

SPECIAL ARTICLE

Temporal Trends in Rates of Patient Harm Resulting from Medical Care

Christopher P. Landrigan, M.D., M.P.H., Gareth J. Parry, Ph.D.,
Catherine B. Bones, M.S.W., Andrew D. Hackbarth, M.Phil.,
Donald A. Goldmann, M.D., and Paul J. Sharek, M.D., M.P.H.

ABSTRACT

BACKGROUND

From the Division of Sleep Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School (C.P.L.); and the Divisions of General Pediatrics (C.P.L., G.J.P.) and Infectious Disease (D.A.G.), Department of Medicine, Children's Hospital Boston and Harvard Medical School — all in Boston; the Institute for Healthcare Improvement, Cambridge, MA (G.J.P., C.B.B., A.D.H., D.A.G.); the Pardee RAND Graduate School, Santa Monica, CA (A.D.H.); and the Division of General Pediatrics, Department of Pediatrics, Lucile Packard Children's Hospital and Stanford University School of Medicine, Stanford, CA (P.J.S.). Address reprint requests to Dr. Landrigan at the Division of Sleep Medicine, Department of Medicine, Brigham and Women's Hospital, 221 Longwood Ave., Boston, MA 02115, or at clandrigan@partners.org.

In the 10 years since publication of the Institute of Medicine's report *To Err Is Human*, extensive efforts have been undertaken to improve patient safety. The success of these efforts remains unclear.

METHODS

We conducted a retrospective study of a stratified random sample of 10 hospitals in North Carolina. A total of 100 admissions per quarter from January 2002 through December 2007 were reviewed in random order by teams of nurse reviewers both within the hospitals (internal reviewers) and outside the hospitals (external reviewers) with the use of the Institute for Healthcare Improvement's Global Trigger Tool for Measuring Adverse Events. Suspected harms that were identified on initial review were evaluated by two independent physician reviewers. We evaluated changes in the rates of harm, using a random-effects Poisson regression model with adjustment for hospital-level clustering, demographic characteristics of patients, hospital service, and high-risk conditions.

RESULTS

Among 2341 admissions, internal reviewers identified 588 harms (25.1 harms per 100 admissions; 95% confidence interval [CI], 23.1 to 27.2). Multivariate analyses of harms identified by internal reviewers showed no significant changes in the overall rate of harms per 1000 patient-days (reduction factor, 0.99 per year; 95% CI, 0.94 to 1.04; $P=0.61$) or the rate of preventable harms. There was a reduction in preventable harms identified by external reviewers that did not reach statistical significance (reduction factor, 0.92; 95% CI, 0.85 to 1.00; $P=0.06$), with no significant change in the overall rate of harms (reduction factor, 0.98; 95% CI, 0.93 to 1.04; $P=0.47$).

CONCLUSIONS

In a study of 10 North Carolina hospitals, we found that harms remain common, with little evidence of widespread improvement. Further efforts are needed to translate effective safety interventions into routine practice and to monitor health care safety over time. (Funded by the Rx Foundation.)

This article was updated on November 24, 2010, at nejm.org.

N Engl J Med 2010;363:2124-34.
Copyright © 2010 Massachusetts Medical Society.

IN DECEMBER 1999, THE INSTITUTE OF MEDICINE (IOM) reported that medical errors cause up to 98,000 deaths and more than 1 million injuries each year in the United States.¹ In response, accreditation bodies, payers, non-profit organizations, governments, and hospitals launched major initiatives and invested considerable resources to improve patient safety.²⁻⁴ Some interventions have been shown to reduce errors, such as implementing computerized provider order-entry systems,^{5,6} limiting residents' work shifts to 16 consecutive hours,⁷⁻⁹ and implementing evidence-based care bundles.^{10,11} However, many of these interventions have not been evaluated rigorously¹² or implemented reliably on a large scale.¹³⁻¹⁶ Unfortunately, it remains unclear whether, in the aggregate, efforts to reduce errors at national, regional, and local levels have translated into significant improvements in the overall safety of patients.

To address this persistent uncertainty,^{17,18} we sought to determine whether statewide rates of harm have been decreasing over time in North Carolina. We chose North Carolina as a site that was likely to have improvement, since it had shown a high level of engagement in efforts to improve patient safety, including a 96% rate of hospital enrollment in a previous national improvement campaign, as compared with an average rate of 78% in other states,^{19,20} and extensive participation in statewide safety training programs and improvement collaboratives.¹⁹

METHODS

STUDY DESIGN

We applied the Institute for Healthcare Improvement's Global Trigger Tool for Measuring Adverse Events to randomly selected medical records of patients who had been discharged between January 2002 and December 2007 in 10 randomly selected hospitals in North Carolina. During the past few years, trigger tools (instruments that facilitate efficient, focused reviews of medical records) have been developed to measure rates of harm resulting from medical care.^{21,22} The trigger tool was developed to provide a reliable hospital-based measure for tracking rates of harm over time.^{23,24}

Data collection and initial analyses were over-

seen by a clinical research organization, Battelle Health and Life Sciences Global Business. We obtained approval for the study from the institutional review boards at Battelle and participating hospitals. A detailed description of the study methods has been reported previously.²⁵ The requirement for written informed consent was waived by the institutional review board, since the study was retrospective and involved record review only.

The study was supported by a grant from the Rx Foundation, which had no role in the design of the study; the collection, analysis, or interpretation of the data; or approval of the manuscript.

HOSPITAL SELECTION

All acute care North Carolina hospitals listed in the American Hospital Association (AHA) database except those providing exclusively pediatric, rehabilitation, or psychiatric care were eligible for selection for the study. These hospitals were stratified according to the AHA's definition of the facility as small, medium, or large; urban or rural; and teaching or nonteaching. The number of hospitals that underwent randomization for inclusion in each stratum reflected the proportion of national discharges from that type of hospital. If an invited hospital declined to participate, another closely matched hospital was randomly invited to participate in its stead.

RECORD SELECTION

In each hospital, 10 randomly selected admissions of at least 24 hours in each quarter from January 2002 through December 2007 (240 records per hospital) were reviewed. The records of patients who were under the age of 18 years and those who were admitted primarily for psychiatric or rehabilitation care were excluded. Reviews of the records with the use of the trigger tool were conducted both by a team of hospital-based (internal) reviewers, who worked in the hospitals where they reviewed charts, and a team of external reviewers, who worked elsewhere and were hired and supervised by Battelle. Both internal and external teams were made up of primary reviewers, typically nurses, and secondary physician reviewers with expertise in hospital care. Internal and external teams were trained in an identical manner, with a standardized series of

Web-based seminars, provided by patient-safety experts and experienced reviewers, that included didactic sessions, practical review exercises, and debriefing sessions.²⁵

RECORD-REVIEW PROCESS

Internal and external review teams independently conducted two-stage reviews of the same records in each hospital. Within each team, a primary reviewer conducted a review of each record using the trigger tool, which consists of 52 triggers, or clues, in patient records that indicate the possibility of medically induced harm. When primary reviewers found a trigger (e.g., administration of naloxone, which is often used to reverse the effects of an inadvertent narcotic overdose), they investigated the chart further to determine whether harm resulting from medical care had apparently occurred. Injuries associated with previous treatment that were identified as present at admission, as well as those that occurred during the index hospitalization, were captured in an effort to determine the total burden of harm resulting from medical care.

The primary review of each record was performed with the use of the trigger tool in a standardized fashion in 20 minutes or less. The order of record review by primary reviewers was randomized (i.e., reviews were not conducted in order of admission date) to prevent any distortion in the results over time by the reviewers' gradual accumulation of experience with the trigger tool. In addition, dates of hospitalization were concealed from the reviewers to prevent any bias in chart review (e.g., the possibility that internal reviewers might have a bias toward seeing improvement over time).

Primary reviewers prepared one- to two-paragraph summaries of all suspected harms, which were presented in a second stage to two independent physician reviewers, who were likewise unaware of dates of hospitalization. The physician reviewers made final determinations about the presence, severity, and preventability of any suspected harms identified. We used the index of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP)²⁶ to evaluate severity, with lower-severity harms defined as those in category E (temporary harms requiring intervention), and higher-severity harms defined as those in category F (temporary harms

requiring initial or prolonged hospitalization), category G (permanent harms), category H (life-threatening harms), or category I (harms causing or contributing to death). Examples of harms in each of the NCC MERP Index categories are provided in the Supplementary Appendix, available with the full text of this article at NEJM.org. We used a Likert scale (with scores ranging from 1 for "definitely not preventable" to 4 for "definitely preventable") to evaluate preventability. Cases in which physician reviewers disagreed were discussed, and consensus was achieved. Interrater reliability was calculated from prediscussion ratings.

RELIABILITY

We assessed the reliability of the abstraction and rating process through multiple checks of interrater and intrarater reliability for each stage of review. In within-team checks on seven of seven reliability tests, internal review teams performed more reliably, with kappa scores for reliability ranging from 0.64 (substantial) to 0.93 (almost perfect), than did external reviewers, with kappa scores ranging from 0.40 (moderate) to 0.72 (substantial).²⁵ Kappa scores for preventability ratings were 0.83 for internal reviewers and 0.54 for external reviewers.

In addition, as previously reported,²⁵ a team of expert reviewers with extensive experience with the trigger tool reviewed a 10% sample of records from each hospital to provide a metric by which to adjudicate any differences in findings between teams. Internal reviewers and experienced reviewers agreed about the presence of harm in 81% of reviews (kappa score, 0.49), as compared with 75% agreement (kappa score, 0.32) between external reviewers and experienced reviewers. Likewise, internal reviewers had a higher kappa score for agreement with experienced reviewers on ratings of severity than did external reviewers (0.53 vs. 0.26).²⁵

STATISTICAL ANALYSIS

We used a Poisson regression model with random effects to account for hospital-level clustering and a term indicating the hospital-admission date (24 quarters during a 6-year period) in order to assess changes in the rate of harm (number of harms per 1000 patient-days and per 100 admissions) over time. To account for the possibility

that changes in harm rates over time were confounded by changes in demographic characteristics of patients or in the severity of illness, we conducted additional Poisson regression analyses, adding terms to adjust for sex, age, race, insurance group, and whether the patient was admitted to an intensive care unit, obstetrical or gynecologic service, or surgical service or had a high risk of harm. We calculated the risk of harm using the Clinical Classification Software of the Agency for Healthcare Research and Quality (AHRQ) to group codes from the *International Classification of Diseases, 9th Revision (ICD-9)* into 200 groups. A high risk of harm was defined as 1 of 20 ICD-9 codes (principal diagnosis) that were associated with at least 50% of the harms in the aggregated data from all 6 years.

On the basis of an anticipated 40 harms per 100 admissions,²¹ the study had a power of 80% to detect a decreasing trend in harms equivalent to a reduction in harms from 40 per 100 admissions in 2001 to 30 per 100 admissions in 2007. A two-sided P value of less than 0.05 was considered to indicate statistical significance.

RESULTS

NUMBER, TYPE, AND SEVERITY OF HARMS

We invited 14 hospitals to participate in the study in order to reach the enrollment goal of 10 hospitals (71% participation rate). Internal teams completed 2341 of 2400 planned record reviews (97.5%) in the 10 study hospitals. A total of 588 harms were identified for 10,415 patient-days that were studied, for a rate of 56.5 harms (95% confidence interval [CI], 52.0 to 61.2) per 1000 patient-days or 25.1 harms (95% CI, 23.1 to 27.2) per 100 admissions. These harms occurred in 423 unique patient admissions (18.1%). Harms that were detected were a consequence of procedures (186), medications (162), nosocomial infections (87), other therapies (59), diagnostic evaluations (7), and falls (5), among other causes (Table 1).

Of 588 harms that were identified, 245 (41.7%) were temporary harms requiring intervention (category E on the NCC MERP Index), and 251 (42.7%) were temporary harms requiring initial or prolonged hospitalization (category F). An additional 17 harms (2.9%) were permanent (category G), 50 (8.5%) were life-threaten-

ing (category H), and 14 (2.4%) caused or contributed to a patient's death (category I) (Fig. 1). A total of 4.4 harms per 100 admissions (17.9%) were present on admission; the remainder, 20.7 per 100 admissions (82.3%), occurred during the studied hospital admission.

External teams completed 2374 of the 2400 planned record reviews (98.9%), identifying 429 harms during 10,675 patient-days, for a rate of 40.2 harms (95% CI, 36.5 to 44.2) per 1000 patient-days (Fig. 1).

PREVENTABLE HARMS

We conducted an analysis of preventable harms on the basis of 588 harms that were identified with the use of the trigger tool. Among these harms, internal reviewers rated 364 (63.1%) as preventable (Table 1). The large majority of identified harms were classified as category E (144) or category F (163) harms. Of the identified preventable harms, 13 caused permanent harm (category G), 35 were life-threatening (category H), and 9 caused or contributed to a patient's death (category I).

CHANGES IN RATE OF HARMS OVER TIME

There was no significant change over time in the rate of harms identified by internal reviewers. Poisson regression that accounted for hospital-level clustering and changes over time showed a nonsignificant 1% reduction per year in the rate of harms per 1000 patient-days (reduction factor, 0.99; 95% CI, 0.95 to 1.04; $P=0.72$) (Fig. 2A). The rate of harms per 100 admissions likewise did not change significantly (Fig. 3A). Moreover, subanalyses of changes in preventable harms (reduction factor, 0.99; 95% CI, 0.93 to 1.05; $P=0.77$) and harms of higher severity (NCC MERP categories F through I) revealed no significant differences over time in rates per 1000 patient-days (Fig. 2C and 2E, respectively) or rates per 100 admissions (Fig. 3C and 3E, respectively).

External reviewers identified fewer harms overall than did internal reviewers, with no significant change over time in the overall rate of harms per 1000 patient-days (reduction factor, 0.97; 95% CI, 0.92 to 1.03; $P=0.33$) (Fig. 2B) or the rate per 100 admissions (Fig. 3B). The rate of preventable harms identified by external reviewers, unadjusted for covariates and risk factors, was reduced from 23.5 harms per 1000 patient-days in 2002 to 15.0 harms per 1000 patient-days

Table 1. All Harms and Preventable Harms, According to Category of Severity, as Reported by Internal Reviewers.*

Type of Harm	All Harms						Preventable Harms					
	E	F	G	H	I	Total	E	F	G	H	I	Total
	<i>number</i>											
Cardiovascular system												
Total events	21	10	1	12	1	45	7	7	1	6	1	22
Cardiac arrest	0	0	0	1	1	2	0	0	0	0	1	1
Hypotension	11	6	1	6	0	24	4	4	1	4	0	13
Hypertension	0	0	0	1	0	1	0	0	0	0	0	0
Shock	0	0	0	1	0	1	0	0	0	1	0	1
Arrhythmias or conduction abnormality	6	1	0	2	0	9	1	0	0	0	0	1
Myocardial ischemia	0	0	0	1	0	1	0	0	0	1	0	1
Other cardiovascular event	4	3	0	0	0	7	2	3	0	0	0	5
Respiratory system												
Total events	7	16	0	17	1	41	4	10	0	13	0	27
Acute respiratory failure	1	2	0	7	0	10	1	2	0	4	0	7
Respiratory distress, not acute failure	0	2	0	4	0	6	0	1	0	3	0	4
Pneumothorax	1	4	0	1	0	6	0	3	0	1	0	4
Atelectasis	1	0	0	0	0	1	1	0	0	0	0	1
Bronchospasm	0	1	0	0	0	1	0	0	0	0	0	0
Aspiration	0	2	0	0	0	2	0	1	0	0	0	1
Pulmonary embolus	1	3	0	2	1	7	0	2	0	2	0	4
Need for reintubation	1	0	0	3	0	4	0	0	0	3	0	3
Other respiratory event	2	2	0	0	0	4	2	1	0	0	0	3
Renal or endocrine system												
Total events	26	17	2	4	3	52	21	15	2	3	2	43
Fluid overload	2	3	0	1	0	6	2	3	0	0	0	5
Dehydration or oliguria	0	2	0	0	0	2	0	1	0	0	0	1
Acute renal failure	1	2	1	0	2	6	0	2	1	0	1	4
Metabolic acidosis	0	0	0	0	1	1	0	0	0	0	1	1
Hyperglycemia	1	1	0	0	0	2	0	1	0	0	0	1
Hypoglycemia	17	1	0	2	0	20	16	1	0	2	0	19
Hyperkalemia	3	1	0	0	0	4	2	1	0	0	0	3
Other renal or endocrine event	2	7	1	1	0	11	1	6	1	1	0	9
Hematologic system												
Total events	25	27	0	0	1	53	12	19	0	0	0	31
Hemorrhage	18	9	0	0	0	27	10	7	0	0	0	17
Thromboembolic venous event	1	2	0	0	0	3	0	1	0	0	0	1
Hematoma	3	2	0	0	0	5	1	1	0	0	0	2
Other hematologic event	3	14	0	0	1	18	1	10	0	0	0	11
Gastrointestinal system												
Total events	11	26	0	2	0	39	4	10	0	1	0	15
Nausea or vomiting	3	9	0	0	0	12	0	1	0	0	0	1
Diarrhea	1	1	0	0	0	2	1	0	0	0	0	1
Constipation	1	0	0	0	0	1	1	0	0	0	0	1
Gastric distention	0	1	0	0	0	1	0	1	0	0	0	1
Pancreatitis	0	1	0	0	0	1	0	1	0	0	0	1
Ileus	0	7	0	0	0	7	0	2	0	0	0	2
Other gastrointestinal event	6	7	0	2	0	15	2	5	0	1	0	8

Table 1. (Continued.)												
Type of Harm	All Harms						Preventable Harms					
	E	F	G	H	I	Total	E	F	G	H	I	Total
	<i>number</i>											
Neurologic system												
Total events	20	20	3	0	1	44	6	12	3	0	1	22
Oversedation	5	9	0	0	0	14	2	7	0	0	0	9
Delirium or encephalopathy	2	2	0	0	0	4	1	0	0	0	0	1
Seizure	1	1	0	0	0	2	0	1	0	0	0	1
Stroke or intracerebral hemorrhage	0	0	1	0	0	1	0	0	1	0	0	1
Inadequate analgesia	0	1	0	0	0	1	0	1	0	0	0	1
Withdrawal symptoms	0	1	0	0	0	1	0	1	0	0	0	1
Other neurologic event	12	6	2	0	1	21	3	2	2	0	1	8
Hospital-acquired infection												
Total infections	39	61	0	3	7	110	30	44	0	3	5	82
Catheter-related bloodstream infection	4	5	0	0	0	9	4	4	0	0	0	8
Sepsis or bacteremia unrelated to catheter	2	7	0	0	1	10	2	5	0	0	1	8
Ventilator-associated pneumonia	0	6	0	0	2	8	0	4	0	0	2	6
Nosocomial pneumonia, not ventilator-related	1	7	0	0	3	11	1	6	0	0	2	9
Urinary tract infection	20	9	0	2	0	31	17	5	0	2	0	24
Surgical-site infection	3	14	0	0	0	17	1	9	0	0	0	10
Endometritis	0	1	0	0	0	1	0	1	0	0	0	1
<i>Clostridium difficile</i> colitis	2	3	0	0	0	5	0	2	0	0	0	2
Other hospital-acquired infection	7	9	0	1	1	18	5	8	0	1	0	14
Surgical or obstetrical event												
Total events	29	40	6	10	0	85	15	24	3	7	0	49
Postoperative hemorrhage	4	0	0	2	0	6	3	0	0	2	0	5
Postoperative hematoma	1	0	0	0	0	1	0	0	0	0	0	0
Laceration or other organ injury	13	2	0	3	0	18	5	2	0	1	0	8
Unplanned removal of organ after intraoperative injury	1	0	1	0	0	2	1	0	1	0	0	2
Vascular injury	0	1	0	1	0	2	0	1	0	1	0	2
Nerve injury	0	0	1	0	0	1	0	0	1	0	0	1
Surgical anastomosis failure	0	3	1	1	0	5	0	2	0	1	0	3
Wound dehiscence	0	2	0	0	0	2	0	2	0	0	0	2
Failed procedure	0	6	0	0	0	6	0	2	0	0	0	2
Unplanned return to surgery	0	14	0	2	0	16	0	6	0	2	0	8
Fetal neonatal complication associated with delivery	2	1	0	0	0	3	0	1	0	0	0	1
Other event	8	11	3	1	0	23	6	8	1	0	0	15
Other types of harm												
Total events	68	41	6	4	0	119	45	22	4	2	0	73
Hypothermia	1	0	0	0	0	1	1	0	0	0	0	1
Pyrexia	0	1	0	0	0	1	0	0	0	0	0	0
Alcohol or drug withdrawal	0	3	0	0	0	3	0	1	0	0	0	1
Allergic reaction	7	2	0	0	0	9	4	0	0	0	0	4
Fall	2	5	1	0	0	8	1	5	1	0	0	7
Pressure ulcer	29	4	2	0	0	35	28	4	2	0	0	34
Rash	3	1	0	0	0	4	0	0	0	0	0	0
Catheter complication	6	2	0	0	0	8	4	2	0	0	0	6
Other type of harm	20	23	3	4	0	50	7	10	1	2	0	20

* The severity categories used by the National Coordinating Council for Medication Error Reporting and Prevention Index are as follows: E, temporary harm to the patient requiring intervention; F, temporary harm to the patient requiring initial or prolonged hospitalization; G, permanent harm to the patient; H, intervention required to sustain life; and I, death of the patient.

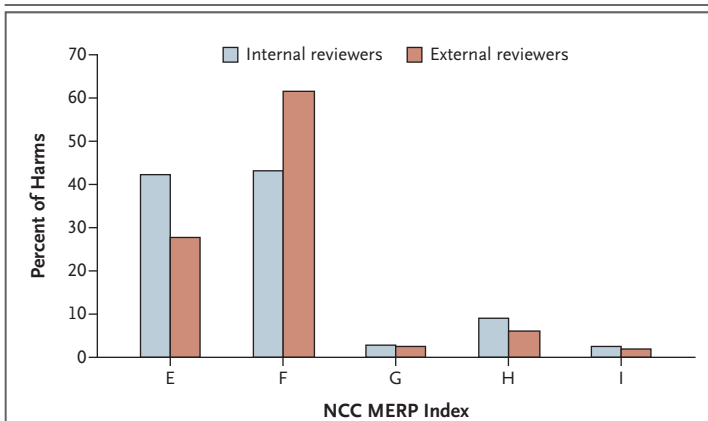


Figure 1. Severity of Harms Detected by Internal and External Reviewers in 10 North Carolina Hospitals (2002–2007).

Harms to patients were rated according to categories of severity used by the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index as follows: E, temporary harm to the patient requiring intervention; F, temporary harm to the patient requiring initial or prolonged hospitalization; G, permanent harm to the patient; H, intervention required to sustain life; and I, death of the patient.

in 2007 (reduction factor, 0.91; 95% CI, 0.84 to 0.994; $P=0.04$) (Fig. 2D). On a per-admission basis, the unadjusted rate of preventable harms also decreased during the study period, from 10.2 harms per 100 admissions in 2002 to 6.5 harms per 100 admissions in 2007 (annual reduction factor, 0.91; 95% CI, 0.84 to 0.99; $P=0.03$) (Fig. 3D). There were no significant changes in rates of higher-severity harms (categories F through I) over time (Fig. 2F and 3F).

RISK ADJUSTMENT

Multivariate analysis of internal reviews with adjustment for demographic features, hospital service, and high-risk conditions had little effect on the primary study results, with a nonsignificant reduction in harms per 1000 patient-days (annual reduction factor, 0.99; 95% CI, 0.94 to 1.04; $P=0.61$). In multivariate analysis of external reviews, there was also a nonsignificant reduction in harms (annual reduction factor, 0.98; 95% CI, 0.93 to 1.04; $P=0.47$). For the rate of preventable harms per 1000 patient-days, external reviews showed a reduction that did not reach statistical significance (reduction factor, 0.92; 95% CI, 0.85 to 1.00; $P=0.06$); internal reviews showed no reduction (reduction factor, 1.00; 95% CI, 0.94 to 1.06; $P=0.92$).

DISCUSSION

In a statewide study of 10 North Carolina hospitals, we found that harm resulting from medical care was common, with little evidence that the rate of harm had decreased substantially over a 6-year period ending in December 2007. Although there was a modest reduction in the rate of preventable harms on the basis of external reviews, the reduction did not reach statistical significance in adjusted analyses. This apparent reduction was not substantiated by the internal reviews, which by all measures were of higher quality than the external reviews (i.e., higher within-team reliability at both primary and secondary review stages and higher agreement with experienced reviewers).²⁵

Our findings validate concern raised by patient-safety experts in the United States¹⁷ and Europe¹⁸ that harm resulting from medical care remains very common. Though disappointing, the absence of apparent improvement is not entirely surprising. Despite substantial resource allocation and efforts to draw attention to the patient-safety epidemic on the part of government agencies, health care regulators, and private organizations,²⁻⁴ the penetration of evidence-based safety practices has been quite modest. For example, only 1.5% of hospitals in the United States have implemented a comprehensive system of electronic medical records, and only 9.1% have even basic electronic record keeping in place; only 17% have computerized provider order entry.¹³ Physicians-in-training and nurses alike routinely work hours in excess of those proven to be safe.^{7-9,27,28} Compliance with even simple interventions such as hand washing is poor in many centers.¹⁴

A reliable measurement strategy is required to determine whether efforts to enhance safety are resulting in overall improvements in care, either locally or more broadly.¹⁸ Most medical centers continue to depend on voluntary reporting to track institutional safety, despite repeated studies showing the inadequacy of such reporting.^{29,30} The patient-safety indicators of the AHRQ are susceptible to variations in coding practices, and many of the measures have limited sensitivity and specificity.^{24,31} Recent studies have shown that the trigger tool has very high specificity, high reliability, and higher sensitivity

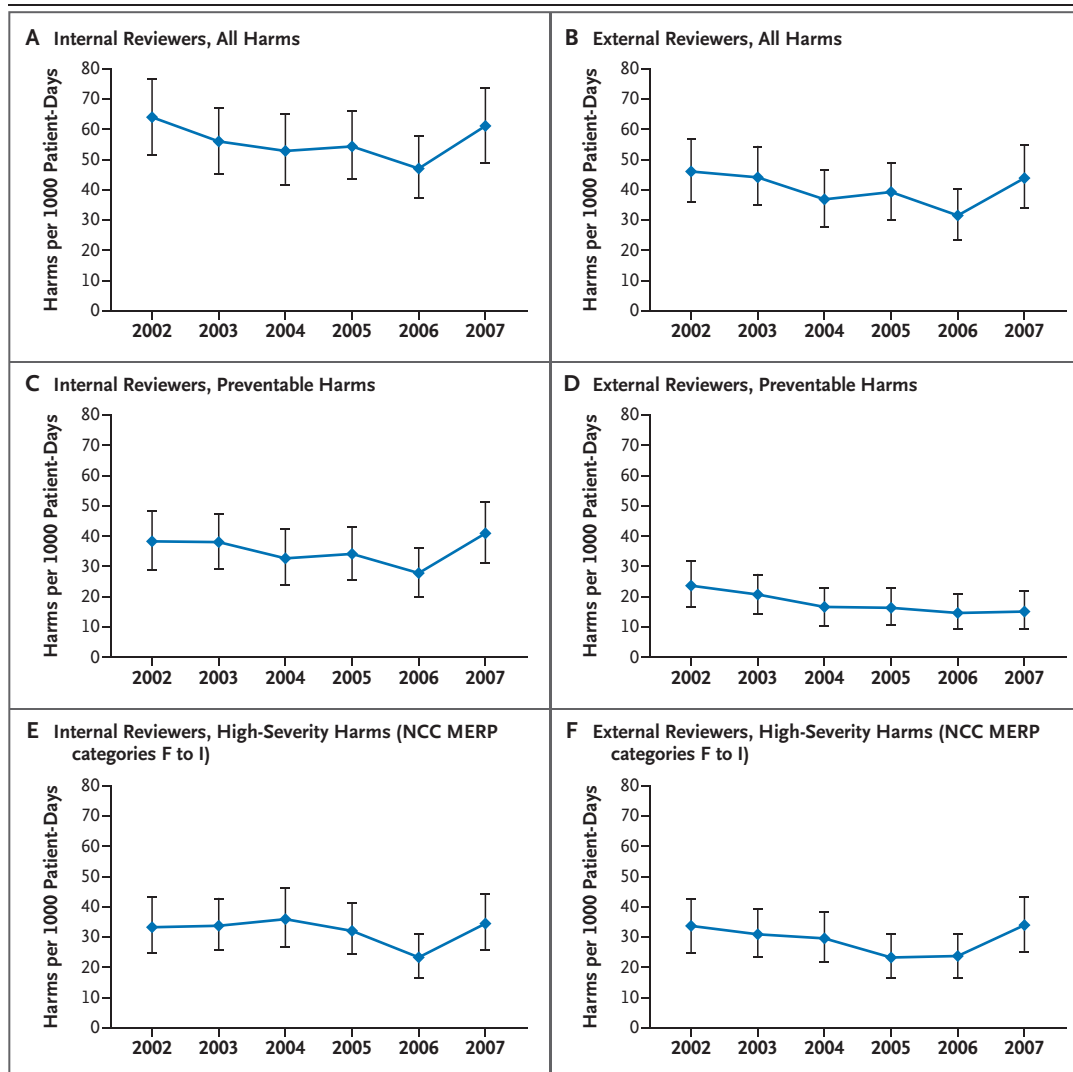


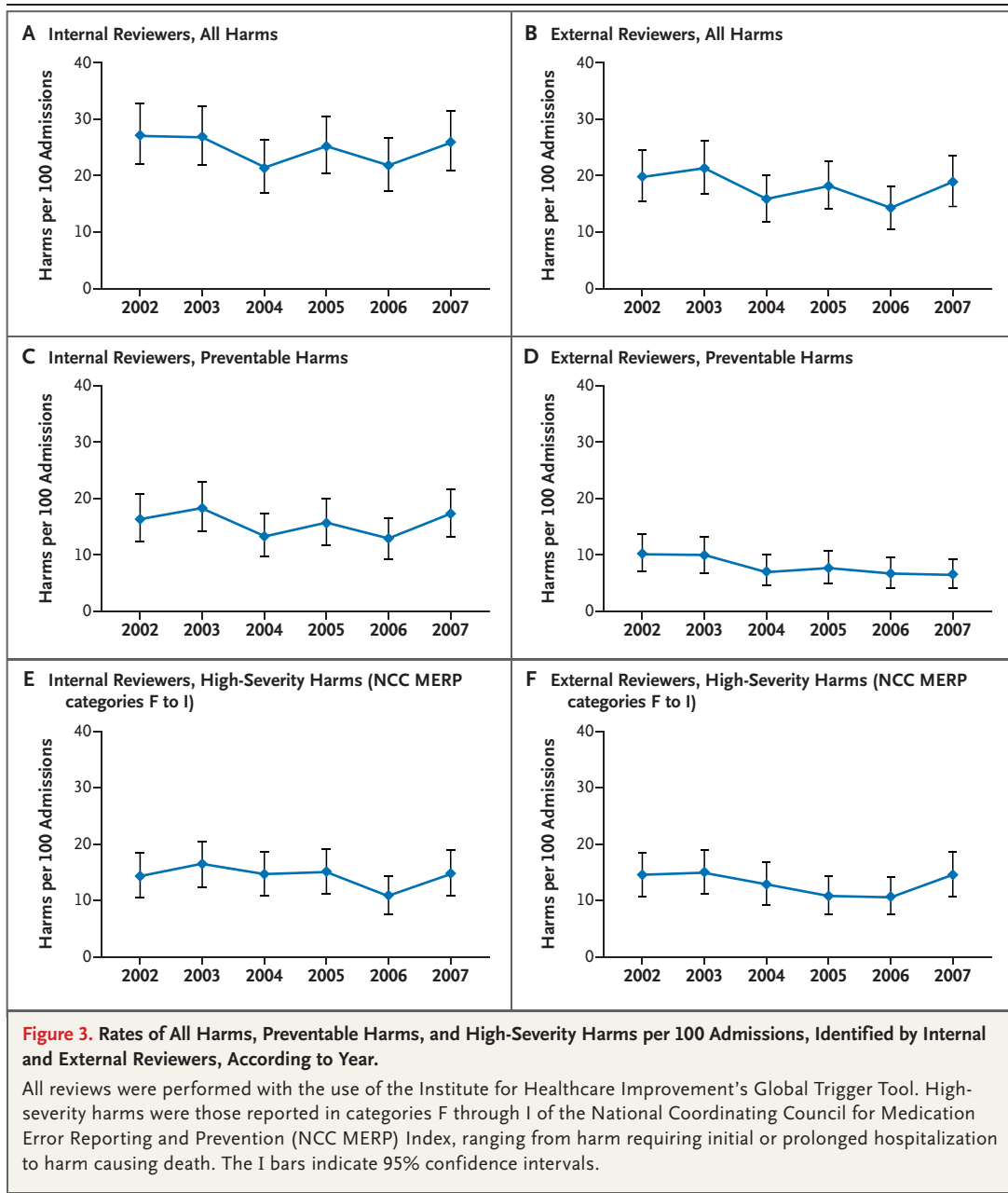
Figure 2. Rates of All Harms, Preventable Harms, and High-Severity Harms per 1000 Patient-Days, Identified by Internal and External Reviewers, According to Year.

All reviews were performed with the use of the Institute for Healthcare Improvement's Global Trigger Tool. High-severity harms were those reported in categories F through I of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index, ranging from harm requiring initial or prolonged hospitalization to harm causing death. The I bars indicate 95% confidence intervals.

than other methods.^{24,25} Manual use of the trigger tool is labor-intensive, but as electronic medical records become more widespread, automating trigger detection could substantially decrease the time required to use this surveillance tool.

Our study has several limitations. First, North Carolina may not be representative of the United States as a whole. We chose North Carolina because of its high level of engagement in efforts to improve patient safety. In addition, the state

has a reputation for being especially proactive regarding patient safety through the North Carolina Hospital Association and the North Carolina Center for Hospital Quality and Patient Safety¹⁹ and was rated as one of the most "engaged" states in the Institute for Healthcare Improvement's harm-reduction campaigns.²⁰ Second, we studied only 10 randomly selected hospitals. Although we sought through our stratification and randomization procedure to ensure



that the selected hospitals were representative, it is possible that these 10 hospitals differ from other North Carolina hospitals in some unrecognized manner. Third, any record review is limited to the information provided in the record. However, the trigger tool has been found to detect harm at higher rates than previous methods of record review,³²⁻³⁴ hospital incident reporting,²⁴ and administrative database algorithms, such as patient-safety indicators of the AHRQ. Although the rates of reliability (both interrater and intra-

rater) and the specificity of internal reviews were high in our study, the newly trained reviewers who participated in the study detected fewer harms than did highly experienced reviewers. Additional monitoring and training may be needed in future studies to bring all reviewers to an expert level of proficiency.³⁵ Finally, our study was powered to detect a 25% reduction in the incidence of harms over a 6-year period, and change in the incidence of all harms, rather than preventable harms, was the primary outcome of

the study, since definitions of preventability are prone to change over time.

Although the lack of a significant reduction in harm suggests that the Institute of Medicine's ambitious goal of a 50% reduction during a 5-year period has not been met,¹ we cannot rule out the possibility of smaller improvements, particularly since the baseline rate of harms that was detected in this study was somewhat lower than anticipated. We also cannot rule out a reduction in harms that was not captured by the trigger tool. The finding in this study of reductions in preventable harms (though not total harms) of borderline statistical significance on the basis of external reviews suggests the possibility that some improvements are beginning to occur, though further longitudinal studies using robust methods will be needed to determine whether this is, in fact, the case. There was some apparent variation among hospitals in rates of change over time, but the study was not powered to examine such variation reliably or to explore the effect of specific hospital-based improvements on rates of harm in particular hospitals. Rather, our goal was to evaluate the aggregate effects of efforts to improve safety across hospitals.

In conclusion, harm to patients resulting from medical care was common in North Carolina, and the rate of harm did not appear to decrease significantly during a 6-year period ending in December 2007, despite substantial national attention and allocation of resources to improve the safety of care. Since North Carolina

has been a leader in efforts to improve safety, a lack of improvement in this state suggests that further improvement is also needed at the national level. Although the absence of large-scale improvement is a cause for concern, it is not evidence that current efforts to improve safety are futile. On the contrary, data have shown that focused efforts to reduce discrete harms, such as nosocomial infections^{10,36} and surgical complications,³⁷ can significantly improve safety. However, achieving transformational improvements in the safety of health care will require further study of which patient-safety efforts are truly effective across settings and a refocusing of resources, regulation, and improvement initiatives to successfully implement proven interventions.

Supported by a grant from the Rx Foundation.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the members of the Scientific Advisory Group, including Jerry Gurwitz, M.D., Donna Isgett, R.N., M.S.N., Brent James, M.D., M.Stat., Bruce Landon, M.D., Lucian Leape, M.D., Elizabeth McGlynn, Ph.D., David Pryor, M.D., Richard Thomson, and James Ware, Ph.D.; David Classen, M.D., for providing guidance on the development of the study protocol; Lee Adler, D.O., Nancy Kimmel, R.Ph., S.S.B.B., Marjorie E. McKeever, R.N., B.S., Diedre A. Rahn, R.N., Frances A. Griffin, R.R.T., M.P.A., and Roger K. Resar, M.D., who conducted the experienced reviews that served as a reference for both internal and external reviews; Catherine M. Murphy, Dale A. Rhoda, M.P.P., Warren J. Strauss, Charles E. Knott, and their colleagues at Battelle Centers for Public Health Research and Evaluation for their help in the conduct of the study and preliminary analyses; the North Carolina Hospital Association for its help in recruiting hospitals; and Frank Davidoff, M.D., and Jane Roessner, Ph.D., for their critical review and assistance in the preparation of the manuscript.

REFERENCES

- Kohn LT, Corrigan JM, Donaldson MS, eds. *To err is human: building a safer health system*. Washington, DC: National Academies Press, 1999.
- Agency for Healthcare Research and Quality. Medical errors & patient safety. Rockville, MD: AHRQ. (<http://www.ahrq.gov/qual/errorsix.htm>.)
- McCannon CJ, Hackbarth AD, Griffin FA. Miles to go: an introduction to the 5 Million Lives Campaign. *Jt Comm J Qual Patient Saf* 2007;33:477-84.
- A journey through the history of the Joint Commission. Oakbrook Terrace, IL: The Joint Commission. (http://www.jointcommission.org/AboutUs/joint_commission_history.htm.)
- Bates DW, Teich J, Lee J, et al. The impact of computerized physician order entry on medication error prevention. *J Am Med Assoc* 1999;281:313-21.
- Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998;280:1311-6.
- Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. *N Engl J Med* 2004;351:1829-37.
- Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med* 2004;351:1838-48.
- Ulmer C, Wolman DM, Johns MME, eds. *Resident duty hours: enhancing sleep, supervision, and safety*. Washington, DC: National Academies Press, 2009.
- Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;355:2725-32. [Erratum, *N Engl J Med* 2007;356:2660.]
- Sharek PJ, McCleod RE Jr, Taketomo C, et al. An intervention to decrease narcotic-related adverse drug events in children's hospitals. *Pediatrics* 2008;122(4):e861-e866.
- Shojania KG, Duncan BW, McDonald KM, Wachter RM. Safe but sound: patient safety meets evidence-based medicine. *JAMA* 2002;288:508-13.
- Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med* 2009;360:1628-38.
- Burke JP. Infection control — a problem for patient safety. *N Engl J Med* 2003;348:651-6.
- Landrigan CP, Barger LK, Cade BE, Ayas NT, Czeisler CA. Interns' compliance with Accreditation Council for Graduate Medical Education work-hour limits. *JAMA* 2006;296:1063-70.
- Longo DR, Hewett JE, Ge B, Schubert

- S. The long road to patient safety: a status report on patient safety systems. *JAMA* 2005;294:2858-65. [Erratum, *JAMA* 2006; 295:164.]
17. Leape LL, Berwick DM. Five years after To Err Is Human: what have we learned? *JAMA* 2005;293:2384-90.
18. Vincent C, Aylin P, Franklin BD, et al. Is health care getting safer? *BMJ* 2008; 337:a2426.
19. North Carolina Center for Hospital Quality and Patient Safety. About us. (<http://www.ncqualitycenter.org/about.lasso>.)
20. Institute for Healthcare Improvement. A network that works! The 100,000 Lives Campaign nodes. Cambridge, MA: IHI, 2006. (<http://www.ihl.org/IHI/Topics/Improvement/SpreadingChanges/ImprovementStories/ANetworkThatWorks100000LivesCampaignNodes.htm>.)
21. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003;12:Suppl 2:ii39-ii45.
22. Sharek PJ, Horbar JD, Mason W, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics* 2006;118:1332-40.
23. Griffin FA, Resar RK. Global Trigger Tool for measuring adverse events: IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement, 2007.
24. Office of the Inspector General. Adverse events in hospitals: methods for identifying events. Washington, DC: Department of Health and Human Services, 2010. (OEI-06-08-00221.) (<http://www.oig.hhs.gov/oei/reports/oei-06-08-00221.pdf>.)
25. Sharek PJ, Parry G, Goldmann DA, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res* 2010 August 16 (Epub ahead of print).
26. National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP). NCC MERP index for categorizing medication errors. (<http://www.nccmerp.org/pdf/indexBW2001-06-12.pdf>.)
27. Rogers AE, Hwang WT, Scott LD, Aiken LH, Dinges DF. The working hours of hospital staff nurses and patient safety. *Health Aff (Millwood)* 2004;23(4):202-12.
28. Page A, ed. Keeping patients safe: transforming the work environment of nurses. Washington, DC: National Academies Press, 2004.
29. Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. *Jt Comm J Qual Improv* 1995; 21:541-8.
30. Sari AB, Sheldon TA, Cracknell A, Turnbull A. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *BMJ* 2007;334:79.
31. Landrigan CP. The safety of inpatient pediatrics: preventing medical errors and injuries among hospitalized children. *Pediatr Clin North Am* 2005;52:979-93.
32. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results from the Harvard Medical Practice Study I. *N Engl J Med* 1991;324:370-6.
33. Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II. *N Engl J Med* 1991;324:377-84.
34. Thomas EJ, Studdert DM, Runciman WB, et al. A comparison of iatrogenic injury studies in Australia and the USA. I. Context, methods, casemix, population, patient and hospital characteristics. *Int J Qual Health Care* 2000;12:371-8.
35. Classen DC, Lloyd RC, Provost L, Griffin FA, Resar R. Development and evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf* 2008;4:169-77.
36. Reduction in central line-associated bloodstream infections among patients in intensive care units — Pennsylvania, April 2001–March 2005. *MMWR Morb Mortal Wkly Rep* 2005;54:1013-6.
37. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009;360:491-9.

Copyright © 2010 Massachusetts Medical Society.

PERSONAL ARCHIVES IN THE JOURNAL ONLINE

Individual subscribers can store articles and searches using a feature on the *Journal's* Web site ([NEJM.org](http://www.nejm.org)) called "Personal Archive." Each article and search result links to this feature. Users can create personal folders and move articles into them for convenient retrieval later.