Cite this article as:
D P Wagner, W A Knaus and E A Draper
The case for adjusting hospital death rates for severity of illness
Health Affairs 5, no.2 (1986):148-153
doi: 10.1377/hlthaff.5.2.148

The online version of this article, along with updated information and services, is available at:
http://content.healthaffairs.org/content/5/2/148.citation

For Reprints, Links & Permissions :
http://content.healthaffairs.org/1340_reprints.php

Email Alertings :
http://content.healthaffairs.org/subscriptions/etoc.dtl

Not for commercial use or unauthorized distribution
To Subscribe: https://fulfillment.healthaffairs.org
The Case For Adjusting Hospital Death Rates For Severity Of Illness

by Douglas P. Wagner, William A. Knaus, and Elizabeth A. Draper

On April 17, 1985 the federal government issued sixty pages of final rules and regulations in the Federal Register regarding the activities of Utilization and Quality Control Peer Review Organizations (PROs).1 A noteworthy section (476.141) allows PROs, for the first time, to “disclose to the public PRO interpretations and generalizations on the quality of health care that identify a particular institution.” Such disclosure cannot directly or indirectly identify individual patients or providers, but can name individual hospitals.

Since the rules and regulations were published, the Health Care Financing Administration (HCFA) has released data identifying hospitals with Medicare patient death rates which were significantly higher or lower than standard. The agency’s methodology for estimating the standard was apparently based on each hospital’s case-mix distribution across diagnosis-related groups (DRGs) and age, and on the portion of patients with cancer—in other words, discharge abstract information aggregated to the hospital level. Publicized in the New York Times and USA Today among other newspapers, the data were sharply criticized by hospitals and state health agencies, while consumer organizations applauded the release. Long before HCFA made its data public, however, there were signs that the issue would become controversial. In 1982, when unidentified hospital death rates were released in Maryland, consumer advocates argued for wider dissemination and linkage with individual institutions.2 Officials from Maryland hospitals responded that the figures were incomplete and potentially misleading.

Though excess death rates may identify poor quality hospitals and drive them out of business, do they truly identify hospitals with a poor quality of care? Referral hospitals are worried that their more serious case-mix will be inadequately documented, leading to unfair and adverse publicity. Hospitals that have practiced conservative medical care, admitting

Douglas P. Wagner is senior staff scientist, William A. Knaus, director, and Elizabeth A. Draper, senior research scientist, at the Intensive Care Unit (ICU) Research Unit of the George Washington University Medical Center. The research unit is part of both the Departments of Anesthesiology and Computer Medicine. Knaus is also an associate professor in both these departments, as well as codirector of the ICU.
primarily patients with advanced disease, could be identified as poor quality because of their adverse case selection. Extensive research on physician practice patterns have already demonstrated large variations in hospital admission rates which are attributed to physician practice style.³

This article briefly examines these concerns using a prospectively collected data base covering 5,030 patients admitted to intensive care units (ICUs) in thirteen tertiary care hospitals. The data set was collected as part of an effort to validate the Acute Physiology and Chronic Health Evaluation (APACHE II) severity of illness classification system.⁴ It includes the broad range of diagnoses among patients commonly admitted to medical and surgical intensive care units with the exceptions of postcoronary artery bypass graft, burn, heart attack, and pediatric patients. The APACHE II severity system is primarily based on physiologic data obtained from vital signs and routine blood tests completed within twenty-four hours of admission into the ICU.

The APACHE II scores were closely correlated with hospital death rates across the range of severity of illness. Exhibit 1 illustrates that observed

---

**Exhibit 1**

Severity-Predicted And Observed Death Rate

<table>
<thead>
<tr>
<th>Severity score (APACHE II) on ICU admissions</th>
<th>Death rate (percent of admissions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 3 6 9 12 15 18 21 24 27 30 33 36 39 42+</td>
<td>Observed</td>
</tr>
<tr>
<td></td>
<td>Severity-predicted</td>
</tr>
</tbody>
</table>

⁴ Based on 5,030 intensive care unit admissions.

⁵ APACHE scores depicted in three point ranges.
death rates are close to predicted death rates across the entire range of severity of illness, from 1.9 percent death rate at the lower end to 94 percent at the upper. The predictions were based on a multivariate analysis which adjusted for each of forty-eight mutually exclusive diagnostic categories as well as APACHE II score.

The advantage of using a severity of illness classification system in combination with diagnosis rather than diagnosis alone is illustrated in Exhibit 2. This exhibit tabulates observed and predicted death rates across the same severity ranges as Exhibit 1, but this time outcome predictions were based on diagnosis alone. Note that in the lower ranges of severity, diagnosis alone predicts higher death rates than observed, and in the more severe regions, diagnosis substantially underpredicts death rates. Though not identical to DRG definitions, the predictions in Exhibit 2 approximate what one would expect using the diagnostic information now available in hospital discharge abstracts.

The discrepancies between predicted and observed death rates in Exhibit 2 would not matter if severely ill patients were equally distri-

---

**Exhibit 2**

Diagnosis-Predicted And Observed Death Rate

<table>
<thead>
<tr>
<th>Death rate (percent of admissions)</th>
<th>Observed</th>
<th>Diagnosis-predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Severity score (APACHE II) on ICU admissions

*Based on 5,030 intensive care unit admissions.

*APACHE scores depicted in three point ranges.
buted across hospitals. But they are not. Exhibit 3 illustrates the implications of the two alternate mortality standards (Exhibit 1 and Exhibit 2) for each of the thirteen hospitals. Each bar on Exhibit 3 represents the

Exhibit 3
Comparison Of Two Standards For Death Rates

Percent difference from average performance\(^a\) between predicted & observed death rates

\(120\)
\(110\)
\(100\)
\(90\)
\(80\)
\(70\)
\(60\)
\(50\)
\(40\)
\(30\)
\(20\)
\(10\)
\(0\)
\(-10\)
\(-20\)
\(-30\)
\(-40\)
\(-50\)
\(-60\)

Hospital

\(^a\)Based on 5,030 intensive care admissions.

\(^b\)Computed as \((\text{predicted} - \text{observed})/\text{observed death rate}\) for each of the two methods of prediction.
difference between the predicted and observed death rates divided by
the observed death rate for each hospital. A percent difference of zero
means the hospital is performing average for this study. Negative per-
centages indicate worse than average performance; positive percentages
indicate better than average performance.

The first of the two parallel bars is the ratio computed from the same
predictive equation which generated Exhibit 1 (diagnosis plus severity),
while the second bar is from the equation which generated Exhibit 2
(diagnosis alone). The hospitals are sorted in descending order by their
diagnosis plus severity predicted mortality ratios. Under this latter stand-
ard, hospital one is the best with 72 percent more deaths predicted than
observed and hospital thirteen the worst, with 37 percent fewer deaths
predicted than observed. It is also evident from this exhibit that there are
substantial discrepancies between a mortality standard based on diagno-
sis alone versus diagnosis plus severity combined. Without further infor-
mation, however, we cannot determine which of the two standards is
closest to the truth.

Fortunately, additional information is available concerning ICU admis-
sion criteria and the quality of care provided at these thirteen hospitals.
The information is best summarized by discussing hospitals one, four,
six, and thirteen. At hospital one, there was a substantial institutional
commitment to good coordination among the highly skilled ICU staff
and the hospital in order to match treatment to patient need. For hospi-
tal thirteen, this survey and a review of individual patient records docu-
mented a high frequency of complications, difficulties in managing
complex patients, and poor coordination between physicians and nurs-
ing personnel. Without information on severity, however, it would be
impossible to suggest that these differences in style of practice were asso-
ciated with variations in the quality of care, specifically the number of
critically ill patients discharged alive from the hospital.

In regard to utilization, hospital six admitted the largest proportion of
noncritically ill ICU patients with very low severity of illness. A mortality
standard based on diagnosis alone, however, indicated that the death
rate at hospital six was 110 percent better than expected—a finding which
is difficult to accept and which was not supported by our review of proc-
ess of care and a mortality projection based on diagnosis plus severity.

In contrast, hospital four had extraordinarily tight intensive care
unit admission criteria with very few stable noncritically ill patients that
were admitted for monitoring. The observed hospital death rate was 38
percent. Again, a diagnostic standard alone did not accurately capture
the complexity of this case-mix. A diagnostic standard would also have
completely missed the fact that, when controlling for severity as well as
diagnosis, hospital four performed better than most other hospitals. Informa-
tion on process of care indicated that hospital four was the one hospi-
tal most similar to hospital one in its processes and procedures. We believe this study suggests that, for ICU patients at least, quality of care differences and variations in patient selection criteria could confound hospital death rate information based on only diagnosis. A combination of diagnosis and admission severity appears to come closer to the truth.

Implications For Public Policy

The publication of hospital specific death rates could provide useful information on expected quality of care. It seems more rational to select hospitals for closure based on inadequate quality than inefficiency or inadequate capital financing. However, the accuracy of this information would be improved substantially by the development of more precise and reproducible measures of patient severity of illness.

Unfortunately, the data reported here is not directly applicable to PRO reporting for a number of reasons: the hospitals were self-selected and do not represent a national or regional standard; and the patients were only intensive care unit admissions with higher death rates and more serious physiologic instability than the average hospital patient. Nevertheless, we believe these results suggest that admission physiologic data could provide a more accurate standard for hospital-specific mortality rates than currently available from discharge abstracts for large numbers of hospital admissions.

This analysis was supported by grants from the National Center for Health Services Research and Health care Technology Assessment (#HS04857) and the Robert Wood Johnson Foundation (#8498).

NOTES