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 S. 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

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.
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=8,524), transfers from a non-VA hospital (n=6,337), transfers subsequent to a contiguous qualifying admission (to avoid double counting admissions within 30 days) (n=4,103), admissions spanning July 1, 2003 (n=1,729), 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=23,888) 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=5,231) 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 post-reform 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-
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 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 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 χ² test, which tests whether the prereform year 2 × resident-to-bed ratio and the prereform year 1 × 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 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 χ² tests, multiailed. 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).

©2007 American Medical Association. All rights reserved.

(Reprinted) JAMA, September 5, 2007—Vol 298, No. 9 987

Table 1. Characteristics of the Study Populationa

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prereform Year</th>
<th>Postreform Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted mortality, %</td>
<td>7128</td>
<td>7056</td>
</tr>
<tr>
<td>No. of cases</td>
<td>4630</td>
<td>4500</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>6.91</td>
<td>6.76</td>
</tr>
<tr>
<td>No. of cases</td>
<td>10091</td>
<td>10317</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>6.62</td>
<td>6.57</td>
</tr>
<tr>
<td>No. of cases</td>
<td>4630</td>
<td>4500</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>2.00</td>
<td>1.93</td>
</tr>
<tr>
<td>No. of cases</td>
<td>15158</td>
<td>15387</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>3.72</td>
<td>3.58</td>
</tr>
<tr>
<td>No. of cases</td>
<td>10091</td>
<td>10317</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>6.91</td>
<td>6.76</td>
</tr>
<tr>
<td>No. of cases</td>
<td>4630</td>
<td>4500</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>2.00</td>
<td>1.93</td>
</tr>
<tr>
<td>No. of cases</td>
<td>15158</td>
<td>15387</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>3.72</td>
<td>3.58</td>
</tr>
<tr>
<td>No. of cases</td>
<td>10091</td>
<td>10317</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>6.91</td>
<td>6.76</td>
</tr>
<tr>
<td>No. of cases</td>
<td>4630</td>
<td>4500</td>
</tr>
<tr>
<td>Unadjusted mortality, %</td>
<td>2.00</td>
<td>1.93</td>
</tr>
</tbody>
</table>

a Included 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).
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 improving 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 × resident-to-bed ratio and prereform year 1 × 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

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

*Percentages 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. 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 χ² 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.

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

<table>
<thead>
<tr>
<th>Patient Categories</th>
<th>Total No. of Cases</th>
<th>Resident-to-Bed Ratio × Postreform Year 1</th>
<th>Resident-to-Bed Ratio × Postreform Year 2</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical conditions</td>
<td></td>
<td>Odds Ratio (95% CI) b</td>
<td>Odds Ratio (95% CI) b</td>
<td>P Value</td>
</tr>
<tr>
<td>Acute myocardial infarction</td>
<td>32,170</td>
<td>1.11 (0.79-1.56) .07 0.48 (0.33-0.71) &lt;.001</td>
<td>0.79 (0.63-0.98) .03 0.79 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>27,434</td>
<td>0.84 (0.57-1.23) .36 0.83 (0.57-1.20) .31</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal bleeding</td>
<td>36,035</td>
<td>1.22 (0.81-1.84) .35 0.75 (0.48-1.18) .21</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>50,266</td>
<td>1.00 (0.74-1.37) .99 0.77 (0.56-1.07) .12</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Combined medical</td>
<td>145,905</td>
<td>1.08 (0.91-1.29) .39 0.74 (0.61-0.89) .002</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Combined medical, excluding acute myocardial infarction</td>
<td>113,735</td>
<td>1.02 (0.82-1.25) .89 0.79 (0.63-0.98) .03</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Surgical categories</td>
<td></td>
<td>Odds Ratio (95% CI) b</td>
<td>Odds Ratio (95% CI) b</td>
<td>P Value</td>
</tr>
<tr>
<td>General surgery</td>
<td>79,509</td>
<td>0.92 (0.64-1.33) .67 0.92 (0.63-1.34) .66</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Orthopedic surgery</td>
<td>73,209</td>
<td>0.87 (0.67-1.33) .67 1.33 (0.71-2.52) .38</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Vascular surgery</td>
<td>21,997</td>
<td>0.74 (0.43-1.27) .27 0.96 (0.55-1.66) .89</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
<tr>
<td>Combined surgical</td>
<td>174,780</td>
<td>0.89 (0.68-1.17) .41 1.02 (0.77-1.36) .38</td>
<td>0.74 (0.63-0.98) .03 0.74 (0.63-0.98) .03</td>
<td></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Resident-to-Bed Ratio</th>
<th>0.05 (Very Minor Teaching)</th>
<th>0.25 (Minor Teaching)</th>
<th>0.60 (Major Teaching)</th>
<th>1 (Very Major Teaching)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Nonteaching)</td>
<td>0.98 (0.98-0.99)</td>
<td>0.93 (0.88-0.97)</td>
<td>0.83 (0.74-0.93)</td>
<td>0.74 (0.61-0.89)</td>
</tr>
<tr>
<td>0.05 (Very minor teaching)</td>
<td>0.94 (0.91-0.98)</td>
<td>0.85 (0.76-0.94)</td>
<td>0.75 (0.63-0.90)</td>
<td></td>
</tr>
<tr>
<td>0.25 (Minor teaching)</td>
<td>0.90 (0.84-0.96)</td>
<td>0.80 (0.69-0.92)</td>
<td>0.80 (0.69-0.92)</td>
<td></td>
</tr>
<tr>
<td>0.60 (Major teaching)</td>
<td>0.89 (0.82-0.96)</td>
<td>0.89 (0.82-0.96)</td>
<td>0.89 (0.82-0.96)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

*The 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.

**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

©2007 American Medical Association. All rights reserved.

(Reprinted) JAMA, September 5, 2007—Vol 298, No. 9
Mortality among Patients in VA 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. The Accreditation Council for Graduate Medical Education duty hour regulations were implemented on July 1, 2003. Prereform year 1 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 χ² 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, differences in staffing models and clinical volume, differing balances between the effects of decreased fatigue and worsened continuity, 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. A larger study 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 biased 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 postadmission 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.

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. 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. Studies observing the effects of the Libby Zion laws in New York State have found no differences in mortality in teaching vs nonteaching hospitals or within the same teaching hospital. However, compliance rates may have been very low.

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

990 JAMA, September 5, 2007—Vol 298, No. 9 (Reprinted) ©2007 American Medical Association. All rights reserved.
hospitals and greater autonomy for residents, a much higher proportion of teaching-intensive than the non-VA system, with greater compliance with duty hour rules. Other possible reasons for the improved outcomes in the VA system include system-wide directives that all VA hospitals must follow.

The sponsors had no role in the study design, data collection, analysis, and interpretation of the data, or in the writing of the manuscript. The manuscript was reviewed and approved by all authors. None reported.

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 morbidity 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, 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 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.

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 (Dr Volpp and Bellini, and Ms Bhringer) and Anesthesiology and Critical Care (Dr Silber), University of Pennsylvania School of Medicine, Philadelphia, Department of Health Care Systems (Dr 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 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, Bhringer, Silber.

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

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

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

Obtained funding: Volpp, Rosen, Rosenbaum, Silber.

Administrative, technical, or material support: Volpp, Even-Shoshan, Bellini, Bhringer, 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-D646602 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 School of Public Health, Boston, Massachusetts, and Center for Health Quality, Outcomes and Economic Research, Veterans Administration Hospital, Bedford, Massachusetts (Dr Rosenbaum); and Division of General Medicine and Center for Healthcare Policy and Research, University of California Davis School of Medicine, Sacramento (Dr Romano).
of Pennsylvania School of Medicine; seminar partic-
pants at the University of Chicago and the University
of Pennsylvania, and the ACGME Committee on
improving the Learning Environment provided com-
ments on earlier drafts. Lyi Cen, MS, Philadelphia Vet-
erans Affairs Medical Center and University of Penn-
sylvania, provided help as a statistical program-
mer employed on this project. Anee Lee, BA, Phila-
delphia Veterans Affairs Medical Center, provided help as
a research assistant employed on this project. Drs
Blumenthal and Dinges did not receive any compen-
sation for their roles as members of the advisory board
for this study.

REFERENCES
  1. Resident duty hours language: final require-
mements. Accreditation Council for Graduate Medical
  2. Report of the work group on resident duty hours and the
learning environment, June 11, 2002. In: The ACG-
ME’s approach to limit resident duty hours: a summary
of achievements for the second year under the common
duty hour standards. Accreditation Council for Gradu-
ate Medical Education Web site. http://www.acgme.org/
acWebsite/dutyHours/dh_achieveSum04-05.pdf. Ac-
cessed October 3, 2006.
  3. Kohn LT, Corrigan JM, Donaldson MS. To err is hu-
mankind: building a safer health system. Washington, DC:
  4. Fletcher KE, Davis SQ, Underwood W, Man-
griksr AS, McMahan LF Jr, Saint S. Systematic re-
view of the literature regarding high resident work
  5. Weinger MB, Ancoli-Israel S. Sleep deprivation and
  6. Charap M. Reducing resident work hours: un-
sure outcome or an opportunity for orthopaedic
  7. Leuconi LJ, ed. Risk Adjustment for Measuring
Care outcomes. 3rd ed. Chicago, IL: Health
Administration Press; 2000.
  8. Smith RS. A survey of residents and faculty regarding
changes in outcomes for internal medicine inpa-
  9. Niederee MJ, Knudtson JL, Byrnes MC, Helmer SD,
Horwitz LI, Kosiborod M, Lin Z, Krumholz HM. A
preliminary analysis of two comorbidity mea-
sures, with and without information from priori-
  10. Southern DA, Quan H, Gahi WA. Comparison of
the Elashvili and Charlson/Death/Neo methods of
comorbidity measurement in administrative data. Med
  11. Loeangi GC, Dick AW, Osler TM, Mukamel DB. Do
data stamping ICD-9-CM codes increase the value of
data in administrative information in administrative data? Health Serv
approaches for defining comorbidities in ICD-9-CM and
ICD-10 administrative data. Med Care. 2005;43
(11):1130-1139.
  13. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinica-
logical comorbidity index for use with ICD-9-CM admin-
613-619.
mortality really lower for black veterans compared to
white veterans? Health Serv Res. 2007;42(4):1613-
  15. Fisher SG, Weber L, Goldberg I, Davis F. Mortal-
ity ascertainment in the veteran population: alterna-
tives to the National Death Index. Am J Epidemiol.
  17. Blumenthal JA, Dinges DF, Woolf R, Northfield TC.
Variation in outcome after acute upper gastrointes-
tinal haemorrhage: the National Audit of Acute Upper
Gastrointestinal Haemorrhage. Lancet. 1995;
  18. Loeangi GC, Dick AW, Osler TM, Mukamel DB. Mort-
2006;21(12):1154-1163.
  19. Ayanian JZ, Weissman JS. Teaching hospitals and
mortality among hospitalized Medicare beneficiaries in the
first 2 years following ACGME resident duty hour
  20. Loeangi GC, Dick AW, Osler TM, Mukamel DB. Mort-
2006;21(12):1154-1163.
TH. Does housestaff discontinuity of care increase the
risk for preventable adverse events? Ann Intern Med.
  22. Laine C, Goldman L, Soukup JR, Hayes JG. The
impact of a regulation restricting medical house staff work-
hours on the quality of patient care. JAMA. 1993;
  23. Veterans Health Administration Handbook 14001.
gov/hospublications/ViewPublication.asp?pubID=
  24. Center for Medicare and Medicaid Services. Guide-
lines for teaching physicians, interns, and residents.
May 19, 2007.
  25. Association of American Medical Colleges. 2005
Medical School Graduation Questionnaire Final All
  26. Langrider 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.
  27. Accreditation Council for Graduate Medical Education.
The ACGME’s approach to limit resident duty
hours 12 months after implementation: a sum-
acWebsite/dutyHours/dh_achieveSum04-05.pdf. Ac-
cessed October 3, 2006.
  28. DeBuono BA, Osten WM. The medical resident
workload: the case of New York State. JAMA. 1998;
A national survey of residents’ self-reported work hours:
thinking beyond specialty. Acad Med. 2003;78(11):
1154-1163.

MORTALITY AMONG PATIENTS IN VA HOSPITALS