

Mortality Among Hospitalized Medicare Beneficiaries in the First 2 Years Following ACGME Resident Duty Hour Reform

Kevin G. Volpp, MD, PhD

Amy K. Rosen, PhD

Paul R. Rosenbaum, PhD

Patrick S. Romano, MD, MPH

Orit Even-Shoshan, MS

Yanli Wang, MS

Lisa Bellini, MD

Tiffany Behringer, MS

Jeffrey H. Silber, MD, PhD

WIDESPREAD CONCERN about the number of deaths in US hospitals from medical errors¹ prompted the Accreditation Council for Graduate Medical Education (ACGME) to implement duty hour regulations effective July 1, 2003, for all ACGME-accredited residency programs.^{2,3} Work limitations for residents included no more than 80 hours per week, with 1 day in 7 free of all duties, averaged over 4 weeks; no more than 24 continuous hours with an additional 6 hours for education and transfer of care; in-house call no more frequently than every third night; and at least 10 hours of rest between duty periods.²

Although there is extensive scientific evidence linking fatigue and impaired cognitive performance,⁴⁻⁹ little empirical data guided the design of the duty hour regulations,^{10,11} and the design itself remains controversial,^{12,13} with contrasting views as to whether the duty

Context The Accreditation Council for Graduate Medical Education (ACGME) implemented duty hour regulations for physicians-in-training throughout the United States on July 1, 2003. The association of duty hour reform with mortality among patients in teaching hospitals nationally has not been well established.

Objective To determine whether the change in duty hour regulations was associated with relative changes in mortality among Medicare patients in hospitals of different teaching intensity.

Design, Setting, and Patients An observational study of all unique Medicare patients (N=8 529 595) admitted to short-term, acute-care, general US nonfederal hospitals (N=3321) using interrupted time series analysis with data from July 1, 2000, to June 30, 2005. All Medicare patients had principal diagnoses of acute myocardial infarction, congestive heart failure, gastrointestinal bleeding, or stroke or a diagnosis related group classification of general, orthopedic, or vascular surgery. Logistic regression was used to examine the change in mortality for patients in more vs less teaching-intensive hospitals before (academic years 2000-2003) and after (academic years 2003-2005) duty hour reform, adjusting for patient comorbidities, common time trends, and hospital site.

Main Outcome Measure All-location mortality within 30 days of hospital admission.

Results In medical and surgical patients, no significant relative increases or decreases in the odds of mortality for more vs less teaching-intensive hospitals were observed in either postreform year 1 (combined medical conditions group: odds ratio [OR], 1.03; 95% confidence interval [CI], 0.98-1.07; and combined surgical categories group: OR, 1.05; 95% CI, 0.98-1.12) or postreform year 2 (combined medical conditions group: OR, 1.03; 95% CI, 0.99-1.08; and combined surgical categories group: OR, 1.01; 95% CI, 0.95-1.08) compared with the prereform years. The only condition for which there was a relative increase in mortality in more teaching-intensive hospitals postreform was stroke, but this association preceded the onset of duty hour reform. Compared with nonteaching hospitals, the most teaching-intensive hospitals had an absolute change in mortality from prereform year 1 to postreform year 2 of 0.42 percentage points (4.4% relative increase) for patients in the combined medical conditions group and 0.05 percentage points (2.3% relative increase) for patients in the combined surgical categories group, neither of which were statistically significant.

Conclusion The ACGME duty hour reform was not associated with either significant worsening or improvement in mortality for Medicare patients in the first 2 years after implementation.

JAMA. 2007;298(9):975-983

www.jama.com

Author Affiliations are listed at the end of this article.
Corresponding Author: Kevin G. Volpp, MD, PhD, Center for Health Equity Research and Promotion

(CHERP), Philadelphia Veterans Affairs Medical Center, 3900 Woodland Ave, Philadelphia, PA 19104-6021 (volpp70@wharton.upenn.edu).

See also pp 984 and 1055.

hour rules adversely affect patient care,¹⁴⁻¹⁶ benefit patient care,¹⁷ or do not go far enough in limiting work hours for physicians in training.^{18,19} A necessary by-product of the reform has been an increase in the number of handoffs between residents, which has led to concerns about the continuity of care postreform.¹⁴⁻¹⁶ Similar work hour reforms in New York State, implemented in 1989, were not associated with changes in mortality rates for patients with heart failure, myocardial infarction, or pneumonia,²⁰ or rates for several procedural complications, although higher rates of some types of complications were observed.^{21,22} A recent evaluation of the ACGME reform, potentially limited by its data source, suggested that improvements in mortality for medical but not surgical patients may have been associated with implementation of the reform.²³

Because of this, we studied the association between changes in the ACGME duty hour rules and mortality rates among Medicare patients hospitalized in short-term, acute-care US nonfederal hospitals. We compared trends in risk-adjusted mortality rates among more vs less teaching-intensive hospitals to examine whether mortality changed differentially among these groups following implementation of the rules. An accompanying article reports a complementary analysis among patients hospitalized within the US Veterans Affairs (VA) health care system.²⁴

METHODS

Approval for this study was obtained from the institutional review boards of the University of Pennsylvania and The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.

Main Outcome Measures

The main outcome measure was death within 30 days of hospital admission for all patients admitted for acute myocardial infarction (AMI), stroke, gastrointestinal bleeding, congestive heart failure (CHF), general surgery, orthopedic surgery, or vascular surgery. The medical conditions were a subset of the

Agency for Healthcare Research and Quality (AHRQ) Quality Indicators for which mortality was a relevant outcome. Although the AHRQ Quality Indicators use only in-hospital mortality, we examined any in-hospital or postdischarge deaths within 30 days of hospital admission, a measure that eliminates bias due to length of stay differences across hospitals or time.²⁵ For these study conditions, there is evidence that mortality varies substantially across institutions and that high mortality may be associated with deficiencies in the quality of care.²⁶⁻²⁸

Study Sample

All Medicare patients admitted to short-term, acute-care, general US nonfederal hospitals from July 1, 2000, to June 30, 2005, with a principal diagnosis of AMI, CHF, gastrointestinal bleeding, or stroke or with a diagnosis related group classification of general, orthopedic, or vascular surgery comprised the sample. The initial sample included 12 052 344 patients from 5736 acute care hospitals that contributed data for all 5 years within the 50 states or Washington, DC.

We excluded patients from hospitals that opened or closed during this period (9600 patients from 33 hospitals), that had fewer than 350 admissions in any year (40 756 patients from 1640 hospitals), that did not have Medicare Cost Report data (14 514 patients from 39 hospitals), or that were missing more than 3 months of data in the prereform period or 2 months worth of data in the postreform period (276 040 patients from 703 hospitals). Hospitals with fewer than 350 total admissions for all Medicare patients per year in any year (a mean of <1 admission per day across all conditions) were removed to screen out hospitals that may not have been acute care facilities as well as those that were so small that they were likely to yield unstable estimates if included in a fixed effects analysis.

We also excluded patients younger than 66 years (n=1 562 532) to allow a 180-day lookback, which utilized information on secondary diagnoses coded in hospitalizations within the

180 days before admission for better risk adjustment; older than 90 years (n=720 191), because the proportion of such patients that are treated aggressively may change over time in ways that cannot be observed well with administrative data; whose hospitalization spanned July 1, 2003 (n=26 856); who were enrolled in a health maintenance organization at some point during the study period (n=128 778); who had dates of death earlier than their discharge dates (n=488); or who were transferred in from other hospitals (n=13 615). Among patients with AMI or stroke, those patients discharged alive in fewer than 2 days (n=50 926) were excluded because such cases may not represent actual AMIs or strokes.²⁸

So that each patient would only be represented once within each analysis, we defined an index admission as the first eligible admission between July 1, 2000, and June 30, 2005, for which there was no prior admission for the medical condition or surgical category within 5 years (using data back to July 1, 1995). We chose the first admission in the past 5 years for each patient to eliminate the possibility of selecting cases for which there would be a higher mortality rate postreform for reasons other than duty hour reform. Including multiple admissions for each patient in a longitudinal evaluation of a policy reform like this would be problematic because for patients with multiple admissions, any admission before the last was by definition an admission they survived. This means that an admission that resulted in death would be more likely to happen postreform, because the reform is confounded by the passage of time. Using only the first admission in the past 5 years excluded 678 453 patients, leaving a total sample for our analysis of 8 529 595 patients from 3321 hospitals.

More than 90% of patients with AMI, stroke, and vascular surgery had a first admission that met these criteria, as did more than 82% of patients with gastrointestinal bleeding, general surgery, and orthopedic surgery. For patients with

CHF, 66% had only 1 admission over the 5-year period. We therefore tested the stability of this approach by also assessing the associations with duty hour reform for patients with CHF who experienced a second admission within 6 months of the first admission, which allowed inclusion of an additional 12% of patients with CHF.

Risk Adjustment and Hospital Control Measures

The risk adjustment approach used was developed by Elixhauser et al.²⁹ The Elixhauser method used 27 comorbidities (excluding fluid and electrolyte disorders and coagulopathy, which should not be used in quality indicator risk adjustment)^{30,31} and has been shown to achieve better discrimination than alternative approaches.^{32,33} This was augmented with adjustments for age and sex. For surgical patients, we also adjusted for diagnosis related groups, grouping diagnosis related groups with and without complications or comorbidities into 1 variable. We performed a 180-day look-back, including data from prior hospitalizations, to obtain more comprehensive information on comorbidities than available by using the index admission alone.³⁴ A series of stability analyses was conducted to assess whether the substantive conclusions were unchanged.³⁵ For patients with AMI, these included additional adjustments for anatomic location of the AMI (*International Classification of Diseases, Ninth Revision [ICD-9]* codes: anterior 410.00-410.19, inferolateral 410.20-410.69, subendocardial 410.7x, other 410.80-410.99). For patients with stroke, this included assessment of whether adjustment for the proportion of hemorrhagic strokes affected the results.

Data

Data on patient characteristics were obtained from the Medicare Provider Analysis and Treatment File (MEDPAR), which includes information on principal and secondary diagnoses, age, sex, comorbidities, and discharge status, including dates of death.³⁶ Data on health maintenance organization status were

obtained from the denominator files obtained from the Center for Medicare and Medicaid Services (CMS). The number of residents at each hospital was obtained from the Medicare Cost Reports from CMS. Because funding for graduate medical education is tied to the number of residents within each hospital under Medicare, hospitals have a strong incentive to report complete data. Data from the American Hospital Association were used to identify hospitals that merged, opened, or closed during the study period.

The primary measure of teaching intensity was the resident-to-bed ratio, calculated at a defined point in time as the number of interns plus residents divided by the mean number of operational beds. The resident-to-bed ratio has been used as an approach to differentiate major teaching, minor teaching, and nonteaching hospitals in previous studies.³⁷⁻³⁹ In cross-sectional comparisons of quality of care in teaching and nonteaching hospitals, findings have been similar regardless of whether the resident-to-bed ratio or measures available through the American Hospital Association such as Council of Teaching Hospitals membership were used.⁴⁰ Teaching hospitals were defined as those hospitals with resident-to-bed ratios of more than 0; major and very major teaching hospitals were those hospitals with a resident-to-bed ratio of 0.250 to 0.599 and 0.60 or higher, respectively.

Resident-to-bed ratios are used by the CMS to calculate indirect medical education (IME) adjustments,⁴¹ supporting legitimacy of the use of the resident-to-bed ratio as a marker of teaching intensity. The IME is the additional amount Medicare pays teaching hospitals for patient care due to the higher costs of care provision in these hospitals. For example, the median IME amounts were much higher for very major teaching hospitals (resident-to-bed ratio, 0.600-1.090; IME, \$15 000 894) compared with major teaching hospitals (resident-to-bed ratio, 0.250-0.599; IME, \$8 878 650), minor teaching hospitals (resident-to-bed ratio, 0.050-0.249; IME, \$1 911 228),

very minor teaching (resident-to-bed ratio, >0-0.049; IME, \$234 549), and nonteaching hospitals (resident-to-bed ratio=0; IME, \$0).

We used the resident-to-bed ratio as a continuous variable, which provided more power for assessing associations with implementing the duty hour rules than dividing hospitals into arbitrary categories.^{42,43} We held the resident-to-bed ratio fixed using the level in prereform year 1 so that any potential behavioral response to the reforms by hospitals (such as changing the number of residents) would not confound estimation of the net effects of the duty hour reform. Resident-to-bed ratios varied little over time. The mean changes from prereform year 3 to prereform year 2, and from prereform year 2 to prereform year 1, were both 0.02. Prereform year 3 included academic year 2000-2001 (July 1, 2000, to June 30, 2001); prereform year 2, academic year 2001-2002; prereform year 1, academic year 2002-2003; postreform year 1, academic year 2003-2004; and postreform year 2, academic year 2004-2005.

Statistical Analysis

We used a multiple time series research design,⁴⁴ also known as difference-in-differences, to examine whether the change in duty hour rules was associated with a change in the underlying trend in patient outcomes in teaching hospitals, an approach that reduces potential biases from unmeasured variables.^{45,46} The multiple time series research design compares each hospital with itself, before and after reform, contrasting the changes in hospitals with more residents to the changes in hospitals with fewer or no residents, making adjustments for observed differences in patient risk factors. It also adjusts for changes in outcomes over time (trends) that were common to all hospitals. This design prevents bias from 3 possible sources. First, a difference between hospitals that is stable over time cannot be mistaken for an effect of the reform, because each hospital is compared with itself, before and

after reform. Because of this, hospital indicators for fixed effects were used in the logistic model. Second, changes over time that affect all hospitals similarly (eg, technological improvements) cannot be mistaken for an effect of the reform. Because of this, year indicators were used in the logistic model. Third, if the mix of patients is changing in different ways at different hospitals, and if these changes are accurately reflected in measured risk factors, this cannot be mistaken for an effect of the reform because the logistic model adjusts for these measured risk factors.

Although the method of difference-in-differences offers these advantages, it has limitations. Any diverging trend in mortality over time for more vs less teaching-intensive hospitals that was already in progress or coincident with the initiation of the reform could be mistaken for an effect of the reform, although we tested extensively whether the prereform trends were similar in more and less teaching-intensive hospitals and adjusted for any observed underlying dif-

ference in prereform trends. Less teaching-intensive hospitals, including all non-teaching hospitals, served as the primary control group for more teaching-intensive hospitals because they were less affected by the duty hour reform but were subject to the same technological quality improvement imperatives, changes in market conditions, and Medicare-specific initiatives such as pay for performance. In addition, they are geographically diverse with large patient populations, and similar patient discharge data are available. Data from July 1, 2000, to June 30, 2003, were used as the prereform period and data from July 1, 2003, to June 30, 2005, were used as the postreform period.

The dependent variable was death within 30 days of hospital admission, using logistic regression to adjust for patient comorbidities, secular trends common to all patients (eg, due to general changes in technology), and hospital site where treated. The effect of the change in duty hour rules was measured as the coefficients of resident-to-

bed ratio interacted with dummy variables indicating postreform year 1 and postreform year 2. These coefficients, presented as odds ratios (ORs), measure the degree to which mortality changed in more vs less teaching-intensive hospitals after adjusting for cross-sectional differences in hospital quality and general improvements in care. They were measured for each year separately because of the possibility of either delayed beneficial effects or early harmful effects. Conditions were assessed both individually and together as combined medical and combined surgical groups.

In the models, baseline mortality levels were allowed to differ between more and less teaching-intensive hospitals and were assumed to have a common time trend until implementation of the duty hour rules, after which the teaching hospital trend was allowed to diverge. To assess whether underlying trends in risk-adjusted mortality were similar in higher and lower teaching-intensity hospitals during the 3 years before the ACGME duty hour reform, a test of controls was performed. Parameters were added to the model for interactions between the resident-to-bed ratio and indicators for prereform year 2 and prereform year 1. A Wald χ^2 test was used to determine whether these interactions were equal to 0. A statistically significant test of controls suggested that teaching and nonteaching hospitals had a diverging trend in mortality in the 3 years prereform that could not have been caused by the reform. When such diverging trends were found by the test of controls, post hoc analyses were conducted in which postreform results were compared with the prereform year 1 as a baseline rather than using data from the entire 3-year prereform period.

To provide examples of the effect of being in hospitals with different degrees of teaching intensity, we converted the regression coefficients into estimated probabilities of mortality for an average patient by using the mean values for each of the covariates and replacing hospital indicators by the resident-to-bed ratio; the examples compared hospitals with

Table 1. Characteristics of the Study Population^a

	Prereform Year			Postreform Year	
	3 (2000-2001)	2 (2001-2002)	1 (2002-2003)	1 (2003-2004)	2 (2004-2005)
Medical conditions					
Acute myocardial infarction					
No. of cases	213 500	198 873	194 158	188 164	175 489
Unadjusted mortality, %	16.3	15.6	14.7	14.4	13.8
Stroke					
No. of cases	206 417	192 314	184 568	179 077	170 849
Unadjusted mortality, %	17.7	17.5	17.4	17.5	17.4
Gastrointestinal bleeding					
No. of cases	160 397	152 358	150 401	153 797	146 812
Unadjusted mortality, %	6.7	6.5	6.3	6.1	5.8
Congestive heart failure					
No. of cases	262 086	242 531	236 757	232 880	222 040
Unadjusted mortality, %	8.5	8.6	8.4	8.5	8.4
Surgical categories					
General surgery					
No. of cases	359 250	345 337	340 770	338 585	323 345
Unadjusted mortality, %	5.2	5.2	5.2	5.2	5.0
Orthopedic surgery					
No. of cases	483 848	488 657	511 002	542 320	552 363
Unadjusted mortality, %	2.5	2.5	2.4	2.3	2.1
Vascular surgery					
No. of cases	83 445	78 943	75 942	73 093	69 227
Unadjusted mortality, %	13.0	13.1	12.8	12.5	12.1

^aIncluded 8 529 595 patients from 3321 hospitals. Prereform and postreform years are academic years (eg, prereform year 3 indicates academic year July 1, 2000, to June 30, 2001).

resident-to-bed ratios of 1 and 0. We tested the stability of the medical and surgical results by (1) eliminating patients admitted to hospitals in New York State, due to earlier passage of the Libby Zion laws^{20,21}; (2) eliminating patients admitted from nursing homes, because such patients may not have been treated aggressively; (3) testing the robustness of the results to analysis without comorbidity adjustment to determine whether changes in the rate of coded comorbidities could explain any of these effects; and (4) examining the degree to which mortality changed in patients with a second CHF admission within 6 months of the first. All *P* values were either 2-tailed or, for χ^2 tests, multitailed. *P* < .05 was considered statistically significant. All statistical analyses were conducted with SAS version 9.1 (SAS Institute Inc, Cary, North Carolina).

RESULTS

The number of admissions for each of the conditions was fairly constant over time, differing by less than 10% per year for any condition (TABLE 1). Approximately 69% of the hospitals were non-

teaching hospitals and approximately 9.2% of hospitals (treating 14.0% of patients) were major or very major teaching hospitals (TABLE 2).

Between prereform year 2 and prereform year 1, unadjusted mortality rates for the combined medical group increased in the very major teaching-intensive hospitals more than in other hospitals; however, mortality subsequently changed at similar rates from prereform year 1 to postreform year 2 (FIGURE 1). The test for controls for stroke indicated that in the 3 years before the reform, stroke mortality had diverged significantly in more vs less teaching-intensive hospitals (Wald χ^2 test, 12.0; *P* = .003); therefore, post hoc

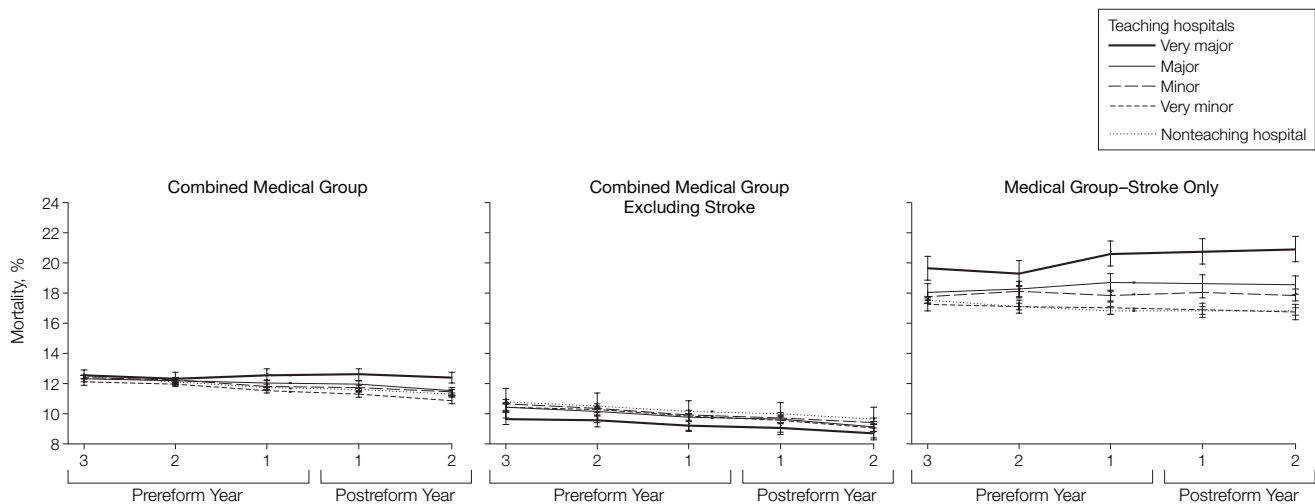
analyses were conducted that examined the associations without stroke separately. Unadjusted mortality for the combined medical group excluding stroke changed at similar rates for patients in hospitals of differing teaching intensity between 2000 and 2005 (Figure 1). This indicates that the prereform divergence was largely due to the changes in stroke mortality. Compared with nonteaching hospitals, hospitals with a resident-to-bed ratio of 1 had no change in the unadjusted odds of mortality for combined medical conditions in postreform year 1 (OR, 1.03; 95% confidence interval [CI], 0.99-1.08) or postreform year 2 (OR, 1.04; 95% CI, 0.99-1.09).

Table 2. Characteristics of Included Hospitals^a

	Resident-to-Bed Ratio, Category (Range)				
	Nonteaching (0)	Very Minor Teaching (>0-0.049)	Minor Teaching (0.050-0.249)	Major Teaching (0.250-0.599)	Very Major Teaching (0.600-1.090)
No. (% of facilities)	2298 (69.2)	308 (9.3)	411 (12.45)	195 (5.9)	109 (3.3)
No. (% of admissions)	4 313 981 (50.6)	1 231 262 (14.4)	1 791 242 (21.0)	797 822 (9.4)	395 288 (4.6)

^aPercentages may not total 100 due to rounding. Included 3321 facilities and 8 529 595 admissions.

Figure 1. Unadjusted Trends in Mortality for Combined Medical Groups by Teaching-Intensive Hospital Status



Error bars indicate 95% confidence intervals, calculated using the binomial distribution.⁴⁷ The Accreditation Council for Graduate Medical Education duty hour regulations were implemented on July 1, 2003. Prereform year 3 included academic year 2000-2001 (July 1, 2000, to June 30, 2001); prereform year 2, academic year 2001-2002; prereform year 1, academic year 2002-2003; postreform year 1, academic year 2003-2004; and postreform year 2, academic year 2004-2005. For combined medical group, a significant divergence was found before the onset of the duty hour reform (by Wald χ^2 test, *P* = .04), due to patterns observed for patients with stroke. No significant divergence was found in the degree to which mortality changed from prereform year 1 to either postreform year for combined medical group or combined medical group excluding stroke. Significance levels assess whether trend from prereform year 1 to postreform years 1 and 2, respectively, differed for more vs less teaching-intensive hospitals.

Among surgical patients, there was a relative increase in unadjusted mortality among patients in more teaching-intensive hospitals from prereform year 3 to prereform year 1, following which mortality in all hospitals declined to a similar degree (FIGURE 2). Compared with nonteaching hospitals, hospitals with a resident-to-bed ratio of 1 had no change in the unadjusted odds of mortality for combined surgical categories

in postreform year 1 (OR, 1.04; 95% CI, 0.97-1.11) or postreform year 2 (OR, 0.99; 95% CI, 0.93-1.07).

In adjusted analyses of both the combined medical and combined surgical groups, there was no evidence of relative increases or decreases in the odds of mortality for patients in more vs less teaching-intensive hospitals in either postreform year 1 or postreform year 2 (TABLE 3). Results remained

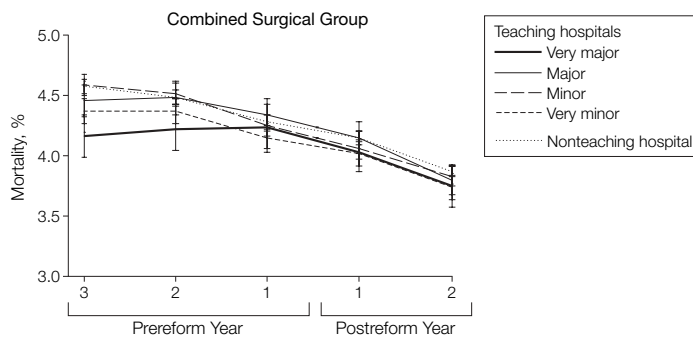
nonsignificant when the 2 postreform years were combined into a single resident-to-bed ratio \times postreform coefficient.

None of the individual conditions other than stroke showed any significant relative changes in mortality in more vs less teaching-intensive hospitals in either postreform year 1 or postreform year 2. There was a statistically significant increase in mortality for patients with stroke in more teaching-intensive hospitals in postreform year 1 (OR, 1.08; 95% CI, 1.00-1.16; $P=.04$) and postreform year 2 (OR, 1.08; 95% CI, 1.01-1.16; $P=.03$) relative to less teaching-intensive hospitals (Table 3). Because of the prereform divergence for stroke mortality, an analysis was conducted in which the referent group was changed from the entire prereform period to prereform year 1 only. This indicated no relative increase in mortality, with the OR for postreform year 1 of 0.98 (95% CI, 0.90-1.08) and the OR for postreform year 2 of 0.99 (95% CI, 0.91-1.08). Analysis of the combined medical group excluding stroke showed no significant association of changes in mortality with teaching status (Table 3).

The test of controls was consistent with divergent prereform trends for the stroke ($P=.003$), combined medical conditions ($P=.048$), general surgery ($P=.01$), and combined surgical categories groups ($P=.02$). Except for stroke, there was no significant differential change in the odds of mortality in more vs less teaching-intensive hospitals after adjusting for the differential prereform trends for combined medical conditions. Using prereform year 1 as the referent period for each of these groups, there were no significant associations between the degree of change in mortality and teaching intensity between prereform year 1 and either postreform year 1 or postreform year 2.

Excluding patients admitted to hospitals in New York State or patients admitted from nursing homes from the statistical analyses did not change the results. We also examined whether changes in the coding of comorbidi-

Figure 2. Unadjusted Trends in Mortality for Combined Surgical Group by Teaching-Intensive Hospital Status



Error bars indicate 95% confidence intervals, calculated using the binomial distribution.⁴⁷ The Accreditation Council for Graduate Medical Education duty hour regulations were implemented on July 1, 2003. Prereform year 3 included academic year 2000-2001 (July 1, 2000, to June 30, 2001); prereform year 2, academic year 2001-2002; prereform year 1, academic year 2002-2003; postreform year 1, academic year 2003-2004; and postreform year 2, academic year 2004-2005. No significant divergence was found between prereform year 1 and either postreform year. Significance levels assess whether trend from prereform year 1 to postreform years 1 and 2, respectively, differed for more vs less teaching-intensive hospitals.

Table 3. Adjusted Odds of Mortality After Duty Hour Reform in More vs Less Teaching-Intensive Hospitals

Patient Categories	Total No. of Cases	Resident-to-Bed Ratio \times Postreform Year 1 ^a		Resident-to-Bed Ratio \times Postreform Year 2 ^a	
		Odds Ratio (95% CI) ^b	P Value	Odds Ratio (95% CI) ^b	P Value
Medical conditions					
Acute myocardial infarction	970 184	0.97 (0.89-1.05)	.46	0.93 (0.86-1.02)	.12
Stroke	933 225	1.08 (1.00-1.16)	.04	1.08 (1.01-1.16)	.03
Gastrointestinal bleeding	763 765	1.12 (0.97-1.30)	.11	0.94 (0.81-1.10)	.44
Congestive heart failure	1 196 294	0.90 (0.80-1.00)	.054	1.03 (0.92-1.14)	.65
Combined medical	3 863 468	1.03 (0.98-1.07)	.30	1.03 (0.99-1.08)	.18
Combined medical, excluding stroke	2 930 243	0.97 (0.91-1.03)	.27	0.95 (0.90-1.01)	.13
Surgical categories					
General surgery	1 707 287	1.04 (0.95-1.15)	.39	0.97 (0.88-1.07)	.58
Orthopedic surgery	2 578 190	1.12 (0.99-1.27)	.07	1.04 (0.92-1.18)	.52
Vascular surgery	380 650	1.00 (0.89-1.13)	.96	1.02 (0.90-1.14)	.81
Combined surgical	4 666 127	1.05 (0.98-1.12)	.15	1.01 (0.95-1.08)	.67

Abbreviation: CI, confidence interval.

^aThe interaction terms (resident-to-bed ratio \times postreform year 1) and (resident-to-bed ratio \times postreform year 2) measure whether there is any relative change in the odds of mortality in more vs less teaching-intensive hospitals. Models are also adjusted for age, sex, comorbidities, common time trends, and hospital site where treated.

^bUnits for odds ratios compare hospitals with a resident-to-bed ratio of 1 with hospitals with a resident-to-bed ratio of 0 (nonteaching).

ties could explain any of these effects. There was only a 0.7% relative decrease in the mean number of comorbidities in more teaching-intensive hospitals relative to nonteaching hospitals from prereform year 1 to postreform year 2, and sensitivity analyses without adjustment for comorbidities produced similar results. For example, the OR for combined medical conditions without risk adjustment was 1.03 (95% CI, 0.99-1.08) vs 1.03 (95% CI, 0.99-1.08) with risk adjustment. There also were no significant differences in the degree to which mortality changed in more vs less teaching-intensive hospitals for patients with CHF who had a second admission within 6 months of the first, in patients with AMI when additional adjustments for anatomic location were added, or in patients with stroke when adjustment for the proportion of hemorrhagic strokes was added.

To illustrate the magnitude of the changes in mortality associated with duty hour reform, we estimated the adjusted risk of mortality in each academic year for a hypothetical person with mean values of all regression covariates at 2 hypothetical hospitals: a nonteaching hospital (resident-to-bed ratio of 0) and a highly teaching-intensive hospital (resident-to-bed ratio of 1). In the combined medical group, mortality in a hospital with a resident-to-bed ratio of 1 did not change (9.06% in prereform year 1 vs 9.06% in postreform year 2), while in a hospital with a resident-to-bed ratio of 0 mortality decreased from 9.48% to 9.06%, a comparative increase of 0.42 percentage points (4.4% relative increase). For a patient representing combined surgical categories, mortality in a hospital with a resident-to-bed ratio of 1 decreased from 1.67% in prereform year 1 to 1.56% in postreform year 2, while in a hospital with a resident-to-bed ratio of 0 mortality decreased from 2.10% to 1.94%, a comparative increase of 0.05 percentage points (2.3% relative increase). None of these changes were statistically significant.

COMMENT

Implementation of the ACGME duty hour rules was arguably one of the largest efforts ever undertaken to reduce errors in teaching hospitals. The results of this study suggest that after 2 years it has not been associated with a significant positive or negative change in mortality among Medicare patients. These results should be considered in the context of 3 other studies.

First, the Healthcare Cost and Utilization Project's (HCUP's) Nationwide Inpatient Sample (NIS) was the data source for a study by Shetty and Bhattacharya²³ that found a small but statistically significant relative improvement in mortality outcomes for medical but not surgical patients in teaching compared with nonteaching hospitals following duty hour reform. However, there are a number of potential limitations in these data. The HCUP NIS samples different hospitals in different years, as well as hospitals from different states in different years⁴⁸; therefore, analyses using these data could be biased by changes in the distribution of sampled patients across hospitals between the prereform and postreform periods. The study by Shetty and Bhattacharya²³ was also limited to the 551 hospitals included in the NIS both before and after reform. In contrast, our study, which includes Medicare data on 3321 hospitals, incorporates data from all hospitals in all years, allowing for more reliable testing for differences in outcomes before and after reform within the same hospital. There are no patient identifiers in the HCUP NIS, so that the data cannot distinguish between a single admission from different patients vs multiple admissions from the same patient. Inclusion of multiple admissions could make it appear that duty hour reform reduced mortality if the percentage of patients with multiple admissions was higher at nonteaching hospitals than at teaching hospitals. Because the HCUP NIS does not include information on out-of-hospital deaths, it was necessary for Shetty and Bhattacharya to use in-hospital mortality as the primary out-

come. In-hospital mortality can change without any change in total mortality rates if patients are discharged faster or slower; discharge rates may have been influenced by the duty hour reform. Our study used 30-day (from admission) mortality, which is not affected by discharge patterns.

Second, a single-site study by Horwitz et al⁴⁹ compared changes in outcomes on the teaching vs nonteaching service before and after duty hour reform. It found no changes in length of stay, 30-day readmission rates, or adverse drug-drug interactions but did not have sufficient power to compare changes in mortality.

Third, we applied methods similar to those used in the present study to the population of patients receiving care in the VA health care system.²⁴ That study may be considered complementary to the present study because in aggregate VA hospitals are the largest site of residency training in the United States, whereas the vast majority of US patients are treated in non-VA hospitals. The VA study found improved mortality for medical patients in more teaching-intensive environments in postreform year 2, although among surgical patients, there were no relative changes for patients in more vs less teaching-intensive hospitals. The finding of reduced mortality for VA medical patients but not Medicare medical patients could reflect that the mean resident-to-bed ratio among VA teaching hospitals is much higher than in non-VA teaching hospitals and that residents may have more autonomy at VA hospitals,^{50,51} both of which could amplify positive or negative effects of the duty hour rules. Resident staffing models and clinical volume at VA hospitals may be different from non-VA hospitals of similar teaching intensity, which could enable higher rates of compliance with duty hour rules at VA hospitals. The lack of significant improvement in mortality among Medicare medical patients in higher teaching-intensity hospitals could reflect a different balance between the effects of decreased fatigue^{18,19,22} and worsened

continuity^{21,52} compared with VA hospitals. Observational studies may always have unmeasured confounders, which could have existed to a greater or lesser degree in the VA analysis than in the Medicare analysis.

Although we found a relative increase in mortality in patients with stroke in higher teaching-intensity hospitals postreform, this increase was not present when the referent group was prereform year 1. Because a relative increase occurred before reform, it is likely that this finding reflects a trend that was independent of duty hour changes. The reasons for this prior trend are not clear.

Coding of comorbidities changed at similar rates in all hospitals over time, with a relative decrease of 0.7% in the mean number of comorbidities in the most teaching-intensive hospitals relative to nonteaching hospitals in postreform year 2. The similarity of our results with and without comorbidity adjustment suggests that there were no major changes in the patient profiles over time for hospitals of different teaching intensity, that our results would likely be robust to the specific method of comorbidity adjustment chosen, and that changes in coding frequency do not explain the measured effects.

The absence of a relative change in mortality among Medicare beneficiaries in more teaching-intensive hospitals may reflect many factors. Lack of compliance with the duty hour rules has been described, although such surveys have been limited to self-reported compliance based on a potentially nonrepresentative sample of about 7% of all interns⁵³ or data collected by the ACGME.⁵⁴ There are strong financial incentives for residency programs to comply to avoid loss of accreditation. Even if programs are not in complete compliance, it seems likely that there has been a reduction in work hours that would be greatest in the subspecialties (eg, surgery) in which residents worked the most hours prereform. However, unless the appropriate resources were put in place at the institutional level to mitigate the impact of residents working

fewer hours, the effect of duty hour reform at the local level would likely be to increase work intensity, which could offset potential benefits of decreased fatigue. Investigators in New York State examined the effects of the Libby Zion laws on mortality but found no differences in mortality in teaching vs nonteaching hospitals²⁰ or within the same teaching hospital.²¹ However, compliance rates were extremely low,^{20,55} and thus the applicability of these studies to the ACGME duty hour rules is unclear.

There are several possible limitations to our study. We focused on one outcome, mortality. Although the duty hour rules were an attempt to reduce deaths from medical errors, measurement of other outcomes such as patient safety indicators may help to explain the relative effects of decreased continuity of care compared with decreased resident fatigue.

We do not have information on actual hours worked at each hospital. Our study may be considered an effectiveness study, in contrast with an efficacy study, because we measured the outcomes associated with the duty hour rules as implemented. Even with the size of the Medicare population, some of the CIs are still somewhat broad and we cannot rule out small and possibly clinically meaningful effects. This is particularly true for the individual condition analyses and for surgical patients because mortality was lower for surgical patients than for medical patients.

Any observational study is susceptible to unmeasured confounding. We used administrative data, so risk adjustment is more limited than with clinical data; however, by comparing outcomes over time within each hospital in more vs less teaching-intensive hospitals, potential bias from unmeasured confounders is markedly diminished. Given the multiple time series difference-in-differences approach, to be a confounder a variable must have a differential rate of change in more vs less teaching-intensive hospitals. Because we adjusted for different baseline rates of change in more and less teaching-

intensive hospitals, confounding would only occur if the prevalence of unmeasured risk factors changed at different rates according to hospitals' teaching intensity.

In conclusion, we found that implementation of duty hours limitations was not associated with any significant change in risk-adjusted mortality among Medicare patients. These results do not address whether the current design of duty hour rules is optimal,¹³ as other work has found significantly lower rates of errors with 16-hour vs 24-hour to 36-hour shifts.^{18,19} Given the lack of evidence of improvements in outcomes in this study, research should focus on examining different approaches to duty hour design as well as measuring resident work intensity and clinically relevant patient outcomes in addition to mortality.

Author Affiliations: Center for Health Equity Research and Promotion, Veterans Administration Hospital, Philadelphia, Pennsylvania (Dr Volpp); Center for Outcomes Research, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Dr Silber and Mss Even-Shoshan and Wang); Departments of Medicine (Drs Volpp and Bellini, and Ms Behringer) and Pediatrics and Anesthesiology and Critical Care (Dr Silber), University of Pennsylvania School of Medicine, Philadelphia; Departments of Health Care Systems (Drs Volpp and Silber) and Statistics (Dr Rosenbaum), The Wharton School, University of Pennsylvania, Philadelphia; The Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia (Drs Volpp and Silber, and Ms Even-Shoshan); Department of Health Policy and Management, Boston University School of Public Health, Boston, Massachusetts, and Center for Health Quality, Outcomes and Economic Research, Veterans Administration Hospital, Bedford, Massachusetts (Dr Rosen); and Division of General Medicine and Center for Healthcare Policy and Research, University of California Davis School of Medicine, Sacramento (Dr Romano).

Author Contributions: Drs Volpp and Silber had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Volpp, Rosen, Rosenbaum, Romano, Bellini, Silber.

Acquisition of data: Volpp, Even-Shoshan, Behringer, Silber.

Analysis and interpretation of data: Volpp, Rosen, Rosenbaum, Romano, Wang, Bellini, Silber.

Drafting of the manuscript: Volpp, Rosenbaum, Silber.

Critical revision of the manuscript for important intellectual content: Volpp, Rosen, Rosenbaum, Romano, Even-Shoshan, Wang, Bellini, Behringer, Silber.

Statistical analysis: Volpp, Rosenbaum, Romano, Wang, Silber.

Obtained funding: Volpp, Rosen, Rosenbaum, Even-Shoshan, Silber.

Administrative, technical, or material support: Volpp, Even-Shoshan, Bellini, Behringer, Silber.

Study supervision: Volpp, Rosen, Rosenbaum, Silber.

Financial Disclosures: None reported.

Funding/Support: This work was supported primarily by grant R01 HL082637 from the National Heart, Lung, and Blood Institute. Additional support was received by grant IIR 04.202.1 from the US Veterans Affairs Health Services Research and Development Service and grant SES-0646002 from the National Science Foundation.

Role of the Sponsors: The sponsors had no role in the design and conduct of the study, in the collection, management, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

Additional Contributions: Hong Zhou, MS, Children's Hospital of Philadelphia, provided assistance in preparing the Medicare data for analysis and Michael Halenar, BA, University of Pennsylvania, provided assistance as a research assistant. Ms Zhou and Mr Halenar received financial compensation for their contributions.

REFERENCES

- Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
- Accreditation Council for Graduate Medical Education Web site. Resident duty hours language: final requirements. <http://www.acgme.org>. Accessed July 27, 2003.
- Accreditation Council for Graduate Medical Education Web site. Report of the work group on resident duty hours and the learning environment. June 11, 2002. In: The ACGME's approach to limit resident duty hours 12 months after implementation: a summary of achievements, 2004. http://www.acgme.org/acWebsite/dutyHours/dh_achieveSum04-05.pdf. Accessed October 3, 2006.
- Buysse DJ, Barzansky B, Dinges D, et al. Sleep, fatigue, and medical training: setting an agenda for optimal learning and patient care. *Sleep*. 2003;26(2):218-225.
- Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. *Semin Neurol*. 2005;25(1):117-129.
- Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*. 1997;20(4):267-277.
- Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*. 2003;26(2):117-126.
- Dinges DF, Baynard M, Rogers NL. Chronic sleep deprivation. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 4th ed. Philadelphia, PA: WB Saunders; 2005:67-76.
- Williamson AM, Feyer AM. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occup Environ Med*. 2000;57(10):649-665.
- Fletcher KE, Davis SQ, Underwood W, Mangrulkar RS, McMahon LF Jr, Saint S. Systematic review: effects of resident work hours on patient safety. *Ann Intern Med*. 2004;141(11):851-857.
- Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. *JAMA*. 2002;287(8):955-957.
- Charap M. Reducing resident work hours: unproven assumptions and unforeseen outcomes. *Ann Intern Med*. 2004;140(10):814-815.
- Pellegrini VD Jr, Peabody T, Dinges DF, Moody J, Fabri PJ. Symposium resident work-hour guidelines: a sentence or an opportunity for orthopaedic education? *J Bone Joint Surg Am*. 2005;87(11):2576-2586.
- Drzen JM. Awake and informed. *N Engl J Med*. 2004;351(18):1884.
- Mukherjee S. A precarious exchange. *N Engl J Med*. 2004;351(18):1822-1824.
- Ofri D. Residency regulations: resisting our reflexes. *N Engl J Med*. 2004;351(18):1824-1826.
- Gaba DM, Howard SK. Patient safety: fatigue among clinicians and the safety of patients. *N Engl J Med*. 2002;347(16):1249-1255.
- Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med*. 2004;351(18):1838-1848.
- Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. *N Engl J Med*. 2004;351(18):1829-1837.
- Howard DL, Silber JH, Jobes DR. Do regulations limiting residents' work hours affect patient mortality? *J Gen Intern Med*. 2004;19(1):1-7.
- Laine C, Goldman L, Soukup JR, Hayes JG. The impact of a regulation restricting medical house staff working hours on the quality of patient care. *JAMA*. 1993;269(3):374-378.
- Poulose BK, Ray WA, Arbogast PG, et al. Resident work hour limits and patient safety. *Ann Surg*. 2005;241(6):847-856.
- Shetty KD, Bhattacharya J. Changes in hospital mortality associated with residency work-hour regulations [published online ahead of print June 4, 2007]. *Ann Intern Med*. 2007;147(2):73-80.
- Volpp KG, Rosen AK, Rosenbaum P, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. *JAMA*. 2007;298(9):984-992.
- Iezzoni LI, ed. *Risk Adjustment for Measuring Health-care Outcomes*. 3rd ed. Chicago, IL: Health Administration Press; 2003:73.
- Rockall TA, Logan RF, Devlin HB, Northfield TC. Variation in outcome after acute upper gastrointestinal haemorrhage: the National Audit of Acute Upper Gastrointestinal Haemorrhage. *Lancet*. 1995;346(8971):346-350.
- Inpatient Quality Indicators Overview. AHRQ Quality Indicators Web site. http://www.qualityindicators.ahrq.gov/iqi_overview.htm. Accessed May 1, 2007.
- Second Report of the California Hospitals Outcomes Project, May 1996. *Acute Myocardial Infarction*. Sacramento, CA: Office of Statewide Health Planning and Development; 1996. <http://repositories.cdlib.org/chsrpc/coshpd/index.html#1996>. Accessed May 1, 2007.
- Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
- Glance LG, Dick AW, Osler TM, Mukamel DB. Does date stamping ICD-9-CM codes increase the value of clinical information in administrative data? *Health Serv Res*. 2006;41(1):231-251.
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130-1139.
- Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care*. 2004;42(4):355-360.
- Stukenborg GJ, Wagner DP, Connors AF Jr. Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. *Med Care*. 2001;39(7):727-739.
- Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613-619.
- Rosenbaum PR. Choice as an alternative to control in observational studies. *Stat Sci*. 1999;14(3):259-278.
- Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of in-hospital complications from claims data: is it valid? *Med Care*. 2000;38(8):785-795.
- Keeler EB, Rubenstein LV, Kahn KL, et al. Hospital characteristics and quality of care. *JAMA*. 1992;268(13):1709-1714.
- Allison JJ, Kiefe CI, Weissman NW, et al. Relationship of hospital teaching status with quality of care and mortality for Medicare patients with acute MI. *JAMA*. 2000;284(10):1256-1262.
- Taylor DH, Whellan DJ, Sloan FA. Effects of admission to a teaching hospital on the cost and quality of care for Medicare beneficiaries. *N Engl J Med*. 1999;340(4):293-299.
- Ayanian JZ, Weissman JS. Teaching hospitals and quality of care: a review of the literature. *Milbank Q*. 2002;80(3):569-593.
- Centers for Medicare & Medicaid Services. Indirect Medical Education (IME). http://www.cms.hhs.gov/AcuteInpatientPPS/07_ime.asp. Accessed June 29, 2007.
- Cox DR. Note on grouping. *J Am Stat Assoc*. 1957;52(280):543-547.
- Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 1968;24(2):295-313.
- Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Dallas, TX: Houghton Mifflin Co; 1963:56-63.
- Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin; 2002:181.
- Rosenbaum PR. Stability in the absence of treatment. *J Am Stat Assoc*. 2001;96(453):210-219.
- Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. New York, NY: Wiley; 2003:28-29.
- Introduction to the HCUP Nationwide Inpatient Sample 2005. Healthcare Cost and Utilization Project (HCUP). <http://www.hcup-us.ahrq.gov/nisoverview.jsp>. Accessed June 28, 2007.
- Horwitz LI, Kosiborod M, Lin Z, Krumholz HM. Changes in outcomes for internal medicine inpatients after work-hour regulations [published online ahead of print June 4, 2007]. *Ann Intern Med*. 2007;147(2):97-103.
- Department of Veterans Affairs. *Veterans Health Administration Handbook 1400.1: Resident Supervision*. Washington, DC: Veterans Health Administration; 2005. http://www1.va.gov/vhapublications/ViewPublication.asp?pub_ID=1289. Accessed May 19, 2007.
- Centers for Medicare and Medicaid Services. Guidelines for teaching physicians, interns, and residents. July 2006. <http://www.cms.hhs.gov/MLNProducts/downloads/gdelinesteachgresfctst.pdf>. Accessed May 19, 2007.
- Petersen LA, Brennan TA, O'Neil AC, Cook EF, Lee TH. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med*. 1994;121(11):866-872.
- Landrigan CP, Barger LK, Cade BE, Ayas NT, Czeisler CA. Interns' compliance with accreditation council for graduate medical education work-hour limits. *JAMA*. 2006;296(9):1063-1070.
- Accreditation Council for Graduate Medical Education. The ACGME's approach to limit resident duty hours 12 months after implementation: a summary of achievements, 2004. http://www.acgme.org/acWebsite/dutyHours/dh_achieveSum04-05.pdf. Accessed October 3, 2006.
- DeBuono BA, Osten WM. The medical resident workload: the case of New York State. *JAMA*. 1998;280(21):1882-1883.