Correcting the Bias of Clerkship Timing on Academic Performance

Julie E. Cho, AB; John M. Belmont, PhD; Cheng T. Cho, MD, PhD

Objectives: To document the time-of-year bias in National Board of Medical Examiners subject examination (NBME) scores in a third-year pediatrics clerkship and to develop a grading method that neutralizes the bias.

Design: Interventional modeling of final grades.

Setting: University-based medical school.

Subjects and Methods: During each of the past 3 academic years, we conducted six 2-month pediatric clerkships for third-year students. To counter the time-of-year bias, NBME scores, clinical evaluations, and departmental examination scores for the current rotation were pooled with those from the rotations from the same time of year during the previous 2 years. Final grades for the current rotation were determined by cutoff points derived from that entire 3-year pool. We analyzed this approach by testing the time-of-year effects on NBME scores, clinical evaluations, and final grades while maintaining step 1 of the US Medical Licensing Examination as a preclinical baseline control.

Results: The scores for step 1 of the US Medical Licensing Examination did not differ significantly by time of year. Clinical evaluations and NBME scores showed significant upward trends as the academic year progressed. By contrast, according to design, final grades showed no significant time-of-year trend.

Conclusions: Our results support previous reports of significant improvements in NBME scores across the academic year. Our method of computing final grades corrects for this time-of-year bias by judging students only in relation to those who took the rotation at the same time of year. It is our belief that the prevalence and significance of the time-of-year trend warrants such an adjustment in grading to help minimize the academic disadvantage faced by students early in their clinical training.


Editor's Note: I'm all for anything that leads to fairer evaluations of anyone. However, I still remain concerned about the emphasis placed on grades in what's supposed to be adult education.

Catherine D. DeAngelis, MD

The National Board of Medical Examiners subject examination (NBME) is used by more than 60% of the pediatrics departments of US medical schools as a component of clerkship grading. The NBME is attractive because it is nationally standardized and is constructed to cover the subject matter broadly. It has become apparent, however, that the NBME presents a challenge for educational measurement: medical educators are reporting significant effects of clerkship timing (ie, time of year) on NBME scores and final grades.

Whalen and Moses found that student scores on the NBME medicine clerkship examination and final grades significantly increased as the academic year progressed. Baciewicz et al and Ripkey et al reported a similar trend for the NBME surgery clerkship examination. Clark and Jelovsek, Manetta et al, and Smith et al reported that mean NBME obstetrics and gynecology clerkship scores were higher later in the year; and Hampton et al reported the time-of-year trend for NBME obstetrics and gynecology scores and for final grades in obstetrics and gynecology and pediatrics. In contrast to these results, the time-of-year effect on NBME psychiatry scores appears to be very weak or nonexistent.

A common speculation about the time-of-year trend is that students coming to a given clerkship later in the year have an advantage because of generalized clinical experience and technical
SUBJECTS AND METHODS

STUDENTS

There were six 2-month pediatrics clerkship rotations (July-August, September-October, November-December, January-February, March-April, and May-June) in each of the following 3 academic years: 1994-1995 (n = 129), 1995-1996 (n = 120), and 1996-1997 (n = 118) (total sample, n = 367). The median number of students per rotation was 21 (range, 12-24; interquartile range, 19-22). Students were assigned to rotations according to a long-standing policy that combines student choice and lottery. This may be regarded as a sample of convenience. No information was available concerning factors that may have influenced individual students’ assignments to any particular rotation. We have observed informally that very few students had strong specialty preferences going into their third year.

MEASURES

The students took step 1 of the US Medical Licensing Examination (USMLE-1) at the end of the second year, before being assigned to any third-year clerkship. The USMLE-1 is highly correlated with NBME scores, and therefore was used as an appropriate preclinical baseline measure to control statistically for possible academic biases associated with time of year.

A departmental written examination was administered midway through each rotation. It was worth 13% of the final grade (results not reported herein). At the end of the rotation, the student took the NBME in pediatrics (35% of the final grade). Before announcement of the NBME results, each student underwent subjective evaluation for clinical ability by an average of 3 attending physicians, 3 senior residents, and 2 junior residents. Each evaluator gave a qualitative grade (superior [roughly equivalent to an “A”], high satisfactory [“B”], satisfactory [“C”], low satisfactory [“D”], and unsatisfactory [“F”]) in each of the following 4 domains: factual knowledge, practical skill, personal and professional behavior, and values. Numeric scores were assigned for each of the 4 domains and then averaged for each evaluator. The evaluators’ scores were averaged within level (attending physician, senior resident, and junior resident), weighted by level, and finally averaged to make up 50% of the final grade.

GRADING

To neutralize the time-of-year trend, we graded the students by pooling their scores with those of all students who had taken the rotation at that same time of year during the past 2 years (or, in cases where data from previous years were not yet sufficiently numerous, the rotation directly before theirs). Each student was thus compared with approximately 60 others, all of whom had taken pediatrics at almost the same time of year. For each measure (departmental examination, NBME score, and clinical evaluation), the approximately 60 scores were standardized to mean (SD) of 86 (5.19). The 3 standardized scores were multiplied by their respective weights (0.15, 0.35, and 0.50), and the 3 weighted scores were summed to determine the final grade. Using the department’s traditional numerical grading policy, this standardization would yield approximately 23% grades of superior (A), 50% grades of high satisfactory (B), and 23% grades of satisfactory (C) for the entire group of 60. The actual distribution of grades departed from this ideal for any particular rotation of approximately 20 students, depending on their strengths relative to the entire group with whom they were standardized.

STATISTICAL ANALYSIS

Descriptive statistics and statistical tests were computed using BMDP PC-90. Data are reported as mean (SD). Time-of-year trends are graphed as mean raw scores. For each outcome measure, the effects of year, time of year, and their interaction were tested using 2-way factorial analysis of variance (ANOVA) with Brown-Forsythe adjustments for unequal variances; post hoc follow-up comparisons were performed using Student-Newman-Keuls tests. Specific hypotheses concerning time-of-year trends were tested separately using linear and quadratic contrasts. Correlations among measures were tested using the Pearson product moment correlation r (α = .05 throughout). Complete data were available for all students for all variables except USMLE-1, for which 6 students’ data (1.6%) could not be identified because of name change (n = 4) or could not be used because of a different scoring system (n = 2).

knowledge acquired during that year’s previous clerkships.

When the NBME score is considered in computing the student’s final grade, the time-of-year trend presents an interesting problem to the medical educator and the medical student. From one perspective, the trend may be accepted, and the student may be encouraged to capitalize on it, rather than fall passive victim to “the accident of scheduling.”12 In this vein, Baciewicz et al6 and Ripkey et al6 recommended that students should take their surgery clerkship later in the year if they wish to maximize their performance in that specialty, and Hampton et al10 maintained that students who aspire to enter obstetrics and gynecology should be aware of the end-of-year grade advantage. Whalen and Moses’11 concurred with respect to the internal medicine rotation.

“Handicapping” of grades and “special compensation” have been mentioned as alternatives to advising students to take advantage of the time-of-year effect. However, no handicapping system seems to have been described. We believe that it can be an actual and unproductive burden on the student to need to play a statistical advantage when specialty choices are far from settled. In our experience, at the end of the second year most students have only a weak idea concerning their ultimate choice of specialty, and strong specialty preferences often develop unexpectedly on particular services long after the accident of rotation scheduling. With this in mind, in the academic year, 1994-1995 we developed a grading system for the pediatrics clerkship that is specifically aimed at neutralizing the time-of-year handicap. We herein report our method and experience during the past 3 years.
Year of clerkship was not significantly related to NBME score (P = .47), clinical evaluation (P = .10), final grade (P = .97), or USMLE-1 (P = .88). Therefore, the data were combined for all 3 years and analyzed by time of year.

### RESULTS

#### Year of Clerkship

Across the 6 rotations, mean (SD) USMLE-1 scores were 210 (20), 208 (15), 204 (18), 207 (15), 205 (20), and 207 (17) (Figure). There were no significant differences or trends for time of year (for the overall ANOVA and the linear and quadratic trends, all P values >.23). This flat preclinical baseline suggests that time-of-year trends observed in other measures did not result from accidental confounding of time of year with academic ability.4-6,8,11

### NBME SCORES

Mean (SD) NBME scores for the 6 rotations were 477 (110), 466 (80), 484 (112), 506 (112), 523 (86), and 526 (83) (Figure). The increase across time of year was significant (linear trend, P<.001), but occurred mostly in the early rotations (quadratic trend, P = .03).

#### Clinical Evaluations

Mean (SD) clinical evaluation scores for the 6 rotations were 87.1 (2.9), 87.6 (3.2), 89.3 (2.2), 89.6 (3.6), 89.7 (3.1), and 89.8 (2.9) (Figure). The increase across time of year was significant (linear trend, P<.001), but occurred mostly in the early rotations (quadratic trend, P = .03).

### WEIGHTED FINAL GRADES

Mean (SD) final grades for the 6 rotations were 86.3 (4.3), 86.4 (4.2), 87.2 (3.6), 86.6 (4.4), 86.5 (4.1), and 86.1 (3.9) (Figure). The time-of-year function was flat, showing no significant changes or trends as the academic year progressed (all P values >.21).

### CORRELATIONS AMONG MEASURES

For the sample as a whole, the 3 outcome measures were significantly intercorrelated. For final grade vs NBME score, r = 0.75 (P<.001); for final grade vs clinical evaluation, r = 0.78 (P<.001). These correlations will not be discussed further because measures that form part of the final grade will necessarily be correlated with it. The clinical evaluation was independent of NBME, however, so their correlation (r = 0.48; P<.001) is of interest.

### COMMENT

Data from our pediatrics clerkship confirm the upward time-of-year trend in NBME scores identified in earlier reports for medicine,4 surgery,5,6 and obstetrics and gynecology clerkships.7,8,10 Our findings also support the data of Hampton et al.,10 who reported that NBME pediatrics scores increased toward the end of the year (although nonsignificantly in their data).

Few studies have examined the relation of clerkship timing and subjective evaluations. Whalen and Moses4 found no timing effect in the percentage of students receiving outstanding or above-average subjective evaluations in internal medicine.4 Baciewicz et al.2 reported likewise for oral examinations in surgery. In contrast, our analysis showed a significant timing bias in subjective clinical evaluations that paralleled that of the NBME pediatrics clerkship score. The reasons for the disparity between our findings and those of Whalen and Moses4 and Baciewicz et al.2 are unclear, as the contents and criteria of the programs’ subjective evaluations cannot be compared with precision. A possible explanation for our findings is that our residents’ and attending physicians’ evaluations may have emphasized the impressions of the students’ factual academic knowledge. This interpretation is supported by the significant correlation found herein between the subjective clinical evaluations and the NBME scores (r = 0.48).

In contrast to the NBME results, our data on final clerkship grades differed dramatically from previous reports of time-of-year final grade increases in medicine,4 obstetrics and gynecology,7,10 and pediatrics.10 Our finding of no time-of-year trend in final grades was precisely the desired outcome of our method of calculating grades to neutralize the time-of-year effect. Thus, the approach seems to realize the handicapping concept that was raised hypothetically by Whalen and Moses,4 Ripkey et al.,9 and Clark and Jelovsek7 but apparently not previously implemented.
Beginning in the academic year 1994-1995, we have used a grading system that takes into account clerkship timing when calculating final grades. Due to the bias in objective testing and subjective evaluations favoring students who complete their pediatrics clerkships later in the academic year, we scale our final scores so that student performance is considered only in relation to students taking the pediatric clerkship at the same time in their medical training (e.g., a student in the January-February rotation of academic year 1996-1997 underwent evaluation relative to the January-February rotations from that year plus the 2 previous years).

Given the weight that residency selection committees can put on medical students’ final grades, we believe it is the responsibility of clerkship directors and faculty to develop and refine grading systems that are responsive to known biases such as clerkship timing. Simply recommending that a student take his pediatrics or surgery clerkship later in the year if he has a hunch that he would like to become a pediatrician or a surgeon is an unsatisfying response. In our experience, at the end of their second year most students do not have a strong specialty preference; and almost regardless of preference, the rotation order is determined by the whim of computer lottery. Moreover, encouraging a student to play an early hunch fails to recognize the important role that medical educators play in opening doors of interest and opportunity for their students during the required clerkship year. Because our grading method is sensitive to clerkship timing, it effectively minimizes at least one disadvantage faced by students early in their clinical training.

Accepted for publication April 30, 1998.

We thank Dwayne A. Ollerich, PhD, for assistance in USMLE-1 data acquisition.

Reprints: Cheng T. Cho, MD, PhD, Department of Pediatrics, University of Kansas Medical Center, 3901 Rainbow Blvd, Kansas City, KS 66160.

REFERENCES

13. Federation of State Medical Boards of the US, Inc, and the National Board of Medical Examiners. USMLE Step 1 General Instructions, Content Description, and Sample Items. Philadelphia, Pa: National Board of Medical Examiners; 1996.