjusting for demographics, US Medical Licensing Examination (USMLE) Step 1 board scores, and other measures of medical school performance.

**Results** For each additional intensive course taken, the RR of obtaining an excellent score per intensive course was 1.05 (95% CI 1.03–1.07,  $P < .001$ ), whereas the RR per nonintensive course taken was 0.99 (95% CI 0.98–1.00,  $P = .03$ ). An association of intensive course work with increased risk of excellent performance was seen across multiple clinical competencies, including medical knowledge (RR 1.08, 95% CI 1.04–1.11); patient care (RR 1.07, 95% CI 1.04–1.10); and practice-based learning (RR 1.05, 95% CI 1.03–1.09).

**Conclusions** For this single institution's cohort of medical interns, increased exposure to intensive course work during the fourth year of medical school was associated with better clinical evaluations during internship.

## Introduction

Many undergraduate medical education programs are redesigning their curriculum and assessment methods to meet a changing practice landscape.<sup>1</sup> Educators largely have focused on the first 3 years of the traditional 4-year undergraduate curriculum to address concerns about the cost of medical education and declining interest in primary care.<sup>2,3</sup> Some medical schools have eliminated the fourth year entirely, while others believe it is critical to professional development and, in support, cite declining board certification performance.<sup>4–6</sup> Residency program directors also raise concerns that interns lack self-reflective skills, leading to underdeveloped professionalism, weak medical knowledge, and lack of preparedness to manage medical emergencies.<sup>7,8</sup> To address these issues, some advocate for a more rigorous undergraduate experience.<sup>9</sup>

At the same time, little attention has been paid to the composition and quality of experiences during a fourth year of medical school, which represents the

last opportunity to expand clinical skills and knowledge before learners become residents.<sup>10,11</sup> The subinternship experience, considered the cornerstone of the fourth year, lacks national standards for content and assessment.<sup>2,11</sup> For many medical schools, the medical subinternship is the only requisite course in the fourth-year curriculum. At a majority of schools, the remainder of the year is largely unstructured, with students choosing from a variety of clinical and nonclinical electives.<sup>12</sup>

We sought to assess the impact of the fourth year on clinical performance during internal medicine internship. We examined 2 consecutive classes of interns to determine how their fourth-year experiences, including the number and intensity of courses, related to their multi-source assessments of performance based on the Accreditation Council for Graduate Medical Education (ACGME) educational milestones.<sup>13</sup>

## Methods

### Design and Participants

We conducted a single center study of interns enrolled in the internal medicine residency program (IMRP) at Beth Israel Deaconess Medical Center (BIDMC) from

DOI: <http://dx.doi.org/10.4300/JGME-D-16-00043.1>

*Editor's Note: The online version of this article contains sample questions taken from the evaluation tools.*

2011 until 2013. BIDMC is a 600-bed academic hospital; the IMRP has approximately 47 categorical and 13 preliminary interns each year.

### Transcript Collection and Coding

The program receives final medical school transcripts for most interns. One author (N.D.) deidentified the transcripts and assigned identification numbers to protect anonymity prior to coding. The transcripts were coded and entered into a REDCap (Research Electronic Data Capture, Vanderbilt University, Nashville, TN) database. The authors a priori defined intensive clinical courses as experiences with a higher order of clinical responsibility and knowledge than the average course. These included subinternships of any variety, intensive care, and surgical and emergency medicine rotations. Research, less relevant patient care specialties (pathology and radiology), didactic courses, and language courses were defined as not clinically intensive. The authors reviewed all categorizations; the 3 instances of difficulty interpreting transcript information were resolved by consensus.

### Multi-Source Assessment

The IMRP maintains assessments of all residents using New Innovations, a confidential online assessment tool. Evaluations are based on the ACGME's 6 competencies and milestones.<sup>13</sup> After most clinical rotations, attendings, residents, fellows, medical students, and nurses evaluate interns using questionnaires specific for the rotation and evaluator.

Evaluations use a 5- or 9-point scale, depending on type and specific question (see online supplemental material for sample evaluations). Based on a strong ceiling effect of the total distribution of scores with clustering at maximal values, we defined "excellent scores" as an 8 or 9 on the 9-point scale and 5 on the 5-point scale. For robustness and ease of interpretation, we also established an outcome of a "poor score" on any individual item as 6 or less (9-point scale) and 3 or less (5-point scale).

### Covariates

We collected demographic information (age, sex, race, and categorical versus preliminary status) and metrics of medical school performance (US Medical Licensing Examination [USMLE] Step 1 score, Alpha Omega Alpha [AOA] membership, and additional degrees) from residency applications. We a priori grouped medical student performance evaluations and medical school ranking into 3 categories based on previous department determinations.

#### What was known and gap

There is interest in enhancing the relevance of the fourth year of medical school for residency and practice.

#### What is new

Intensive clinical course work taken during the fourth year was associated with enhanced excellent performance in the domains of medical knowledge, patient care, and practice-based learning.

#### Limitations

Single specialty, single elite institution sample reduces generalizability.

#### Bottom line

Increased exposure to intensive clinical course work during the fourth year was associated with better clinical evaluations during internship.

This project was reviewed by the Committee on Clinical Investigations at BIDMC and was approved with exemption from full review.

### Statistical Methods

To account for the multiple questionnaires within-intern and within-rater, we performed all analyses using generalized estimating equations, with the individual item as the unit of analysis. We estimated odds ratios (ORs) for the likelihood of the primary outcome, excellent scores, using binomial error structures, a log link, and an exchangeable correlation matrix, with hierarchical clustering by both intern and rater. For robustness, we estimated OR for the likelihood of poor scores similarly, using a logit rather than log link. In all cases, we constructed both models that only included the number of intensive and nonintensive courses and models that further adjusted for outlined covariates.

We categorized the intensity of course loads in multiple complementary ways by examining the proportion of time spent in intensive or nonintensive activities. We also assessed the number of courses taken, adjusting for intensive and nonintensive coursework. We examined the individual course type as described above, and treated the proportion and number of intensive courses as linear variables (tests of curvature using quadratic terms were not significant). We also present deciles for illustrative purposes.

## Results

### Demographics, Course Load, and Evaluations

Of 115 interns eligible for participation in the study, we were able to obtain 83 medical school transcripts, of which 5 were not interpretable and excluded. The demographics are summarized in TABLE 1; 3 interns held additional degrees (PhD/MS). A summary of total completed fourth-year courses and breakdown

TABLE 1

Demographic Information of 78 Interns Included in Final Analysis and 32 Interns With Transcripts Not Available

Characteristic	Transcript Available	Transcript Unavailable	P Value
Age (SD), y	28.6 (2.6)	29 (2.6)	.26 <sup>a</sup>
Step 1 score (SD)	243 (14)	242 (15)	.57 <sup>a</sup>
Sex, No. (%)			
Male	43 (55)	16 (50)	.68 <sup>b</sup>
Female	35 (45)	16 (50)	
Dean's rank, No. (%)			
Unknown	20 (25)	11 (34)	.84 <sup>b</sup>
Top 10	26 (33)	10 (31)	
Top 33	27 (34)	9 (28)	
Bottom 33	5 (6)	2 (6)	
Medical school tier, No. (%)			
Top	14 (16)	7 (29)	.92 <sup>b</sup>
Middle	36 (46)	14 (44)	
Bottom	28 (37)	11 (34)	
Intern class, No. (%)			
2011–2012	33 (42)	23 (71)	.005 <sup>b</sup>
2012–2013	45 (58)	9 (28)	
Training type, No. (%)			
Categorical medicine	66 (85)	20 (63)	.009 <sup>b</sup>
Preliminary	12 (15)	12 (37)	

Note: Mean values with either percentage or standard deviations are listed above. The group's average Step 1 score was 243 (SD = 14) compared to the national average of 228.

<sup>a</sup> Continuous variables by Wilcoxon rank sum test.

<sup>b</sup> Categorical variables by Fischer's exact test.

of course type is found in TABLE 2. A total of 69 641 individual points from 2350 completed evaluations were available, with a median of 30 evaluations per intern (range 19–56). Of these, 42 203 (61%) assessments met the criteria for excellent and 5724 (8%) for poor.

### Relative Risks and OR of Excellent and Poor Scores

When examined continuously, the relative risk (RR) of an excellent score per intensive course was 1.05

(95% CI 1.03–1.07,  $P < .001$ ), while the corresponding RR per nonintensive course was 0.99 (95% CI 0.98–1.00,  $P = .03$ ); these RRs differed significantly ( $P < .001$ ). When adjusted for demographics, the RR of an excellent score was 1.05 (95% CI 1.03–1.08,  $P < .001$ ) per intensive course and 1.00 (95% CI 0.98–1.01,  $P = .40$ ) per nonintensive course; these again differed significantly from each other ( $P < .001$ ).

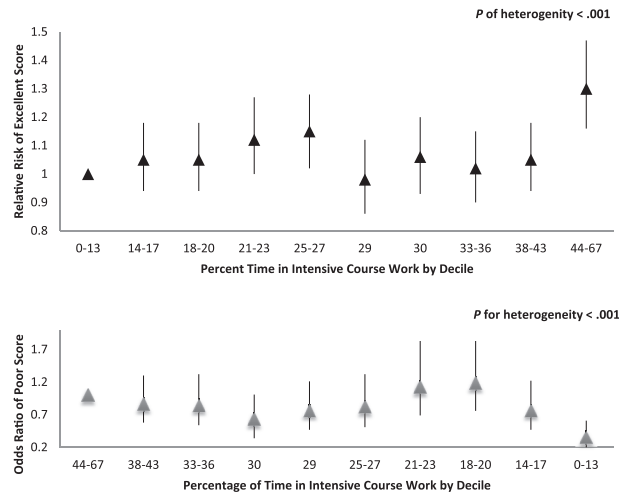
TABLE 2

Breakdown of Course Work Taken During the Fourth Year of Medical School

Courses/Internships	Median	Minimum	Maximum
Total clinical courses	6	2	12
Medical courses (sub-I, ICU, subspecialties)	5	1	9
Intensive medical courses (sub-I and ICU)	2	0	4
Medical subinternships	1	0	3
Total subinternships (medical and surgical sub-I)	1	0	3
ICU courses (medical and surgical ICU)	0	0	1
Total intensive courses (all sub-I, all ICU, surgical clerkships, and emergency medicine)	2	0	6
Nonintensive courses	7	2	12

Abbreviations: sub-I, subinternship; ICU, intensive care unit.

Note: The majority of interns completed at least 1 subinternship, but further intensive course work was less common; interns completed a median of 2 intensive courses (range 0–6). Categorical and preliminary interns did not differ in the number of intensive courses or medical subinternships they took.



**FIGURE**  
Outcomes of Obtaining Both Excellent and Poor Scores by Decile of Time Spent Pursuing Intensive Course Work During Fourth Year of Medical School

A second analysis accounting for variable course lengths (median 4 weeks) assessed the relationship of percentage time in intensive courses with evaluations. The upper chart in the **FIGURE** depicts the adjusted RR of obtaining an excellent score pursuing intensive course work, with the referent being the lowest decile of intensive course work (decile 1). To determine if the positive association of intensive course work with performance was driven by any of the individual components, we determined the adjusted RR of excellent evaluations based on individual course types. No single type of intensive course work accounted for our findings (**TABLE 3**).

The positive influence of intensive course work was seen in all competencies except professionalism, and our measurement of global assessment independent of any of the ACGME Milestones (**TABLE 4**).

To determine the robustness of these associations, we performed a sensitivity analysis using poor scores. The unadjusted OR for obtaining a poor score per intensive course was 0.92 (95% CI 0.84–1.01,  $P = .07$ ), whereas the OR per nonintensive course

was 1.04 (95% CI 1.00–1.08,  $P = .04$ ). These 2 ORs were significantly different from one another ( $P = .02$ ); these differences were similar but were no longer statistically significant after adjustment for demographics ( $P = .12$ ). Similarly, there was a persistent decrease in the OR of a poor evaluation with increasing time spent taking intensive course work ( $P < .001$ ; **FIGURE**).

We performed an additional sensitivity analysis of the 2350 evaluations; 532 (23%) were uniformly excellent. The RR of such an evaluation for each additional intensive course was 1.13 (95% CI 1.05–1.21,  $P = .001$ ). The corresponding RR for each nonintensive course was 0.97 (95% CI 0.94–1.01,  $P = .01$ ). These 2 estimates differed significantly from each other ( $P < .001$ ).

## Discussion

In this study of medical interns in 1 large residency program, the quantity and intensity of medical school courses taken in the fourth year had a small, but significant and dose-dependent, association with

**TABLE 3**  
Adjusted Relative Risk for Excellent Evaluation per 10% of Time Spent in Individual Course Types<sup>a</sup>

Course Types	Relative Risk	Confidence Interval	P Value
All clinical courses	0.99	0.98–1.00	.17
Intensive courses	1.05	1.03–1.07	< .001
Nonintensive courses	0.95	0.93–0.97	< .001
Medical courses	1.01	0.99–1.03	.23
All subinternships	1.01	0.98–1.05	.35
Medical subinternships	1.01	0.97–1.05	.77
Intensive care units	1.04	0.99–1.08	.12

<sup>a</sup> Adjusted for age, sex, minority status, dean's rank, medical school tier, intern year, categorical, and Step 1 score.

TABLE 4

Adjusted Relative Risk of Excellent Score for Each Additional Intensive Course in Each ACGME Clinical Competency and Global Assessment Metric<sup>a</sup>

Clinical Competency	Relative Risk	Confidence Interval	P Value
Patient care	1.07	1.04–1.10	< .001
Medical knowledge	1.08	1.04–1.11	< .001
Systems-based practice	1.06	1.02–1.10	.002
Communication and interpersonal skills	1.04	1.02–1.07	.002
Professionalism	1.01	0.99–1.03	.48
Practice-based learning	1.05	1.03–1.09	< .001
Global assessment	1.11	1.06–1.15	< .001

Abbreviation: ACGME, Accreditation Council for Graduate Medical Education.

<sup>a</sup> Adjusted for age, sex, minority status, dean's rank, medical school tier, intern year, categorical, Step 1 score, and number of intensive courses.

clinical performance during internship. This effect was seen across all ACGME competencies except professionalism, and persisted when corrected for potential confounders. The association of intensive courses with better performance differed significantly from the corresponding association with nonintensive courses and strengthens the plausibility that intensive course work has a measurable impact on intern performance.

The observed effect from intensive courses was robust and seen for all types of evaluations, for most ACGME clinical competencies, and for global performance, where the relationship was the strongest. The 1 exception was professionalism. Other studies suggest a fundamental difference between professionalism and other competencies, theorizing it is more difficult to teach, correct, and change over time.<sup>14–17</sup>

Our program uses robust assessment tools based on the ACGME core competencies that considers input from students to attending physicians. We analyzed nearly 70 000 points of assessment, which afforded the power to detect subtle differences in the performance of high-performing medical interns. Our data not only support the argument that the fourth year should be maintained, but also that its clinical intensity should be strengthened to produce “clinical-ready” graduates.

We do not believe that nonintensive course work, as we have defined here, is without value. There is certainly benefit to research and nonclinical specialty exposure. In this regard, our performance measures may not capture the value of such courses, as we focused on intern clinical performance—which has the most direct relationship to courses—but not on long-term success or satisfaction. Nonetheless, medical students could reasonably be advised to take intensive courses during their fourth year to improve their clinical performance during internship.

Our study has limitations. We studied a single academic residency program, and results may not generalize to other programs. However, the interns in this study represented 46 different medical schools, enhancing our generalizability. In our analysis, poor scores were rarely given for intern performance. We observed a low median number of fourth-year courses taken by our intern classes. Without comparable information from other programs, we cannot necessarily extrapolate our results to the incremental value of intensive courses when students take a larger number of courses.

Interns at BIDMC are academically talented, with high levels of AOA membership and above-average USMLE Step 1 scores.<sup>18</sup> This had the expected consequence of a ceiling effect in evaluations, minimizing the variability of performance within the cohort. This tends to reduce our ability to detect differences among interns, leading to a possible underestimate of benefit.

Another limitation of observational studies like ours is the ability to infer causality in the presence of confounding. Students with strong clinical backgrounds may disproportionately select demanding fourth-year programs. Although we controlled for several potential markers of performance (AOA, USMLE Step 1 score, and reputation of medical school), none of these factors combined or individually materially confounded our primary estimates of association. We were limited to these proxy markers of achievement; we did not have access to USMLE Step 2 scores, and the heterogeneity of medical school grading precluded the use of honors. Even more subjective concepts, such as medical student motivation, are not readily measured by any routinely used instrument. Ultimately, the only way to fully control for all forms of identified and unidentified bias would be to perform a randomized trial, which, in this setting, is unlikely to occur. Of note, our results do

identify variables of potential importance for clinical training that would be helpful to program directors.

## Conclusion

In a single institution's cohort of medical interns, the selection of clinically intensive course work during the fourth year of medical school had a small, but significant, dose-dependent, and wide-ranging impact on clinical evaluations of performance. This association was not influenced by other potential predictors of high performance, and was not matched by improved performance with less intensive courses.

## References

1. Cosgrove EM, Ryan MJ, Wenrich MD. Empowering fourth-year medical students: the value of the senior year. *Acad Med.* 2014;89(4):533–535.
2. Walling A, Merando A. The fourth year of medical education: a literature review. *Acad Med.* 2010;85(11):1698–1704.
3. Stevens CD. Taking back year 4: a call to action. *Acad Med.* 2010;85(11):1663–1664.
4. Emanuel EJ, Fuchs VR. Shortening medical training by 30%. *JAMA.* 2012;307(11):1143–1144.
5. American Board of Internal Medicine. Internal Medicine Certification Exam. <http://www.abim.org/exam/certification/internal-medicine.aspx>. Accessed November 2, 2016.
6. New England Journal Knowledge+. ABIM pass rates: behind the declines. <http://knowledgeplus.nejm.org/abim-pass-rates-behind-declines/>. Accessed November 2, 2016.
7. Lyss-Lerman P, Teherani A, Aagaard E, et al. What training is needed in the fourth year of medical school? Views of residency program directors. *Acad Med.* 2009;84(7):823–829.
8. McEvoy MD, DeWaay DJ, Vanderbilt A, et al. Are fourth-year medical students as prepared to manage unstable patients as they are to manage stable patients? *Acad Med.* 2014;89(4):618–624.
9. Langdale LA, Schaad D, Wipf J, et al. Preparing graduates for the first year of residency: are medical schools meeting the need? *Acad Med.* 2003;78(1):39–44.
10. Barzansky B, Simon FA, Brotherton SE. The fourth-year medical curriculum: has anything changed in 20 years? *Acad Med.* 2001;76(suppl 10):36–38.
11. Sprague C. Articulation of a largely elective fourth year with the traditional internship. Association of American Medical Colleges. Presented at the 78th Annual Meeting of the AAMC; New York, NY; October 29, 1967.
12. Green EH, Fagan MJ, Reddy S, et al. Advances in the internal medicine subinternship. *Am J Med.* 2002;113(9):769–773.
13. Accreditation Council for Graduate Medical Education. Implementing Milestones and Clinical Competency Committees. April 24, 2013. <https://www.acgme.org/acgmeweb/Portals/0/PDFs/ACGMEMilestones-CCC-AssessmentWebinar.pdf>. Accessed November 2, 2016.
14. Arnold L. Assessing professional behavior: yesterday, today and tomorrow. *Acad Med.* 2002;77(6):502–515.
15. ten Cate O, Durning S. Peer teaching in medical education: twelve reasons to move from theory to practice. *Med Teach.* 2007;29(6):591–599.
16. Small PA Jr, Stevens CB, Duerson MC. Issues in medical education: basic problems and potential solutions. *Acad Med.* 1993;68(suppl 10):89–98.
17. Schönrock-Adema J, Heijne-Penning M, van Duijn MA, et al. Assessment of professional behavior in undergraduate medical education: peer assessment enhances performance. *Med Educ.* 2007;41(9):836–842.
18. National Resident Matching Program; Association of American Medical Colleges. Charting outcomes in the Match: characteristics of applications who matched to their preferred specialty in the 2011 main residency match. <http://www.nrmp.org/wp-content/uploads/2013/08/chartingoutcomes2011.pdf>. Accessed November 2, 2016.



**Christopher J. Richards, MD**, is a Fellow, Pulmonary and Critical Care Medicine, Harvard Combined Fellowship Program, Beth Israel Deaconess Medical Center (BIDMC) and Massachusetts General Hospital; **Kenneth J. Mukamal, MD, MPH**, is Associate Professor of Medicine, Harvard Medical School, Visiting Scientist, Harvard School of Public Health, and Associate Director, Internal Medicine Residency Program, BIDMC; **Nikki DeMelo, MS**, is Program Administrator, Internal Medicine Residency Program, BIDMC; and **C. Christopher Smith, MD, FACP**, is Associate Professor of Medicine, Harvard Medical School, and Director, Internal Medicine Residency Program, BIDMC.

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

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

Corresponding author: Christopher J. Richards, MD, Beth Israel Deaconess Medical Center, Pulmonary Medicine, Kirstein Basement 23, 330 Brookline Avenue, Boston, MA 02215, 617.667.5864, fax 617.667.4849, crichar5@bidmc.harvard.edu

Received January 16, 2016; revision received June 13, 2016; accepted August 15, 2016.