

Mortality Among Patients in VA Hospitals 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

Anne Canamucio, MS

Lisa Bellini, MD

Tiffany Behringer, MS

Jeffrey H. Silber, MD, PhD

THE ACCREDITATION COUNCIL for Graduate Medical Education (ACGME) implemented duty hour regulations on July 1, 2003, for all ACGME-accredited residency programs,^{1,2} following concern about deaths associated with medical errors in US hospitals.³ These regulations limited the number of work hours per week, the number of continuous work hours, and the frequency of in-house call, and provided for a minimum amount of time between duty periods and days free of work duties.¹

In the absence of much empirical data supporting the duty hour regulations,^{4,5} the design has been controversial.⁶⁻⁸ There has been concern about effects on continuity of care that could offset any beneficial effects of reduced fatigue on mortality.⁹⁻¹¹ Most studies of duty hour reforms have not examined their effect on quality of care and relied exclusively on surveys and resident self-reports of the impact of the reforms.^{12,13} Recent studies that have

Context Limitations in duty hours for physicians-in-training in the United States were established by the Accreditation Council for Graduate Medical Education (ACGME) and implemented on July 1, 2003. The association of these changes with mortality among hospitalized patients has not been well established.

Objective To determine whether the change in duty hour regulations was associated with relative changes in mortality in hospitals of different teaching intensity within the US Veterans Affairs (VA) system.

Design, Setting, and Patients An observational study of all unique patients (N=318 636) admitted to acute-care VA hospitals (N=131) using interrupted time series analysis with data from July 1, 2000, to June 30, 2005. All patients had principal diagnoses of acute myocardial infarction (AMI), 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 postreform year 1, no significant relative changes in mortality were observed for either medical or surgical patients. In postreform year 2, the odds of mortality decreased significantly in more teaching-intensive hospitals for medical patients only. Comparing a hospital having a resident-to-bed ratio of 1 with a hospital having a resident-to-bed ratio of 0, the odds of mortality were reduced for patients with AMI (odds ratio [OR], 0.48; 95% confidence interval [CI], 0.33-0.71), for the 4 medical conditions together (OR, 0.74; 95% CI, 0.61-0.89), and for the 3 medical conditions excluding AMI (OR, 0.79; 95% CI, 0.63-0.98). Compared with hospitals in the 25th percentile of teaching intensity, there was an absolute improvement in mortality from prereform year 1 to postreform year 2 of 0.70 percentage points (11.1% relative decrease) and 0.88 percentage points (13.9% relative decrease) in hospitals in the 75th and 90th percentile of teaching intensity, respectively, for the combined medical conditions.

Conclusions The ACGME duty hour reform was associated with significant relative improvement in mortality for patients with 4 common medical conditions in more teaching-intensive VA hospitals in postreform year 2. No associations were identified for surgical patients.

JAMA. 2007;298(9):984-992

www.jama.com

examined the association of changes in mortality and teaching status following the duty hour regulation have either been underpowered¹⁴ or have used data sources that may limit the validity.¹⁵

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 975 and 1055.

We therefore studied the association between changes in the ACGME duty hour rules and mortality rates. Trends in risk-adjusted mortality rates among more vs less teaching-intensive hospitals were compared to assess whether mortality improved differentially among these groups following implementation of the rules. We used data on patients with a broad range of clinical conditions hospitalized within the US Veterans Affairs (VA) health care system, the single largest provider of residency training in the United States. An accompanying article reports a complementary analysis among all US hospitalized Medicare patients.¹⁶

METHODS

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

Main Outcome Measures

The main outcome measure was death within 30 days of hospital admission for all patients admitted with diagnoses of acute myocardial infarction (AMI), stroke, gastrointestinal bleeding, congestive heart failure (CHF), general surgery, orthopedic surgery, or vascular surgery. These medical conditions were a subset of the Agency for Healthcare Research and Quality (AHRQ) Quality Indicators for which mortality was a relevant outcome; for these conditions, there is evidence that mortality varies substantially across institutions and that high mortality may be associated with deficiencies in the quality of care.¹⁷⁻¹⁹ Although the AHRQ Quality Indicators use only in-hospital mortality, we examined any in-hospital or postdischarge deaths within 30 days of hospital admission to eliminate bias due to length of stay differences across hospitals or time.²⁰

Study Sample

All patients admitted to acute-care VA 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 459 321 admissions from 132 hospitals, which contributed data for all 5 years. Admissions to hospitals outside of the 50 states or Washington, DC (n=8524), transfers from a non-VA hospital (n=6337), transfers subsequent to a contiguous qualifying admission (to avoid double counting admissions within 30 days) (n=4103), admissions spanning July 1, 2003 (n=1729), or admissions with dates of death earlier than their discharge dates (n=2) were excluded, as were those admissions for patients older than 90 years (n=2388) because the proportion of such patients that are treated aggressively may change over time in ways that cannot be observed well with administrative data. Among patients with AMI and stroke, those discharged alive in fewer than 2 days (n=5231) were excluded because such cases may not represent actual AMIs or strokes.¹⁸ These exclusions resulted in data from 431 007 patients from 131 hospitals.

An index admission was defined as the first eligible admission between July 1, 2000, and June 30, 2005, for which there was no prior admission for the medical conditions or surgical categories within 5 years (using data back to July 1, 1995). This ensured that each patient would only be represented once within each analysis. The first admission in the past 5 years for each patient was chosen to eliminate the possibility of selecting cases for which there would be a higher mortality rate postreform for reasons other than duty hour reform. For patients with multiple admissions, any admission before the last would be an admission they had survived. Inclusion of multiple admissions would therefore lead to a confounding bias due to the passage of time, in which admissions that resulted in death would be more likely to happen postreform.

Using index admissions, the available sample was 318 636 patients from

131 hospitals. More than 90% of patients with each condition had a first admission that met these criteria except for patients with CHF, for whom this was approximately 88%. For this reason, we tested the stability of this approach in patients with multiple admissions by assessing the associations with duty hour reform for patients with CHF who experienced a second or third admission within 6 months of the first admission.

Risk Adjustment and Hospital Control Measures

Risk adjustment was performed according to the Elixhauser method,²¹⁻²³ which included the original 29 comorbidities except for fluid and electrolyte disorders or coagulopathy.^{24,25} Analyses were also adjusted for age and sex. For patients with AMI, we tested the sensitivity of including 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 surgical patients, we also adjusted for diagnosis related groups that were aggregated to include related groups with and without complications or comorbidities. We performed a 180-day look-back, in which data on secondary diagnoses recorded in hospitalizations within 180 days of the index hospitalization were assessed to obtain more comprehensive information on comorbidities than available using the index admission alone.²⁶

Data

Data on patient characteristics were obtained from the VA patient treatment file, which includes information on principal and secondary diagnoses, age, sex, and discharge disposition. Mortality data were obtained from the patient treatment file for in-hospital deaths and from the VA beneficiary identification and record locator subsystem file for out-of-hospital deaths.²⁷⁻²⁹ The VA hospital characteristics were obtained from the Veterans Health Administration Account Level Budgeter Cost Cen-

ter, an administrative database containing nurse staffing information, and from the Veterans Health Administration Support Service Center Occupancy Rate Reports, which contain data on number of beds per facility.

The number of residents at each hospital was obtained from the VA Office of Academic Affiliations, which oversees all graduate medical education within the VA and through which payment for resident service is provided. Funding for graduate medical education is tied to the number of residents within each hospital in the VA system, providing an incentive for reporting complete data. These data are audited regularly.

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 to differentiate major teaching, minor teaching, and nonteaching hospitals in previous studies.³⁰⁻³² Cross-sectional comparisons of quality of care in teaching and nonteaching hospitals have shown similar results using the resident-to-bed ratio or measures available through the American Hospital Association.³³ 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.

To support the validity of the resident-to-bed ratio as a marker of teaching intensity, we verified that hospitals with higher resident-to-bed ratios within VA have not only more residents but a much broader array of teaching programs. For example, hospitals in the lowest tertile of resident-to-bed ratio (0-0.179) have a mean of 2.5 residency programs in different specialties, hospitals in the middle tertile of resident-to-bed ratio (0.180-0.551) have a mean of 19.5 residency programs in different specialties, and hospitals in the highest tertile of resident-to-bed ratio (≥ 0.552) have a mean of

26.7 residency programs in different specialties.

We used the resident-to-bed ratio as a continuous variable to provide more power than dividing hospitals into arbitrary categories.^{34,35} We held the resident-to-bed ratio fixed using the level in prereform year 1 so that a potential response by hospitals to duty hour reforms of changing the number of residents would not confound estimation of the net effects of the reforms. Resident-to-bed ratios varied little over time. The mean change from prereform year 3 to prereform year 2 was -0.001 and from prereform year 2 to prereform year 1 was 0.001 . 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,³⁶ 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.^{37,38} 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, due to technological improve-

ments) 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 difference-in-differences method 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 difference in prereform trends. Less teaching-intensive hospitals, including all nonteaching hospitals, served as the primary control group for more teaching-intensive hospitals because they are subject to the same technological and VA-wide quality improvement imperatives, they are geographically diverse with large patient populations, and similar patient discharge data are available. Nonteaching hospitals could not be exclusively used as the control group for this analysis because only about 15% of VA hospitals are nonteaching hospitals. 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-intensive hospitals before the ACGME duty hour reform, we tested whether the rate of change in mortality was different in the more vs less teaching-intensive hospitals in the 3 years prereform (test of controls). This was performed by using a Wald χ^2 test, which tests whether the prereform year 2 \times resident-to-bed ratio and the prereform year 1 \times resident-to-bed ratio 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 a diverging trend was found for a condition 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. We tested the stability of the medical and surgical results by (1) eliminating patients admitted to hospitals in New York

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	7102	6736	6427	6084	5821
Unadjusted mortality, %	10.41	10.26	10.05	9.80	7.94
Stroke					
No. of cases	5408	5216	4908	4965	4888
Unadjusted mortality, %	10.97	10.33	10.09	8.60	8.35
Gastrointestinal bleeding					
No. of cases	7617	7420	7105	6997	6896
Unadjusted mortality, %	6.12	6.40	5.18	5.43	4.25
Congestive heart failure					
No. of cases	10 091	10 317	9580	10 075	10 203
Unadjusted mortality, %	6.62	6.57	6.00	5.83	5.28
Surgical categories					
General surgery					
No. of cases	15 204	15 639	15 555	16 314	16 862
Unadjusted mortality, %	3.72	3.58	3.25	3.29	2.80
Orthopedic surgery					
No. of cases	14 158	13 874	14 104	15 102	15 971
Unadjusted mortality, %	2.00	1.93	1.69	1.58	1.40
Vascular surgery					
No. of cases	4630	4500	4203	4353	4311
Unadjusted mortality, %	6.91	6.76	6.26	5.77	5.36

^aIncluded 318 636 patients and 131 hospitals. Prereform and postreform years are academic years (eg, prereform year 3 indicates academic year July 1, 2000, to June 30, 2001).

State, due to earlier passage of the Libby Zion laws; (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 or third CHF admission within 6 months of the first. All *P* values were either 2-tailed or, for χ^2 tests, multi-tailed. *P* < .05 was considered statistically significant. All 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 8% per year for any condition (TABLE 1). VA hospitals were teaching intensive, with approximately 85% of the hospitals being

teaching hospitals and more than 50% being major or very major teaching hospitals (resident-to-bed ratio >0.25) (TABLE 2).

Unadjusted mortality rates for the combined medical group improved similarly in hospitals in all quartiles of resident-to-bed ratios from prereform year 3 to postreform year 1 (FIGURE 1). For the combined medical group, in postreform year 2, however, there was relative improvement in the mortality rate in hospitals with the highest resident-to-bed ratios. The unadjusted OR for mortality for combined medical conditions in postreform year 1 was 1.09 (95% confidence interval [CI], 0.92-1.29) and in postreform year 2 was 0.74 (95% CI, 0.62-0.89). In contrast, among surgical patients, no apparent difference was observed in hospitals with different resident-to-bed ratios (Figure 1). The unadjusted OR for mortality for combined surgical categories in postreform year 1 was 0.94 (95% CI, 0.72-1.23) and in postreform year 2 was 1.05 (95% CI, 0.79-1.39).

Adjusted analyses indicate that for all medical conditions in postreform year 1, there was no statistically significant shift in the odds of mortality comparing more and less teaching-intensive hospitals (resident-to-bed ratio interaction with postreform year 1 in TABLE 3). However, in postreform year 2, the risk of death decreased further in more teaching-intensive hospitals for AMI and for all 4 medical conditions combined (resident-to-bed ratio interaction with postreform year 2 in Table 3). Because AMI was the only individual condition for which there was a significant relative reduction in the odds of mortality, we conducted a post hoc analysis of the 3 medical conditions combined excluding AMI; this group also showed a significant relative reduction in the odds of mortality in more vs less teaching-intensive hospitals. Among patients admitted for general, orthopedic, or vascular surgery, the relative odds of mortality in more vs less

teaching-intensive hospitals did not change in either postreform year 1 or postreform year 2 (Table 3). C statistics for these models ranged between 0.72 and 0.88.

The ORs in Table 3 are scaled to contrast resident-to-bed ratios of 1 and 0, but other hypothetical comparisons for combined medical conditions are shown in TABLE 4. For example, the odds of mortality for a patient admitted to a hospital with a resident-to-bed ratio of 0.60 (major teaching) in postreform year 2 compared with prereform would have improved 17% more than a similar person admitted to a nonteaching hospital in those periods, after adjustment for baseline differences in outcomes.

For each of the individual medical conditions except for AMI and for each of the surgical categories, the test of controls showed no evidence that prereform trends were different in more vs less teaching-intensive hospitals. However, for AMI, outcomes were improv-

ing more rapidly in more teaching-intensive hospitals prereform, albeit to a much smaller extent than that observed in postreform year 2. We tested the stability of the AMI results to adjusting for these differential trends by including the prereform year 2 \times resident-to-bed ratio and prereform year 1 \times resident-to-bed ratio interaction terms in the model and, alternatively, using prereform year 1 as the baseline year, and found no qualitative difference in our results.

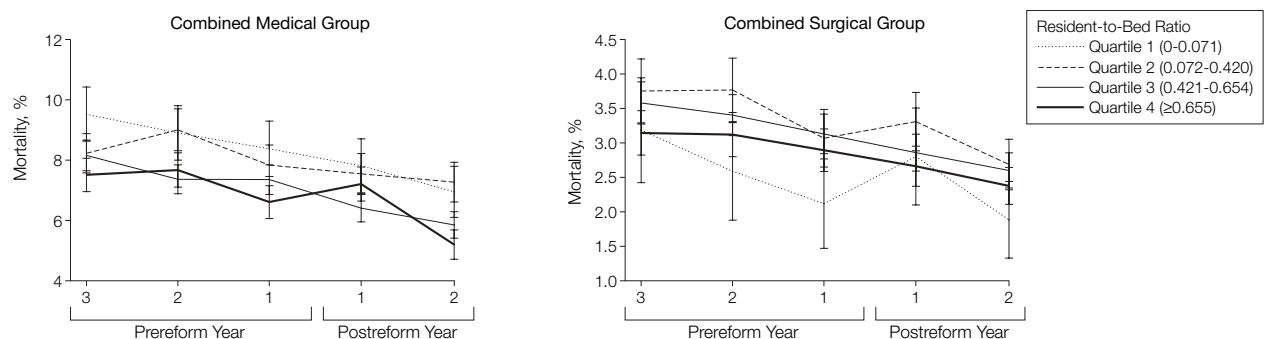
Excluding patients admitted to hospitals in New York State or patients admitted from nursing homes from the analysis did not change the results. We also examined whether changes in the coding of comorbidities could explain any of these effects. Although we found that there was a 2% to 4% relative increase in the coding of comorbidities in more teaching-intensive hospitals relative to nonteaching hospitals in postreform year 2, sensitivity analyses with-

Table 2. Characteristics of Included US Veterans Affairs Hospitals^a

	Resident-to-Bed Ratio, Percentile (Range)				
	0-25 (0-0.071)	26-50 (0.072-0.420)	51-75 (0.421-0.654)	76-90 (0.655-0.867)	>90 (0.868-1.794)
No. (%) of facilities	32 (24.4)	34 (26.0)	33 (25.2)	19 (14.5)	13 (9.9)
No. (%) of admissions	27 378 (8.6)	65 801 (20.7)	127 096 (39.9)	59 006 (18.5)	39 355 (12.4)

^aPercentages may not total 100 due to rounding. Included 131 facilities and 318 636 admissions.

Figure 1. Unadjusted Trends in Mortality in US Veterans Affairs Hospitals for Combined Medical and Surgical Groups by Resident-to-Bed Ratio Quartile



Error bars indicate 95% confidence intervals, calculated using the binomial distribution.³⁹ 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 for postreform year 2 only (by Wald χ^2 test, $P = .001$). For combined surgical group, no significant divergence was found in either postreform year. Significance levels assess whether trend from the prereform period to postreform year 1 and 2, respectively, differed for more vs less teaching-intensive hospitals.

out adjusting for comorbidities produced similar results. For example, the OR for combined medical conditions without risk adjustment was 0.74 (95% CI, 0.62-0.89) vs 0.74 (95% CI, 0.61-0.89) with risk adjustment. In analyses of combined medical without AMI, the OR in postreform year 2 without risk adjustment was 0.82 (95% CI, 0.66-1.01) vs 0.79 (95% CI, 0.63-0.98) 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 AMI with adjustment for anatomic location of AMI or for patients with CHF who had a second or third admission within 6 months of the first.

FIGURE 2 shows the estimated probability of mortality in each year for an average patient in hospitals at the 25th (resident-to-bed ratio of 0.07), 50th (resident-to-bed ratio of 0.42), 75th (resident-to-bed ratio of 0.65), and 90th (resident-to-bed ratio of 0.87) percentile of teaching intensity. For combined medical conditions, adjusted mortality trends tracked in parallel among hospitals with different teaching intensity between prereform year 3 and prereform year 1; however, they diverged significantly in postreform year 2. From prereform year 1 to postreform year 2, an average patient in a hospital at the 90th percentile would have experienced an improvement in mortality from 5.15% to 3.74%, an improvement of 1.41 percentage points. Compared with the underlying rate of improvement for an average patient in a hospital at the 25th percentile, for whom mortality improved by 0.53 percentage points (from 6.29% to 5.76%),

this represents an absolute difference of 0.88 percentage points (13.9% relative decrease). An average patient treated at a hospital at the 75th percentile would have experienced an absolute improvement in mortality that was 0.70 percentage points more than at a hospital at the 25th percentile (11.1% relative decrease). In contrast with the medical conditions, we found no apparent differential trend in mortality rates between surgical patients in more vs less teaching-intensive hospitals.

COMMENT

It has been argued that duty hour reform would improve outcomes by reducing resident fatigue,⁴¹⁻⁴⁴ although other studies have suggested that decreased continuity of care would worsen

outcomes.^{8,45,46} Our results suggest significant relative improvements in mortality rates for patients with 4 common, high-mortality medical conditions in more teaching-intensive VA hospitals following implementation of the duty hour rules.

These results should be considered in light of other examinations of duty hour reform and outcomes. A parallel analysis that we performed on Medicare recipients nationally showed no significant relative changes in mortality for either medical or surgical patients in more vs less teaching-intensive hospitals.¹⁶ Possible reasons for the differences in findings are detailed in that study, but briefly they include the markedly greater mean resident-to-bed ratios at VA teaching

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 × Postreform Year 1 ^a		Resident-to-Bed Ratio × Postreform Year 2 ^a	
		Odds Ratio (95% CI) ^b	P Value	Odds Ratio (95% CI) ^b	P Value
Medical conditions					
Acute myocardial infarction	32 170	1.11 (0.79-1.56)	.57	0.48 (0.33-0.71)	<.001
Stroke	27 434	0.84 (0.57-1.23)	.36	0.83 (0.57-1.20)	.31
Gastrointestinal bleeding	36 035	1.22 (0.81-1.84)	.35	0.75 (0.48-1.18)	.21
Congestive heart failure	50 266	1.00 (0.74-1.37)	.99	0.77 (0.56-1.07)	.12
Combined medical	145 905	1.08 (0.91-1.29)	.39	0.74 (0.61-0.89)	.002
Combined medical, excluding acute myocardial infarction	113 735	1.02 (0.82-1.25)	.89	0.79 (0.63-0.98)	.03
Surgical categories					
General surgery	79 574	0.92 (0.64-1.33)	.67	0.92 (0.63-1.34)	.66
Orthopedic surgery	73 209	0.87 (0.47-1.63)	.67	1.33 (0.71-2.52)	.38
Vascular surgery	21 997	0.74 (0.43-1.27)	.27	0.96 (0.55-1.66)	.89
Combined surgical	174 780	0.89 (0.68-1.17)	.41	1.02 (0.77-1.36)	.88

Abbreviation: CI, confidence interval.

^aThe interaction terms (resident-to-bed ratio × postreform year 1) and (resident-to-bed ratio × 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.

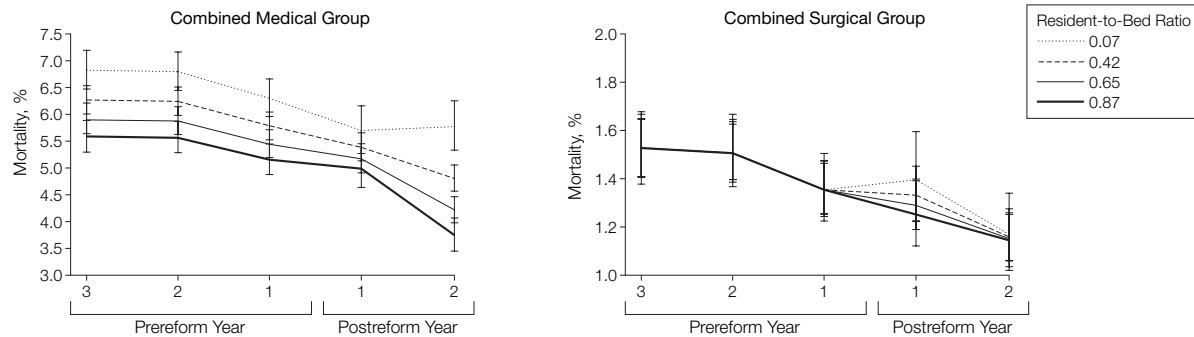
^bUnits for odds ratios compare hospitals having a resident-to-bed ratio of 1 with hospitals having a resident-to-bed ratio of 0 (nonteaching).

Table 4. Comparison of Changes in Odds of Mortality Between Prereform Period and Postreform Year 2 for Combined Medical Conditions for Hospitals With Different Levels of Teaching Intensity^a

Resident-to-Bed Ratio	Relative Change in Odds of Mortality (95% CI)			
	0.05 (Very Minor Teaching)	0.25 (Minor Teaching)	0.60 (Major Teaching)	1 (Very Major Teaching)
0 (Nonteaching)	0.98 (0.98-0.99)	0.93 (0.88-0.97)	0.83 (0.74-0.93)	0.74 (0.61-0.89)
0.05 (Very minor teaching)		0.94 (0.91-0.98)	0.85 (0.76-0.94)	0.75 (0.63-0.90)
0.25 (Minor teaching)			0.90 (0.84-0.96)	0.80 (0.69-0.92)
0.60 (Major teaching)				0.89 (0.82-0.96)

Abbreviation: CI, confidence interval.

^aEach cell indicates the ratio of mortality change in the more teaching-intensive hospital to that in the less teaching-intensive hospital. For example, very major teaching hospitals lowered their odds of mortality by 20% more than minor teaching hospitals.

Figure 2. Estimated Probability of Mortality for an Average Patient in Combined Medical and Surgical Groups by Resident-to-Bed Ratio

Plots show changes in mortality for a patient with mean values of all the comorbidities in hospitals with a resident-to-bed ratio at each percentile of teaching intensity. Error bars indicate 95% confidence intervals, derived from the logit model using the Delta method.⁴⁰ 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 for postreform year 2 only (by Wald χ^2 test, $P=.002$). For combined surgical group, no significant divergence was found in either postreform year.

hospitals compared with non-VA teaching hospitals, potentially greater autonomy for residents at VA hospitals,^{47,48} differences in staffing models and clinical volume, differing balances between the effects of decreased fatigue⁴¹⁻⁴³ and worsened continuity,^{45,46} and potentially different degrees of unmeasured confounders.

A single-site study that compared changes in outcomes on the teaching vs nonteaching service before and after duty hour reform was not large enough to have adequate power to compare changes in mortality.¹⁴ A larger study¹⁵ using the Healthcare Cost and Utilization Project's (HCUP's) Nationwide Inpatient Sample (NIS) found a small but statistically significant relative improvement in mortality outcomes for medical but not surgical patients in teaching hospitals compared with nonteaching hospitals following duty hour reform. However, the HCUP NIS samples varying sets of hospitals and states each year, with possible variation in the proportion of data from each included hospital in different periods. This may affect the validity of the comparisons over time. The HCUP NIS does not include individual patient identifiers, precluding a distinction between a single admission for different patients vs multiple admissions for the same patient, which could lead to a bi-

ased measure of relative changes in mortality over time. The HCUP NIS does not include any information on out-of-hospital deaths, limiting the outcomes to in-hospital mortality; therefore, differential changes in discharge rates may bias the degree to which in-hospital mortality reflects deaths within a fixed period of time from admission. Our study avoided these limitations by including data on all hospitals in all years, limiting participants to a single first-eligible admission for a condition, and measuring total 30-day post-admission mortality.

Although this is an observational study and we cannot be certain that the reduction in mortality was caused by the reform, the findings are nonetheless reassuring. Some or all of the improvement may have been due to hospital efforts to realign service delivery in response to duty hour reform rather than from reduced fatigue among medical house staff. However, relative improvements were observed for medical but not surgical patients in more teaching-intensive hospitals. This suggests that other initiatives that might have led to relative improvements in outcomes for all patients in more teaching-intensive hospitals did not confound our results. These findings are important because the VA is the single largest provider of residency training in

the United States, providing at least some of the training for approximately one-third of all residents each year. More than two-thirds of all US physicians have received some of their training at VA facilities.⁴⁹

The significant relative improvement in mortality observed in the postreform year 2 but not the postreform year 1 is consistent with recent work, suggesting that rates of noncompliance with duty hour rules were high but improving during the postreform year 1.⁵⁰ However, a study by the ACGME found that only 3.3% of residents surveyed in 2004-2005 (postreform year 2) reported working more than 80 hours per week in the previous month compared with 3.0% during 2003-2004.⁵¹ Studies observing the effects of the Libby Zion laws in New York State have found no differences in mortality in teaching vs nonteaching hospitals⁴⁴ or within the same teaching hospital.⁴⁵ However, compliance rates may have been very low.^{44,52}

It is unclear why mortality improved for medical but not surgical programs within VA hospitals. One potential explanation is that for surgical residents, duty hour reform resulted in a relative worsening in continuity of care that offset any improvements from decreased fatigue. Medical training programs may have developed better mechanisms for sign-out and increas-

ing attending involvement to offset losses in continuity. Another possibility is that medical programs within VA hospitals were more compliant with the duty hour rules than surgical programs. It is possible that surgical trainees were less affected by duty hour rules than medical trainees, although the larger number of hours worked by surgical residents pre-reform⁵³ suggested that the degree of change in work hours would have been larger. Further work is needed using different types of measures and data on hospital behavioral responses to duty hour reform to better understand the difference in these effects between medical and surgical patients.

Potential limitations to our study must be considered. We investigated associations with patient mortality, whereas there may be other effects of duty hours of clinical importance to patient care or resident training that were affected by duty hour changes. These outcomes are important in their own right, but also may have acted as mediators of the mortality findings. Our study may be properly considered an effectiveness study rather than an efficacy study.

We do not have information on actual hours worked at each hospital, although the risk of loss of ACGME accreditation may provide an incentive for adherence to the duty hour regulations. The VA system is more teaching intensive than the non-VA system, with a much higher proportion of teaching hospitals and greater autonomy for residents,^{47,48} so findings from the VA may not generalize to the non-VA environment.

We may not have had sufficient power to rule out a clinically significant effect of duty hour reform in surgical training programs, despite the inclusion of more surgical than medical patients. Because mortality rates were much lower among surgical patients, a similar relative change in the odds of mortality for surgical as for medical patients cannot be ruled out. However, the point estimates do not suggest a relative change in the odds of mortality for surgical patients in more vs less teaching-intensive hospitals. If such an effect

did exist, the magnitude of the absolute change in mortality would be small.

Although changes in coding of comorbidity over time must be considered, the results were similar with and without comorbidity adjustment. In addition, during the 5 years of this study, there was little change in the mean number of comorbidities, the number of patients with 0, 1, 2, or 3 comorbidities, or the mean age in hospitals of low, moderate, or high teaching intensity. This suggests that there were no major changes in the patient profiles over time for hospitals of different teaching intensity.

Unmeasured confounding must always be considered. Use of administrative data limits risk adjustment, but the multiple time series difference-in-differences approach reduces the likelihood of important unmeasured confounding. Our analysis adjusted for differences in baseline rates of change in more and less teaching-intensive hospitals. Therefore, confounding would only occur if the prevalence of unmeasured risk factors changed at different rates according to hospitals' teaching intensity. Confounding from other quality improvement initiatives is less likely to be a problem within the VA system, because quality improvement initiatives such as electronic medical records and efforts to improve care for specific conditions tend to occur through system-wide directives that all VA hospitals must follow.

Although our study suggests relative improvements in outcomes for medical patients within the VA system from duty hour reform, there continues to be controversy about whether the current duty hour rules are sufficient for regulating work hours.⁷ Other work has found significantly lower rates of errors with 16-hour vs 24-hour to 36-hour shifts.^{41,42} The question of the relative effectiveness and cost-effectiveness of different approaches to duty hour regulation in improving patient outcomes can only be answered through further experimentation with alternative designs accompanied by rigorous evaluation.

In conclusion, we found that the duty hour reforms were associated with a significant improvement in mortality in more teaching-intensive VA hospitals for patients with medical conditions. Furthermore, we did not find an increase in mortality associated with the new rules affecting surgical patients. Further assessment of how the reforms affected other clinical and educational outcomes in both VA and non-VA settings would be important before modification of the current duty hour standards.

Author Affiliations: Center for Health Equity Research and Promotion, Veterans Administration Hospital, Philadelphia, Pennsylvania (Dr Volpp and Ms Canamucio); Center for Outcomes Research, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Dr Silber and Ms Even-Shoshan); 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: Dr Volpp had full access to all of the data in the study and takes 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, Canamucio, Behringer, Silber.

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

Drafting of the manuscript: Volpp, Rosenbaum, Canamucio, Silber.

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

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

Obtained funding: Volpp, Rosen, 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 IIR 04.202.1 from US Veterans Affairs Health Services Research and Development Service. Support was also received from grants R01 HL082637 from the National Heart, Lung, and Blood Institute and 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: David Blumenthal, MD, Harvard Medical School; David Dinges, PhD, University

of Pennsylvania School of Medicine; seminar participants at the University of Chicago and the University of Pennsylvania; and the ACGME Committee on Improving the Learning Environment provided comments on earlier drafts. Liyi Cen, MS, Philadelphia Veterans Affairs Medical Center and University of Pennsylvania, provided help as a statistical programmer employed on this project. Anee Lee, BA, Philadelphia Veterans Affairs Medical Center, provided help as a research assistant employed on this project. Drs Blumenthal and Dinges did not receive any compensation for their roles as members of the advisory board for this study.

REFERENCES

- Resident duty hours language: final requirements. Accreditation Council for Graduate Medical Education Web site. <http://www.acgme.org>. Accessed July 27, 2003.
- 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: a summary of achievements for the second year under the common duty hour standards. Accreditation Council for Graduate Medical Education Web site. http://www.acgme.org/acWebsite/dutyHours/dh_achieveSum04-05.pdf. Accessed October 3, 2006.
- Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
- 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.
- Gaba DM, Howard SK. Patient safety: fatigue among clinicians and the safety of patients. *N Engl J Med*. 2002;347(16):1249-1255.
- Drazen 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.
- Niederee MJ, Knudtson JL, Byrnes MC, Helmer SD, Smith RS. A survey of residents and faculty regarding work hour limitations in surgical training programs. *Arch Surg*. 2003;138(6):663-669, 669-671.
- Biller CK, Antonacci AC, Pelletier S, et al. The 80-hour work guidelines and resident survey perceptions of quality. *J Surg Res*. 2006;135(2):275-281.
- 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.
- 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.
- Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. *JAMA*. 2007;298(9):975-983.
- 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.
- 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.
- Iezzoni LI, ed. *Risk Adjustment for Measuring Health-care Outcomes*. 3rd ed. Chicago, IL: Health Administration Press; 2003:73.
- Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27.
- 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.
- 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.
- 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.
- 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.
- Volpp KG, Stone R, Lave JR, et al. Is thirty-day hospital mortality really lower for black veterans compared to white veterans? *Health Serv Res*. 2007;42(4):1613-1631. doi:10.1111/j.1475-6773.2006.00688.x.
- Fisher SG, Weber L, Goldberg J, Davis F. Mortality ascertainment in the veteran population: alternatives to the National Death Index. *Am J Epidemiol*. 1995;141(3):242-250.
- Page WF, Braun MM. Ascertainment of mortality in the U.S. veteran population: World War II veteran twins. *Mil Med*. 1995;160(7):351-355.
- 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.
- 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.
- Agresti A. *An Introduction to Categorical Data Analysis*. New York, NY: Wiley; 2007:108-109.
- 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.
- Poulose BK, Ray WA, Arbogast PG, et al. Resident work hour limits and patient safety. *Ann Surg*. 2005;241(6):847-856, 856-860.
- Howard DL, Silber JH, Jobses DR. Do regulations limiting residents' work hours affect patient mortality? *J Gen Intern Med*. 2004;19(1):1-7.
- 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.
- 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.
- Veterans Health Administration Handbook 1400.1. Resident Supervision. July 27, 2005. http://www1.va.gov/vhapublications/ViewPublication.asp?pub_ID=1289. Accessed May 19, 2007.
- Center 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.
- Association of American Medical Colleges. 2005 Medical School Graduation Questionnaire Final All Schools Report. <http://www.aamc.org/data/gq/allschoolsreports/2005.pdf>. Accessed June 1, 2007.
- 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.
- Baldwin DC Jr, Daugherty SR, Tsai R, Scotti MJ Jr. A national survey of residents' self-reported work hours: thinking beyond specialty. *Acad Med*. 2003;78(11):1154-1163.