Estimating Hospital Deaths Due to Medical Errors

Preventability Is in the Eye of the Reviewer

Rodney A. Hayward, MD
Timothy P. Hofer, MD, MS

Context Studies using physician implicit review have suggested that the number of deaths due to medical errors in US hospitals is extremely high. However, some have questioned the validity of these estimates.

Objective To examine the reliability of reviewer ratings of medical error and the implications of a death described as "preventable by better care" in terms of the probability of immediate and short-term survival if care had been optimal.

Design Retrospective implicit review of medical records from 1995-1996.

Setting and Participants Fourteen board-certified, trained internists used a previously tested structured implicit review instrument to conduct 383 reviews of 111 hospital deaths at 7 Department of Veterans Affairs medical centers, oversampling for markers previously found to be associated with high rates of preventable deaths. Patients considered terminally ill who received comfort care only were excluded.

Main Outcome Measures Reviewer estimates of whether deaths could have been prevented by optimal care (rated on a 5-point scale) and of the probability that patients would have lived to discharge or for 3 months or more if care had been optimal (rated from 0%-100%).

Results Similar to previous studies, almost a quarter (22.7%) of active-care patient deaths were rated as at least possibly preventable by optimal care, with 6.0% rated as probably or definitely preventable. Interrater reliability for these ratings was also similar to previous studies (0.34 for 2 reviewers). The reviewers' estimates of the percentage of patients who would have left the hospital alive had optimal care been provided was 6.0% (95% confidence interval [CI], 3.4%-8.6%). However, after considering 3-month prognosis and adjusting for the variability and skewness of reviewers' ratings, clinicians estimated that only 0.5% (95% CI, 0.3%-0.7%) of patients who died would have lived 3 months or more in good cognitive health if care had been optimal, representing roughly 1 patient per 10000 admissions to the study hospitals.

Conclusions Medical errors are a major concern regardless of patients' life expectancies, but our study suggests that previous interpretations of medical error statistics are probably misleading. Our data place the estimates of preventable deaths in context, pointing out the limitations of this means of identifying medical errors and assessing their potential implications for patient outcomes.

JAMA. 2001;286:415-420 www.jama.com

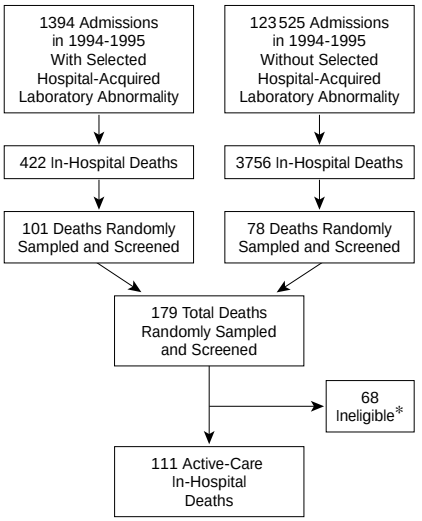
THE NUMBER OF DEATHS IN US hospitals that are reportedly due to medical errors is disturbingly high. A recent Institute of Medicine report quoted rates estimating that medical errors kill between 44000 and 98000 people a year in US hospitals.[1] These widely quoted statistics have helped create initiatives directed at patient safety throughout the United States. The numbers are undeniably startling; they suggest that more Americans are killed in US hospitals every 6 months than died in the entire Vietnam War, and some have compared the alleged rate to 3 fully loaded jumbo jets crashing every other day.[2] Widely disseminated quotes include, "medical mistakes kill 180000 people a year in US hospitals"[3] and "medical errors may be the 5th leading cause of death."[4] If these inferences are correct, the health care system is a public health menace of epidemic proportions.

These statistics are generally based on peer review using structured implicit review instruments. Physicians are trained to review hospital medical records and give their opinion on the occurrence of adverse events and the quality of hospital care and its impact on patient outcomes. Although the wording of the question used to assess hospital deaths has differed somewhat among studies, the studies have produced very similar conclusions. Perhaps the most often quoted study is the Harvard Medical Practice Study, which assessed negligence related to adverse events, including deaths, in New York.[5] However, several other studies have asked whether deaths would have been preventable by optimal quality of care[1,6-9] and have found similar results.

Author Affiliations are listed at the end of this article.
Corresponding Author and Reprints: Rodney A. Hayward, MD, PO Box 130170, Ann Arbor, MI 48113-0170 (e-mail: Rod.Hayward@med.va.gov).

©2001 American Medical Association. All rights reserved.

Figure. Patient Deaths Selected for Review



Patients with hospital-acquired digoxin toxicity, hyperkalemia, hypokalemia, hyponatremia, and renal failure were oversampled. Asterisk indicates 68 cases were determined to be ineligible because patients were admitted for comfort care (n=66) or it was unclear whether death had occurred during the acute hospital stay (n=2).

In an exchange about the validity of these estimates,[10,11] McDonald et al argued on theoretical grounds that these statistics are likely overestimates. They were particularly concerned about the lack of consideration of the expected risk of death in the absence of the medical error. Indeed, these statistics have often been quoted without regard to cautions by the authors of the original reports, who note that physician reviewers do not believe necessarily that 100% of these deaths would be prevented if care were optimal.[12] So, the questions remain: when a reviewer classifies a death as definitely or probably preventable or due to medical errors, is there a 90% chance or a 10% chance that a death would have actually been prevented if care had been optimal? How long would patients have lived if care had been optimal? How does the interrater reliability of reviewers' ratings affect these estimates? To examine these questions, we trained physician reviewers to assess medical records and identify medical errors documented in the care of patients who died at 7 Department of Veterans Affairs (VA) medical centers and asked reviewers to estimate the probability that these deaths could have been prevented by optimal medical care.

METHODS

A total of 4198 patients died at the 7 VA medical centers from 1995 to 1996 and were identified through a uniform hospital discharge data set. Cases with hospital-acquired renal failure, hyperkalemia, hypokalemia, hyponatremia, or digoxin toxicity were identified through the computerized laboratory system at each facility and were oversampled (representing 101 deaths [56%; FIGURE]), since previous research and a pilot study suggested that fluid and electrolyte abnormalities and drug toxicities have a higher rate of preventable death.[1,7,8] Random selection of these cases was stratified by hospital. A total of 201 cases were sampled and 179 (89%) were available for review. Initial screening was done by one of us (T.P.H.) and excluded 66 (37%) of the 179 patients who had died because they had been admitted for end-of-life comfort care. Of the 113 cases reviewed, 2 were excluded from analyses because it was unclear if death occurred during the acute inpatient stay. Study facilities ranged from those that had very close university affiliations to those that had no or only a loose university affiliation and ranged in size from about 3000 admissions per year to more than 13 000 admissions per year.

Fourteen board-certified internists with extensive experience in inpatient medicine were trained in the use of the implicit review instrument, reviewed sample charts, and discussed these reviews. After training established that the reviewers understood the review instrument and that disagreements were based on differences in opinion, not differences in understanding of the review instrument or overlooking information available in the chart, reviewers were allowed to review actual study charts.

Reviewers were blinded to the study question addressed in this article and which charts were selected for duplicate, independent review. Individual reviewers never reviewed the same chart twice and reviewers never reviewed charts of patients they had cared for. All charts were assigned to reviewers in a systematic fashion, with reviewers and those assigning charts blinded to results from previous reviews. Evidence of unbiased chart assignment includes no evidence of a substantial reviewer or temporal effect (meaning that average ratings of preventability did not significantly vary by individual reviewer or whether a review occurred earlier or later in the study period). The sample consisted of 383 reviews of 111 cases with 62 cases undergoing duplicate review, of which 33 had 2 reviews, 6 had 3 to 4 reviews, 8 had 7 to 8 reviews, 11 had 11 to 12 reviews, and 4 had 14 reviews. Of these, 35 cases had undergone duplicate review as part of a larger study on interrater reliability.[13]

The review instrument has been described previously[6] and is summarized briefly herein. In structured implicit review, reviewers are asked a series of questions about specific aspects of care, such as "the timeliness of diagnostic evaluation for presenting problem(s)". Near the end of the review, the reviewers for our study were asked, "Was the patient's death preventable by better quality of care?" and, in a separate question, were asked to rate the "overall quality of medical care." The structured approach focuses the reviewer on different aspects of care and is believed to make the reviews more valid and reliable.[5,6,8,9,14-18] The question on preventable death was rated on a 5-point scale (1=definitely; 2=probably; 3=uncertain; 4=probably not; 5=definitely not). The reviewers were also asked, "What do you estimate the likelihood of the prevention of death to be if care had been optimal?" rated from 0% to 100%. They were also asked to rate the probability, if care had been optimal, of the patient having left the hospital alive and having lived 3 months or more, and to estimate the probability that the patient would have had "good physical functioning" and "good cognitive functioning." The reviewers were told that good functioning corresponded to a level of function that would "allow a reasonable quality

 ©2001 American Medical Association. All rights reserved.

of life and meaningful social functioning." Reviewers were instructed, when assessing "better" or "optimal" care, not to use hindsight to second-guess reasonable clinical judgments but to focus on care that falls below standard of care. Furthermore, they were instructed not to be concerned about who was at fault or whether other aspects of care were good and were told that system errors, in which no single individual was at fault, should still be rated as errors.

Statistical Analysis

The reliability of reviewer ratings (ie, interrater reliability) was assessed by the intraclass correlation coefficient, derived from the within- and between-group variation in the hierarchical analyses.[13,19,20] The hierarchical model accounted for the unbalanced design (not all reviewers had reviewed all charts) and for the clustering of reviews by patient.[13,19,20] Rather than try to "resolve" disagreements by discussion, which previous research suggests is a flawed approach,[13] we examined the pattern of disagreements. The estimated number of preventable deaths was obtained by a weighted sum of each reviewer's estimate of survival to hospital discharge. The weights account for the number of reviews per patient and the sampling probabilities of each case. The SEs were adjusted for the clustering of reviews by patient. These analyses and the 95% confidence intervals (CIs) reflect the statistical power of our overall sampling frame, including the stratified random sampling of charts, the number of total charts reviewed, the number of charts that had duplicate reviews, and the total number of duplicate reviews. For all estimates of rates at the 7 hospitals, sampling weights were used to correct estimates for oversampling cases with hospital-acquired laboratory abnormalities and by hospital (Figure), although unadjusted data are also reported for the main results.

We conducted a Monte Carlo simulation of the effect of interrater reliability on estimates of the preventability of deaths. We estimated the expected mean and median reviewer estimates of preventability, simulating 100 reviewers per case by drawing repeatedly from the distribution of all estimated parameters in the random-effects hierarchical model used to estimate interrater reliability. The log odds of the reviewer estimates of preventability were normally distributed. Further details of the simulation are available from the authors. Analyses were conducted using Stata version 7.0 (Stata Corp, College Station, Tex) and MLwin version 1.02.002 (Multilevel Models Project, Institute of Education, London, England).

RESULTS

Characteristics of the patients who died in the hospital and were included in this study are shown in TABLE 1. The mean patient age at the time of death was 69 years but varied widely (SD, 11 years; range, 32-95 years). Although 67.6% of patients (n=75) had a do-not-resuscitate order at the time of death, only 13.5% had one within the first 2 days of admission.

Among the 383 reviews of the 111 deaths, overall care was rated as substandard in 7.0% of reviews and 6.0% of deaths. Care was rated as borderline in an additional 14.1% of reviews and 10.2% of deaths. Deaths were rated as having at least uncertain or possible preventability in 25.6% of reviews and 22.7% of deaths. Deaths were rated as definitely or probably preventable in 8.6% of reviews and 6.0% of deaths. These rates of reported quality and preventable deaths are similar to those found in previous reports.[5-7,9,13,15] The interrater reliability of ratings of whether deaths were related to errors was also similar to previous studies (intraclass correlation coefficient=0.34 for 2 reviewers compared with 0.24 in the Harvard Medical Practice Study).[5,9,13,14,22]

Although our study found an interrater reliability that is comparable with or better than that in most previous reports, it is not high. If one reviewer rated a death as definitely or probably preventable, the probability that the next reviewer would rate that case as definitely not preventable (18%) was actually slightly higher than the probability that the second reviewer would agree with the first (16%). (The probability that the next reviewer would rate the death as possibly preventable was 18%.) TABLE 2 shows the expected mean and median ratings of simulated reviewers for the 111 cases, produced by Monte Carlo simulation. Several results should be noted. First, the average rating of the probability of the patient leaving the hospital alive given optimal care did not differ substantially by quartile of preventability (estimated mean predicted preventability in the highest quartile of patients, 8.3%, and in the lowest quartile, 3.9%). Second, the mean estimates of preventable deaths were heavily influenced by outlier opinions of reviewers who believed that a major error had occurred, creating skewness, as shown by the median estimate of preventability being much lower than the mean estimate. For example, for the quartile of deaths with the highest preventability ratings, the median simulated reviewer would rate the probability of preventability as being only 2.2%, whereas the mean rating of 8.3% probability is heavily influenced by the 13.9% of reviewers who felt that optimal care had a greater than 50% chance of preventing the death (Table 2). Finally, even cases that were rated as having the lowest preventability still had some aspects of care that were rated as problematic by many reviewers. For example, for deaths in the lowest quartile of estimated preventability, the simulated mean rating

Table 1. Characteristics of Active-Care Patients Who Died in the Hospital (n = 111)*

Age, mean (SD), y	69 (11)
Male, No. (%)	110 (99)
Hospital length of stay, median (interquartile range), d	10 (4-24)
Admission source, No. (%)	
Home/relatives/friends	74 (66.7)
Assisted living	12 (10.8)
Other/uncertain	25 (22.5)
Do-not-resuscitate order within 1-2 d of admission, No. (%)	15 (13.5)
Major diagnostic category, No. (%)	
Circulatory DRG	34 (30.6)
Respiratory DRG	31 (27.9)
Digestive DRG	17 (15.3)
Other	29 (26.1)
Major surgical procedure, No. (%)	36 (32.4)

*Overall mortality rate at study hospitals was 3.4%, comparable with the 3.5% mortality rate found for adults aged 30 years or older in the National Discharge Survey in 1996.[21] DRG indicates diagnosis related group.

©2001 American Medical Association. All rights reserved.

Table 2. Influence of Interrater Reliability and Skewness on Estimates of Preventability of Death*

Patient Quartiles of Estimated Preventability	Stimulated Average Reviewer Report of Probability That Patient Would Have Left Hospital Alive With Optimal Care		
	Estimated Mean, %	Estimated Median, %	% of Reviewers Who Rated Preventability as ≥50%
Quartile 1 (25% of cases with lowest predicted preventability)	3.9	0.7	5.5
Quartile 2	4.9	1.0	8.3
Quartile 3	6.2	1.4	10.3
Quartile 4 (25% of cases with highest predicted preventability)	8.3	2.2	13.9

*Estimates are based on a Monte Carlo simulation using the level of reliability found in the hierarchical regression of 383 physician implicit reviews of 111 active-care deaths. This simulation gives the estimated results, based on the observed reliability, if 100 reviewers had reviewed each case.

Table 3. Reviewers' Estimates of Patient Prognosis and Probability That Death Was Preventable by Optimal Care

	Active-Care Deaths Estimated as Preventable, % (95% Confidence Interval)*	Preventable Deaths per 10 000 Admissions†
Rated as at least possibly preventable (unadjusted)	22.7 (13.5-31.8)‡	23-61
Adjusted for probability of leaving the hospital alive	6.0 (3.4-8.6)§	6-16
Adjusted for probability of leaving the hospital and		
Reliability and skewness of reviews	1.3 (1.0-1.5)	2-3
Reliability and probability of living ≥3 mo in good cognitive health	0.5 (0.3-0.7)	0.5-1.4

*Estimates are based on 383 physician implicit reviews of 111 active-care deaths.
†Estimates are based on the overall mortality rate at study hospitals, the percentage of deaths that received active care, and the 95% confidence intervals of the preventable death estimates.
‡Unweighted results showed that 98 (25.6%) of 383 reviews rated death as at least possibly preventable.
§Unweighted result was 6.4% (SD, 14%).

was that optimal care would have a 3.9% probability of preventing death and 5.5% of reviewers thought that optimal care would have reduced the chance of death by more than 50%. These findings suggest that given enough reviewers, almost all active-care deaths would have some reviewers who believe that an error caused the death, but they would usually represent an outlier opinion.

TABLE 3 demonstrates the impact of adjusting for various sources of potential error in estimating preventable deaths. While 22.7% of deaths reviewed (unweighted, unadjusted estimate [SD], 20.2% [21%]) were estimated as being at least possibly preventable by optimal care, the mean reviewer estimate (for all cases) was that optimal care would have resulted in only 6.0% of patients leaving the hospital alive (unweighted, unadjusted estimate [SD], 6.4% [14%]; Table 3). The estimate was lower when discharge from the hospital was taken into account because when reviewers rated a case as at least possibly preventable, these same reviewers reported, on average, that there was only a 20% (95% CI, 12%-27%) chance that these patients would have left the hospital alive if care had been optimal (unweighted, unadjusted estimate (SD), 20% [21%]). Even when the analysis was limited to only those cases rated as definitely or probably preventable, the mean reviewer estimate of the likelihood that these patients would have left the hospital alive given optimal care was 43% (95% CI, 35%-51%; unweighted, unadjusted estimate [SD], 39% [24%]). Therefore, when a reviewer rated a death as "preventable," that physician reviewer believed that optimal care would have prevented the death and the patient would have survived to discharge less than half of the time.

Table 3 also shows the effect of adjusting the overall estimates for the variability and skewness in reviewer ratings. Using the estimated median rating, rather than the mean, to adjust for the reliability and skewness of reviewer ratings reduces the estimate that the death could have been prevented from 6.0% to 1.3%. Finally, past studies have not considered the underlying prognosis and health of the patients who died. Reviewers estimated that only about one third of the patients judged to survive to discharge with optimal care would have been expected to live 3 months or longer in good cognitive health, or 0.5% of all deaths (Table 3). This would suggest, based on these physician reviews, that optimal care at the study hospitals would result in roughly 1 additional patient of every 10 000 admissions living 3 months or more in good cognitive health.

COMMENT

Studies using implicit review to estimate the impact of medical errors on hospital deaths have been widely quoted and have generated national policy proposals and debate. Our reviewers estimated similar numbers of preventable deaths as that of previous studies, including rating almost a quarter of hospital deaths as at least possibly preventable.[5-7,9,15,23] However, this is the first study to our knowledge to question reviewers about the likelihood of death in the absence of the error, to examine the patients' underlying short-term prognosis, and to consider the effect of variability in reviewers' ratings on these estimates.

As predicted on theoretical grounds,[10-12,24] many deaths reportedly due to medical errors occur at the end of life or in critically ill patients in whom death was the most likely outcome, either during that hospitaliza-

©2001 American Medical Association. All rights reserved.

tion or in the coming months, regardless of the care received. However, this was not the only—or even the largest—source of potential overestimation. Previously, most have framed ratings of preventable deaths as a phenomenon in which a small but palpable number of deaths have clear errors that are being reliably rated as causing death. Our results suggest that this view is incorrect—that if many reviewers evaluate charts for preventable deaths, in most cases some reviewers will strongly believe that death could have been avoided by different care; however, most of the "errors" identified in implicit chart review appear to represent outlier opinions in cases in which the median reviewer believed either that an error did not occur or that it had little or no effect on the outcome.

These results do not suggest that medical errors are unimportant. Simply because implicit review suggests that errors may rarely result in preventable deaths does not excuse mistakes or suggest that they are inconsequential. First, we only evaluated the fatal complications; morbidity due to medical errors and the resultant costs are undoubtedly manyfold greater than the number of preventable deaths.[1,5,11,23] Second, this study did not evaluate errors after hospital discharge or in the outpatient setting, and many hospital errors are likely unidentifiable in the medical record.[11] Third, whether errors warrant systems changes should not be based on the overall impact of all errors but, rather, on a careful examination of specific errors and the effectiveness and costs of a policy directed at error prevention. There are other reasons to be cautious in interpreting our study's results. These VA hospitals cannot be assumed to be representative of US hospitals in general. If these hospitals cared for sicker patients or have better-than-average quality and patient care, the number of preventable deaths could be underestimated. Although the overall mortality rates and the preventable death rate estimates are very similar to those in previous studies, VA hospitals do tend to care for sicker and older patients, and this could have affected our results related to short-term survival. However, this would not affect the adjustment of estimates for reviewer reliability and skewness, and it was this source of overestimation that had the largest effect on the preventable death estimates in our study (Table 3).

Although our study helps clarify some issues regarding medical errors, whether physician reviewers can accurately make such assessments from the medical record remains uncertain. Our study uses the same basic methods as previous studies, structured implicit review, and suggests that if this is accepted as a valid way of addressing this issue, statistics taken from previous studies[1] are probably overestimated. We agree with investigators[10,12,24] who note that we must be very cautious in making causal assertions from retrospective reviews. However, we are not confident that currently available instruments to adjust for severity of illness are adequate to assess the overall impact of medical errors on outcomes (although severity adjustment and rigorous methods may help produce estimates for specific processes of care).[8] Given the complexity of hospital care, in the foreseeable future implicit review may be the best source of estimating the overall impact of errors.

Implicit review could underestimate medical errors. Reviewers may be reluctant to second-guess the care of fellow clinicians, and many errors may not be documented in the medical record or identifiable by chart review.[11,25,26] Our study also may overestimate the consequences of medical errors. First, although we instructed our reviewers to not second-guess reasonable clinical judgments, hindsight bias is part of human nature[27] and empirical evidence exists that this occurs in physician implicit review.[28] Unlike the clinicians who cared for these patients, our chart reviewers had the advantage of knowing the final diagnoses and outcomes. Chart reviewers may consciously or subconsciously allow this privileged knowledge to result in second-guessing reasonable decisions and inflate the true merits of alternative choices and decisions. Another possible bias for reviewers' estimates is that physicians tend to overestimate how long sick patients will live, often dramatically so.[29-33] Although the previous studies were conducted on physicians who were providing care to the patients,[29-33] if our chart reviewers, who did not know the patients, similarly overestimated the probability of short-term survival, this would result in further overestimation of the impact of optimal care on truly preventable deaths.

The statistics on preventable deaths have captured the public's attention and, to the extent that the current patient safety initiative fosters an efficient and effective approach to error reduction, it has great promise to improve the health care system and produce positive outcomes. However, as demonstrated by this study, the statistics that brought much of this attention do not support the tenet that hospitals are unsafe for patients, as some interpretations of these statistics have suggested. Furthermore, while some well-publicized cases[1] have been patients with long life expectancies, if our results can be generalized to other hospitals, they suggest that most of the cases that make up the dramatic statistics occur in substantially different situations. While deaths due to medical errors are still extremely important even when patients have very short life expectancies, the correct understanding of these errors may differ substantially from how they have been publicly portrayed to date.

Our study also suggests that finding patterns of care that result in truly preventable deaths may prove more difficult than previously believed. It is sometimes implied that the egregious errors that make the media headlines (like unintentionally amputating the wrong leg) are representative of the types of errors found in implicit review studies. If that were true, the interrater reliability of implicit review should be much greater than 0.25 for 2 reviewers. In all general medical and surgical chart review studies to date,[5-7,9,13-15] reviewers have had a difficult time agreeing on whether an error caused an adverse event or even on whether something was an error at all.

©2001 American Medical Association. All rights reserved.

Reviewer agreement is usually even worse when specific processes of care are evaluated (as opposed to overall care)[6] and attempts at improving the true reliability of implicit review by discussion between reviewers have been unsuccessful.[13] Under such circumstances, finding patterns can prove difficult, and trying to fix problems in complex settings using hindsight and anecdotes can lead to changes that may increase, not decrease, errors.[34,35] Finally, these results have direct implications for using risk-adjusted hospital mortality rates to assess hospital quality. Past research suggests that the correlation between ratings of "preventable deaths" and actual prevention of deaths would have to be very high for disease-specific hospital mortality rates to be an accurate measure of hospital quality.[36]

In conclusion, we found that our physician reviewers often reported medical errors and frequently reported deaths as being preventable by better care (at a rate similar to previous studies). However, 3 caveats were identified that have implications for preventable deaths: (1) the probability that the error actually caused the death was often considered to be low; (2) reviewer assessment of errors had poor reliability and was usually skewed; and (3) the underlying short-term prognosis of the person who died was often judged to be very limited. Medical errors are undoubtedly common and contribute to many adverse outcomes. However, if our results can be generalized to other hospitals, the statistics on deaths due to medical errors do not accurately reflect the view of most physician chart reviewers. Our results suggest that these statistics are probably unreliable and have substantially different implications than has been implied in the media and others. Most importantly, this study demonstrates the limitations of this means of identifying errors and highlights that caution is warranted when establishing causal relationships between errors and patient outcomes.

Author Affiliations: Department of Veterans Affairs, VA Center for Practice Management and Outcomes Research, VA Ann Arbor Healthcare System, and Departments of Internal Medicine and Health Management and Policy, University of Michigan Schools of Medicine and Public Health, Ann Arbor.
Author Contributions: *Study concept and design:* Hayward, Hofer.
Analysis and interpretation of data: Hayward, Hofer.
Drafting of the manuscript: Hayward, Hofer
Critical revision of the manuscript for important intellectual content: Hayward, Hofer.
Statistical expertise: Hayward, Hofer.
Obtained funding: Hayward, Hofer.
Administrative, technical, or material support: Hayward, Hofer.
Study supervision: Hayward.
Funding/Support: This study was supported by VA Health Services Research and Development Service Merit Review Grant 94-131. Dr Hofer is supported by a VA Career Development Grant from Health Services Research and Development, Office of the Department of Veterans Affairs.

REFERENCES

1. Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
2. Preventing fatal medical errors. *New York Times.* December 1, 1999:A22.
3. "CNN Headline News." Factoid. January 16, 1999.
4. Weiss R. Medical errors blamed for many deaths; as many as 98,000 a year in US linked to mistakes. *Washington Post.* November 30, 1999:A1.
5. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991;324:370-376.
6. Hayward RA, McMahon LF, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? *Ann Intern Med.* 1993;118:550-556.
7. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med.* 1988;109:582-589.
8. Hofer TP, Bernstein SJ, Hayward RA, DeMonner S. Validating quality indicators for hospital care. *Jt Comm J Qual Improv.* 1997;23:455-467.
9. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers: how well do peer review organization methods detect hospital care quality problems? *JAMA*. 1992;267:2349-2354.
10. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. *JAMA*. 2000;284:93-95.
11. Leape LL. Institute of Medicine medical error figures are not exaggerated. *JAMA*. 2000;284:95-97.
12. Brennan TA. The Institute of Medicine report on medical errors—could it do harm? *N Engl J Med*. 2000;342:1123-1125.
13. Hofer TP, Bernstein SJ, DeMonner S, Hayward RA. Discussion between reviewers does not improve reliability of peer review of hospital quality. *Med Care.* 2000;38:152-161.
14. Goldman RL. The reliability of peer assessments of quality of care. *JAMA*. 1992;267:958-960.
15. Hayward RA, Bernard AM, Rosevear JS, Anderson JE, McMahon LF. An evaluation of generic screens for poor quality of hospital care on a general medicine service. *Med Care*. 1993;31:394-402.
16. Rubenstein LV, Kahn KL, Reinisch EJ, et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. *JAMA*. 1990;264:1974-1979.
17. Kahn KL, Rubenstein LV, et al. *Structured Implicit Review for Physician Measurement of Quality of Care: Development of the Form and Guidelines for Its Use*. Santa Monica, Calif: RAND Corp; 1989.
18. Butler JJ, Quinlan JW. Internal audit in the department of medicine of a community hospital: two years' experience. *JAMA*. 1958;167:567-572.
19. Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. *J Clin Epidemiol.* 1992;44:381-390.
20. Evans WJ, Cayten CG, Green PA. Determining the generalizability of rating scales in clinical settings. *Med Care.* 1981;19:1211-1220.
21. National Hospital Discharge Survey, 1970-1998. Available at: http://www.sscnet.ucla.edu/issr/da/index/techinfo/h04001.htm. Accessed April 30, 2001.
22. Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgments concerning adverse events suffered by hospitalized patients. *Med Care.* 1989;27:1148-1158.
23. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery*. 1999;126:66-75.
24. Sox HC, Woloshin S. How many deaths are due to medical error? getting the number right. *Effective Clin Pract.* 2000;6:277-283.
25. Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events: implications for prevention. *JAMA*. 1995;274:29-34.
26. Andrews LB, Stocking C, Krizek T, et al. An alternative strategy for studying adverse events in medical care. *Lancet.* 1997;349:309-313.
27. Thompson SC, Armstrong W, Thomas C. Illusions of control, underestimations, and accuracy: a control heuristic explanation. *Psychol Bull*. 1998;123:143-161.
28. Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgments of appropriateness of care. *JAMA*. 1991;265:1957-1960.
29. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ*. 2000;230:469-473.
30. Christakis NA. Predicting patient survival before and after hospice enrollment. *Hosp J.* 1998;13:71-87.
31. Addington-Hall JM, MacDonald LD, Anderson HR. Can the Spitzer quality of life index help to reduce prognostic uncertainty in terminal care? *Br J Cancer.* 1990;62:695-699.
32. Evans C, McCarthy M. Prognostic uncertainty in terminal care: can the Karnofsky index help? *Lancet.* 1985;148:1204-1206.
33. Forster LE, Lynn J. Predicting life span for applicants to inpatient hospice. *Arch Intern Med.* 1988;148:2540-2543.
34. Perrow C. *Normal Accidents: Living With High-Risk Technologies*. 2nd ed. Princeton, NJ: Princeton University Press; 1999.
35. Hofer TP, Kerr EA, Hayward RA. What is an error? *Effective Clin Pract.* 2000;3:1-10.
36. Hofer TP, Hayward RA. Identifying poor quality hospitals: can hospital mortality rates be useful? *Med Care.* 1996;34:737-753.

©2001 American Medical Association. All rights reserved.