Association Between Implementation of a Medical Team Training Program and Surgical Mortality

Julia Neily, RN, MS, MPH
Peter D. Mills, PhD, MS
Yinong Young-Xu, ScD, MA, MS
Brian T. Carney, MD
Priscilla West, MPH
David H. Berger, MD, MHCM
Lisa M. Mazzia, MD
Douglas E. Paull, MD
James P. Bagian, MD, PE

ADVERSE EVENTS RELATED TO surgery continue to occur despite the best efforts of clinicians.[1] Teamwork and effective communication are known determinates of surgical safety.[2-6] Previous efforts at demonstrating the efficacy of patient safety initiatives have been limited because of the inability to study a control group.[7] For example, the use of the World Health Organization Safe Surgery checklist has been evaluated, but its overall efficacy remains uncertain because no control group was studied to clearly demonstrate this instrument's effectiveness.[6]

The Veterans Health Administration (VHA) is the largest national integrated health care system in the United States, with 153 hospitals, 130 of which provide surgical services. The VHA implemented a national team training program and studied the program's effect on patient outcomes. The VHA began piloting team training that

Context There is insufficient information about the effectiveness of medical team training on surgical outcomes. The Veterans Health Administration (VHA) implemented a formalized medical team training program for operating room personnel on a national level.

Objective To determine whether an association existed between the VHA Medical Team Training program and surgical outcomes.

Design, Setting, and Participants A retrospective health services study with a contemporaneous control group was conducted. Outcome data were obtained from the VHA Surgical Quality Improvement Program (VASQIP) and from structured interviews in fiscal years 2006 to 2008. The analysis included 182 409 sampled procedures from 108 VHA facilities that provided care to veterans. The VHA's nationwide training program required briefings and debriefings in the operating room and included checklists as an integral part of this process. The training included 2 months of preparation, a 1-day conference, and 1 year of quarterly coaching interviews

Main Outcome Measure The rate of change in the mortality rate 1 year after facilities enrolled in the training program compared with the year before and with nontraining sites.

Results The 74 facilities in the training program experienced an 18% reduction in annual mortality (rate ratio [RR], 0.82; 95% confidence interval [CI], 0.76-0.91; $P=.01$) compared with a 7% decrease among the 34 facilities that had not yet undergone training (RR, 0.93; 95% CI, 0.80-1.06; $P=.59$). The risk-adjusted mortality rates at baseline were 17 per 1000 procedures per year for the trained facilities and 15 per 1000 procedures per year for the nontrained facilities. At the end of the study, the rates were 14 per 1000 procedures per year for both groups. Propensity matching of the trained and nontrained groups demonstrated that the decline in the risk-adjusted surgical mortality rate was about 50% greater in the training group (RR, 1.49; 95% CI, 1.10-2.07; $P=.01$) than in the nontraining group. A dose-response relationship for additional quarters of the training program was also demonstrated: for every quarter of the training program, a reduction of 0.5 deaths per 1000 procedures occurred (95% CI, 0.2-1.0; $P=.001$).

Conclusion Participation in the VHA Medical Team Training program was associated with lower surgical mortality.

JAMA. 2010;304(15):1693-1700   www.jama.com

Author Affiliations: National Center for Patient Safety, Department of Veterans Affairs (Mss Neily and West and Drs Mills, Young-Xu, Carney, Mazzia, Paull, and Bagian) and Department of Psychiatry, Dartmouth Medical School (Drs Mills and Young-Xu), Hanover, New Hampshire; Department of Surgery, Baylor College of Medicine, and Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas (Dr Berger); Preventive Medicine and Community Health at the University of Texas Medical Branch, Galveston, and Military and Emergency Medicine—Uniformed Services University of the Health Sciences, F. Edward Hebert School of Medicine, Bethesda, Maryland (Dr Bagian).
Corresponding Author: Julia Neily, RN, MS, MPH, 215 N Main St, White River Junction, VT 05009 (Julia .Neily@va.gov).

incorporated checklists that were used to drive preoperative briefings and postoperative debriefings in 2003. In 2006, based on the pilot experience, the VHA implemented a nationwide Medical Team Training program. The goal of the present study, which includes more than 100 facilities and more than 100 000 sampled procedures over 3 years, was to analyze surgical mortality for facilities that received the VHA training program compared with those that had not yet received it. We hypothesized that facilities where the program was implemented would have improved surgical mortality compared with their own baseline and with facilities that had not yet received the training. We also hypothesized that a higher degree of implementation would be associated with lower surgical mortality.

METHODS

Intervention

The Medical Team Training program[8,9] includes 2 months of preparation and planning with each facility's implementation surgical care team. This is followed by a day-long onsite learning session. To allow surgical staff to attend as a team (surgeons, anesthesiologists, nurse anesthetists, nurses, and technicians), the operating room (OR) is closed. Using the crew resource management theory from aviation adapted for health care,[10] clinicians were trained to work as a team; challenge each other when they identify safety risks; conduct checklist-guided preoperative briefings and postoperative debriefings; and implement other communication strategies such as recognizing red flags, rules of conduct for communication, stepping back to reassess a situation, and how to conduct effective communication between clinicians during care transitions. The learning session included lecture, group interaction, and videos. After the learning session, 4 quarterly follow-up structured telephone interviews were conducted with the team for 1 year to support, coach, and assess the Medical Team Training implementation. Fol-

low-up calls were usually conducted with the OR nurse manager or an OR nurse, a surgeon or chief of surgery, and other staff nurses, and administrative support staff also frequently participated.

The training and follow-up support formally started in August 2006 with a small number of volunteers; then as of January 2007, the program was mandated for facilities that performed surgical procedures. Specifically, the VHA deputy undersecretary for operations and management issued a memorandum on October 2, 2006, stating that all facilities that performed surgery and that had 1 or more intensive care units would receive the program (eAppendix 1 available at http://www.jama.com). To plan the national rollout, time blocks were created for the Veteran Service Integrated Networks and then the network directors were asked to rank their time-block preferences based on their individual facility's readiness for training (eTable 1 available at http://www.jama.com). Directors were required to submit their preferences by November 1, 2006. The National Center for Patient Saftey made every effort to assign first or second time-block preferences. Logistical considerations determined the order in which facilities and Veteran Service Integrated Networks were trained.

The VHA Surgical Quality Improvement Program (VASQIP) mortality outcome data were not a factor in determining the order in which facilities were selected for training because the data were not available to the training team at that time. Selection order was not random because the primary focus for the program was to establish a national program that worked for each facility's scheduling needs and readiness and were thus implemented over a 2-year period. Although the final selections for Veteran Service Integrated Networks were communicated to directors to be implemented over a 2-year period, there were some alterations to the initial implementation timing plan. One hundred thirty facilities were slated for the program. Ten facilities that had

participated in both the pilot and the formal program were excluded from the analysis to ensure that the facilities that received training and the ones that did not were similar at baseline. Twelve facilities scheduled for the training list but with no VASQIP data were excluded from the analysis.

Facilities that had received the training were required to implement briefings and debriefings with the intent to improve communication and surgical safety. They were provided with sample checklists and referred to an internal VHA Web site that contained several briefing and debriefing tools or checklists being used at VHA facilities. Facilities adapted these for their needs and most developed specialty-specific checklists.[11]

Study Design

This was a retrospective health services cohort study using a contemporaneous control group. Surgical mortality data included fiscal years 2006 to 2008. Follow-up data from facility team interviews included fiscal years 2007 and 2008.

Main Outcome Measures

The VASQIP, formerly known as National Surgical Quality Improvement Program, provides reliable, valid, risk-adjusted (surgical complexity, patient comorbidity, and sociodemographic characteristics), and observed 30-day mortality rates for major noncardiac surgery performed at VA medical centers.[12-15] The measures say that "[m]ajor operations performed under general, spinal, or epidural anesthesia are candidates for entry into the database. At low-volume centers, all eligible operations are included. To eliminate sampling bias at higher volume centers, the first 36 consecutive eligible operations are entered in each 8-day cycle, beginning with a different day of the week each cycle."[15(p837)] Surgical specialties included general, orthopedics, urology, vascular, neurosurgery, otolaryngology, non–cardiac thoracic, plastic, and other noncardiac subspecialties.[14] "*CPT* [*Current Procedural*

Terminology] codes of procedures with known low morbidity and mortality rates or transurethral resections of the prostate (TURPs), transurethral resections of the bladder tumor (TURBTs) and herniorraphies exceeding the limit of five per week" were excluded.[14(p495)] Mortality was defined as patient death in or out of the hospital from any cause within 30 days after the operation and risk adjustment was performed based on each patient's characteristics. The VASQIP data are considered the gold standard for measuring surgical quality.[16] Because Medical Team Training interventions were implemented on a facility level, we selected outcome measures to be aggregated mortality rates at the facility level. In other words, the unit of analysis was the facility. Using VASQIP data, we included surgical mortality rates (observed and risk adjusted) as our primary end points for each of the 108 facilities for fiscal years 2006 to 2008. Unless otherwise stated, all years are meant to be VHA fiscal years. For example, year 2008 means fiscal year 2008 (from October 1, 2007, to September 30, 2008).

Although used only as a covariate in our analyses, VASQIP also provides a performance measure called the O to E ratio, where *O* represents the total number of observed events (deaths or complications) and *E*, the number of events expected on the basis of the compendium of the preoperative risk factors prevalent in the patient population. Daley et al[17] found that high O to E ratio outlier hospitals are more likely to have inferior structures and processes of care and that low outlier hospitals are more likely to have superior structures and processes of care.

To examine baseline characteristics of the sites between those that participated in the training and those that did not, we compared the following: rural or urban status, complexity, VASQIP surgical volume, baseline observed and risk-adjusted mortality rate, and O to E mortality ratio. The VHA 2005 complexity model designates VHA facilities into 3 categories: level 1 represents high complexity;

level 2, medium complexity; and level 3, low complexity. These designations are based on a composite involving the number of patients seen, patient risk, number of physician specialists, teaching status, research dollars, and intensive care unit capability.[18] The VHA urban, rural, and highly rural classification is based partly on census tract and partly on population density. Facilities located in US census tracts designated as "urban" are considered urban. All others are considered rural, except for facilities located in a county with a population density of less than 7 people per square mile, which are considered highly rural.[19] For our analysis, hospitals designated by VHA as either rural or highly rural were considered rural.

Intervention Measures

The measurement variables for the intervention (described below) were enrollment in the Medical Team Training program, number of quarters participating in the program (training and follow-up), and degree of briefing and debriefing.

Follow-up quarterly interviews were scheduled at intervals of 1, 4, 8, and 12 months after the learning session. During these semistructured interviews, we asked participating facilities about various aspects of the implementation of the program. Although the program was designed to improve patient safety, we also assessed whether the program affected OR efficiency. The structured interview tool used for this assessment is included as eAppendix 2 (available at http://www.jama.com).[11]

Interview Coding

Narrative responses that required interpretation were coded. The research team identified themes in the responses and then developed a codebook. Interrater reliability was achieved at a κ of 0.76 .

Training and Follow-up

The VASQIP mortality rate data were only available by year. We were only provided the total number of proce-

dures and surgical deaths per facility per year. However, the training program included quarterly intervals of assessment and follow-up. Therefore, we created a yearly measure of training and follow-up. This represented the number of quarters in the fiscal year during which a facility had received the training program and was receiving follow-up.

Degree of Briefing and Debriefing Guided by a Checklist

We not only wanted to train and follow up with facilities, we also wanted to determine, to the greatest extent possible, the degree to which implementation of briefing and debriefing occurred. During the quarterly follow-up interviews, staff members in participating facilities reported on which surgical specialties were conducting briefings and debriefings, how many procedures had been performed during a specified time, and how many of these procedures had a briefing and a debriefing.

The briefing and debriefing process was more comprehensive and above and beyond the required time-out process. The briefings offered the team the opportunity to set the stage for how they would communicate during the case (procedure); the training taught them to encourage all team members to speak up if they had a safety concern. Briefings also were intended to methodically review key aspects about the patient and what was needed for the procedure. The debriefing was intended to give team members a chance to voice what worked well and what needed to be improved for future cases. These were categorized into 4 ordinal categories of briefing and debriefing: (0) none, (1) some services or some cases, (2) some cases in all services or all cases in some services, but not both; and (3) all cases in all services. The quarterly briefing and debriefing scores were averaged into a yearly measure, to match the VASQIP mortality data that were only available by year.

Information on use of a briefing and debriefing checklist was obtained from

Table 1. Baseline Facility Characteristics (N = 108)

	Received Medical Team Training		*P* Value
	No (n = 34)	Yes (n = 74)	
Observed mortality rate per 1000 procedures per year[a]			
Mean (SD)	15 (8)	17 (8)	.35
Median (range)	16 (0-34)	16 (0-42)	
Risk-adjusted mortality rate per 1000 procedures per year[a]			
Mean (SD)	15 (6)	17 (6)	.30
Median (range)	14 (2-35)	15 (1-39)	
O to E ratio, mean (SD) [range][b]	0.92 (0.53) [0-2.13]	0.97 (0.38) [0-1.95]	.62
VASQIP surgical volume, cases per facility[b]	881 (424)	850 (427)	.54
Median (range), No.	870 (138-1686)	831 (64-2067)	
Facility No. (%)[c]			
0 Mortality	3 (9)	3 (4)	.32
High-complexity	19 (56)	38 (51)	.66
Urban	29 (85)	62 (84)	.87

Abbreviation: VASQIP, Veterans Health Administration Surgical Quality Improvement Program.
[a] For variables that are not normally distributed, we used a nonparametric method (Mann-Whitney) for the testing equality between training and nontraining groups.
[b] See "Methods" section for a definition of the O to E ratio.
[c] For comparison of proportions (facilities with 0 mortality, high-complexity facility, and urban facility), we used χ^2 test.

follow-up interviews. Checklist tools included a variety of approaches such as laminated checklist cards, whiteboards (some had sliders to indicate a completed item), paper forms, and wall-mounted posters. Some participants referred to these as guides or tools. For the purposes of this article, we refer to them as checklists.

Statistical Analysis

In general, we modeled the count data of the number of surgical deaths using the Poisson distribution. The link function was the logarithm, and surgical volume was the offset. Independent variables included number of quarters of the Medical Team Training and the degree to which briefings and debriefings were conducted (all aggregated by year to match yearly VASQIP mortality rates), all with surgical risk as a covariate in the model. All analyses were performed using SAS statistical software version 9.2 (PROC GENMOD, SAS Institute Inc, Cary, North Carolina). All reported *P* values are 2-sided, at a significance level of .05.

The yearly mortality rate was defined as number of surgical mortalities (as defined by VASQIP above) divided by the number of procedures. The primary outcome measure was change in mortality rate during the year that facilities were enrolled in the program compared with the year before. Continuous variables that were not distributed normally were also expressed as medians and ranges and were compared using the Mann-Whitney test. Pearson χ^2 tests were used to compare proportions (TABLE 1). Multivariable Poisson generalized estimating equations (GEEs)[20,21] were used to assess associations of training with outcome while adjusting for secular trends as well as propensity scores. After the creation of the propensity score, we performed a full Poisson GEE model while matching on propensity scores by stratification to evaluate the association of the training and mortality rate.[22,23]

Rate ratios (RRs) and accompanying 95% confidence intervals (CIs) were calculated to represent the strength of association between training exposure and mortality rates, estimated using either Poisson regression or Poisson GEE model. The GEE method was used to account for the repeated longitudinal nature of the yearly collected data on the outcomes (mortality rates). We controlled for baseline characteristics that included complexity, size, urbanicity, baseline O to E ratio, and mortality and morbidity rates.

We compared not only facilities that received the training with those that did not, but we also compared before and after mortality rates within each facility and with the other participating facilities. As a result, some facilities served as their own controls as well as controls for others.

Propensity Score

Because training was not randomly assigned in this study, potential confounding and selection biases were accounted for by developing a propensity score for training. The propensity score is the probability of receiving the program for a facility with specific baseline characteristics. The propensity score was constructed based on the following variables measured at baseline: the observed and risk-adjusted surgical mortality rates, the observed and risk-adjusted surgical morbidity rates, the average number of sampled procedures per facility, hospital complexity, and urbanicity. These are summarized into 1 propensity score (through a full nonparsimonious logistic model; TABLE 2). Based on a propensity score calculation, participating facilities were stratified into 4 categories. Within-propensity score strata, covariates in training and nontraining groups are similarly distributed; furthermore, stratifying by propensity score removes more than 90% of the overt bias due to the covariates used to estimate the score.[24] Once groups are stratified by propensity score, they can again be separated into training vs nontraining groups to detect differences in baseline variables such as observed and adjusted surgical mortality to suggest an imbalance. Propensity scores cannot remove hidden biases except to the extent that unmeasured variables are correlated with the measured covariates used to compute the score.[22-24]

Propensity Score Matching

Propensity score matching was used to select nontraining facilities that are similar to facilities receiving the training program with respect to propensity score and other covariates, thereby

matching on many confounders simultaneously.[25] As a result, by construction, in each propensity score stratum defined by this procedure, the covariates were balanced and the training assignment could be considered random. Then within each stratum, a Poisson analysis was performed to compute the training program's effect. The total program effect was finally obtained as a summary of the effect estimate of each stratum. This provides a more valid estimate of program effect because they compare facilities with similar baseline characteristics.

The study was approved by the Research and Development Committee at the VA Medical Center in White River Junction, Vermont, and considered exempt by the Dartmouth College institutional review board.

RESULTS

A total of 108 facilities were analyzed. If a facility had not yet received Medical Team Training in a particular fiscal year, it was counted as having 0 quarters of training for that year. Seventy-four facilities had 0 quarters at baseline (**FIGURE**).

We analyzed 3 years of available VASQIP data: 2006, 2007, and 2008. The baseline VASQIP mortality rate measure for the 42 facilities that underwent program implementation in 2007 was their 2006 rate, and the baseline rate for the 32 that underwent it in 2008 was their 2007 rate (5 of the 42 facilities had their learning session in the fourth quarter of fiscal year 2006). Thirty-four facilities did not receive training during those 3 years. Because the majority of the 74 facilities had initiated the program in 2007, we analyzed 2006 and 2007 for the 34 untrained facilities to control for existing secular trends. The baseline for the 34 nontrained facilities that served as the contemporaneous control group was 2006 and their follow-up year was 2007.

The characteristics of facilities at baseline did not differ significantly statistically (Table 1). The trained facilities had on average higher observed and

risk-adjusted mortality rates at baseline than the nontrained facilities.

Table 2 displays the 4 groups of facilities stratified by propensity scores for each baseline mortality variable (observed and adjusted surgical mortality rates). Mean Medical Team Training selection propensity scores ranged from 0.33 to 0.88 across propensity quartiles, with discrimination between both training groups (*C* Statistic=0.74). The distribution of key potential confounders—observed and risk-adjusted surgical mortality rate at baseline—was similar within propensity quartiles for trained and nontrained facilities. This indicates that at baseline, none of the 4 propensity score strata showed dif-

Table 2. Mortality Rates at Baseline by Propensity Score Quartiles, According to Medical Team Training Selection

	Quartiles of Propensity Score, Mean (Range)[a]			
	0.33 (0.24-0.38)	0.52 (0.40-0.60)	0.68 (0.60-0.80)	0.88 (0.80-0.97)
Observed-mortality rate per 1000 procedures per year				
Nontrained	10	13	14	17
Trained	10	13	14	18
P value	.51	.51	.78	.65
Risk-adjusted mortality rate per 1000 procedures per year				
Nontrained	9	13	14	18
Trained	8	13	15	17
P value	.33	.68	.80	.63
Observed-morbidity rate per 1000 procedures per year				
Nontrained	60	78	80	106
Trained	80	71	88	110
P value	.66	.69	.63	.81
Risk-adjusted morbidity rate per 1000 procedures per year				
Nontrained	69	83	97	101
Trained	80	84	94	108
P value	.63	.91	.65	.29
Mean No. of procedures sampled per facility				
Nontrained	443	729	937	1012
Trained	915	781	789	980
P value	.26	.74	.32	.88
Urban facilities, %				
Nontrained	100	88	78	83
Trained	67	100	79	86
P value	.27	.25	.95	.88
High-complexity facilities, %				
Nontrained	33	63	78	67
Trained	67	60	55	54
P value	.41	.90	.22	.58

[a] Propensity analysis was based on the observed and risk-adjusted surgical mortality rates, the observed and risk-adjusted surgical morbidity rates, the average number of sampled procedures per facility, hospital complexity, and urbanicity.

Figure. Quarters of Risk-Adjusted Surgical Mortality Rate

No. of Facilities: 74, 16, 20, 24, 14

Linear trend line was fitted to the average risk-adjusted mortality rate (n=5). Error bars indicate 95% confidence intervals.
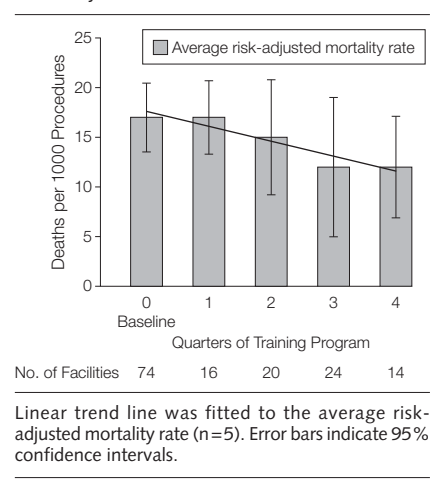
Table 3. Improvements Reported by Medical Team Training Facilities From Structured Interviews

Reported Improvements	No. (%) of Facilities (n = 74)
Communication among operating room staff	35 (47.2)
Staff awareness	34 (46.0)
Overall efficiency	49 (66.2)
Equipment use during surgery	44 (59.9)
Reduced length of procedures	15 (20.3)
Improved first-case start times	30 (40.5)
Other types of efficiency improvements[a]	6 (8.1)

[a] For example, reduced delays for surgical consent, decreased turnover time between cases, reduction in staff overtime hours.

ferences in mortality rates between the training groups. In other words we were not able to discern any overt selection bias based on these 2 key baseline variables in our model after matching on propensity score.

The risk-adjusted mortality rates at baseline were 17 per 1000 procedures per year for the trained facilities and 15 per 1000 procedures per year for the nontrained facilities. At the end of the study, the rates were 14 per 1000 procedures per year for both groups.

After controlling for baseline differences, the 74 trained facilities experienced a significant decrease of 18% in observed mortality (RR, 0.82; 95% CI, 0.76-0.91; $P = .01$). Mortality decreased by 7% (RR, 0.93; 95% CI, 0.80-1.06; $P = .59$) in the nontrained facilities.

Raw and risk-adjusted annual mortality rates were unchanged in the 34 nontrained facilities. Propensity-matched mortality assessment showed an almost 50% greater decrease in annual mortality in the trained group (RR, 1.49; 95% CI, 1.10-2.07; $P = .01$) than in the nontrained group.

After adjusting for surgical risk and volume, we found a dose-response relationship for increasing quarters: for every quarter of training, the mortality rate decreased 0.5 per 1000 procedure deaths (95% CI, 0.2-1.0; $P = .001$; Figure). The degree of reported briefing and debriefing in each facility showed that for every increase in degree of briefing and debriefings mortality rate was reduced by 0.6 per 1000 procedures (95% CI, 0.3-0.8; $P = .001$).

Thirty-five training facilities (47.2%) reported at their final interview that they had improved communication among their OR staff. Similarly, 34 (46.0%) reported improved OR staff awareness, and 48 (64.9%) reported an improvement in OR teamwork. (**TABLE 3**) Additional interview results are provided in eTable 2 (available at http://www.jama.com).

COMMENT

The VHA Medical Team Training program was associated with a statistically significant reduction in surgical mortality rate. We used VASQIP mortality data, which is recognized as the gold standard and has been adopted by the American College of Surgeons as its principal quality metric.[16] Although others have shown that team training resulted in improved teamwork, safety attitudes, communication, and reduced errors,[4,5,7,26-31] this is the first large study, to our knowledge, with a contemporaneous control group that demonstrates an association between a medical team training program and reduced surgical mortality rate, both observed and risk adjusted.

Haynes et al[6] reported a decrease in mortality associated with a surgical safety checklist in a project involving 8 hospitals. The primary intervention in that study was the use of a standardized presurgical checklist, but the intervention in the current study was the implementation of the VHA Medical Team Training program. A required component of the program was the implementation of briefings and de-

briefings. Facilities were instructed to develop a checklist to facilitate this process. Taken together, the results of both studies suggest that the use of preoperative checklists, especially to guide a preoperative discussion of the case, may be helpful in lowering surgical mortality. The training program facilitated more open communication in the OR.

Of interest is the dose-response relationship between the number of quarters the training program had been implemented and the rate of surgical mortality. As facilities implemented longer, their rate of surgical mortality decreased further. This suggests that it is critical not only to provide training but also to ensure that the tools are fully integrated into the surgical service. The year-long follow-up was helpful in ensuring that OR clinicians adopted the training tools and changed practice patterns.

During the quarterly facility interviews, we collected detailed adherence data regarding the degree of briefing and debriefing. As a result, we propose some mechanisms regarding how team training strategies and tools may contribute to decreasing mortality. It is our hypothesis that conducting preoperative briefings is a key component in reducing mortality because it provides a final chance to correct problems before starting the case. The use of conducting briefings and debriefings requires a more active participation and involvement than sometimes occurs when a checklist is used by itself. During follow-up interviews, facilities provided specific examples of having avoided adverse events because of the briefing. Surgical teams shared stories such as discovering during the briefings that a patient was anticoagulated or that a patient required cardiac clearance, while others identified the need for additional equipment or implants. These catches could help avoid potential adverse events. We have reported in another study[32] that such discoveries as learning that the correct size of an implant was not available in the OR. Equipment unavailability could

increase the time a patient is under anesthesia thus risking his/her risk of surgical complications.

Teams also shared the value of voicing problems in the debriefing. They reported resolving issues in a timely manner as an improvement attributable to the team training program. Examples included fixing broken equipment or instruments, ordering extra or backup sets of instruments to prevent intraoperative delays, and improving collaboration with radiology for quicker response times. The resolution of such issues likely also prevented potential adverse events. Some specific examples are provided in eAppendix 3 (available at http://www.jama.com).

Limitations

This study has several limitations. One natural concern involves the baseline imbalance in the average mortality rate between the intervention and the control facilities: 17 vs 15 deaths per 1000 procedures (Table 1). Because the study was not randomized, this could indicate the existence of bias in the formation of study groups. For example, the first facilities to complete the training program may also have been those facilities with the greatest likelihood of improvement. To address these concerns, we used propensity score matching to approximate an unbiased design. This method created scores based on baseline characteristics. Then we grouped the facilities by severity scores. We conducted the analyses within these groups to control for these confounders. Without propensity score matching, the difference between facilities in terms of reduction of the mortality rate was almost 2 and half fold (18% in the training group vs 7% in the nontraining group). After using propensity score matching to correct for selection bias, we estimated the difference to be about 50% (RR, 1.49; 95% CI, 1.10-2.07; $P = .01$).

This is a retrospective cohort study and not a prospective randomized trial, although all the mortality data were collected prospectively by researchers who were blind to the study hypothesis.

Mortality data were available and analyzed by facility, and the absence of individual patient data for the 2 groups could be viewed as a limitation. At the same time, the goal of the training program was to change the safety culture in each facility's OR, so analyzing by facility level is consistent with this approach. Unmeasured potential confounders are likely to exist in the nonrandomized study design.

Although the design of the training is not complex, because it was assessed on a quarterly basis while VASQIP mortality rates were provided annually, we made the necessary adjustments. In addition to estimating the marginal reduction in mortality rate (ie, intervention vs control over the same calendar period), we attempted to estimate the dose response where time since enrollment in the Medical Team Training itself was modeled. Because the intervention was grouped on a quarterly basis, this time unit naturally becomes the proxy measure for dose. In short, we carried out 2 types of analyses—one based on calendar time (to address secular trend) and another based on implementation time (to address dose issue). The calendar time analysis showed an almost 50% greater reduction in mortality rate among the trained facilities than those that had not been trained. The implementation time analysis showed a dose effect of 0.5 deaths per 1000 procedures for every additional quarter of the program.

We adopted the more flexible and versatile longitudinal GEE model to analyze the data in order to ameliorate some of the limitations in our study design. For example, in our longitudinal GEE analyses, each facility served as its own control thus enabling us to remove some extraneous, but unavoidable, sources of variability among individual facilities, such as facility location, size, structure, etc. Although the VHA introduced the program because facilities needed to improve communication, it is nevertheless possible that facilities could have started to implement some aspects of the pro-

gram before initiating it. The best way that we could address this limitation was to use facilities as their own controls in the analysis and thus remove the effect of the heterogeneity.

Because there are many factors that could reduce surgical mortality, the inclusion of a contemporaneous control group that was similar to the trained facilities after being matched on propensity scores should have decreased the chance of potential confounding due to existing secular trends. The dose-response relationship between the training program and reduced surgical mortality, together with the inclusion of a contemporaneous control, provides support that observed changes are due to the training rather than other environmental or cultural influences that may have occurred.

Another potential limitation was that information collected about program implementation was by self-report. Self-reports were not confirmed by audits; however, the effect of overreporting or underreporting implementation of briefings and debriefings would be to wash out the effect of the program on changes in mortality and reduce differences between the 2 groups. Finally, the applicability of this study to the general population may be limited because patients who receive treatment from VA facilities have been found to differ from patients in the private sector.[12]

CONCLUSION

Participation in the VHA Medical Team Training program was associated with lower surgical mortality.

REFERENCES

1. Joint Commission. Joint Commission Sentinel Event. http://www.jointcommission.org/NR/rdonlyres/377FF7E7-F565-4D61-9FD2-593CA688135B/0/Statswithallfieldshidden31March2010.pdf. March 31, 2010. Accessed May 23, 2010.
2. Lingard L, Regehr G, Orser B, et al. Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Arch Surg*. 2008;143(1):12-17.
3. Davenport DL, Henderson WG, Mosca CL, Khuri SF, Mentzer RM Jr. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg*. 2007;205(6):778-784.
4. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care*. 2004;13(suppl 1):i85-i90.
5. Awad SS, Fagan SP, Bellows C, et al. Bridging the communication gap in the operating room with medical team training. *Am J Surg*. 2005;190(5):770-774.

6. Haynes AB, Weiser TG, Berry WR, et al; Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009;360(5):491-499.
7. Sorbero M, Farley D, Mattke S, Lovejoy S. *Outcome Measures for Effective Teamwork in Inpatient Care*. Santa Monica, CA: Rand Corp; 2008. http://www.rand.org/pubs/technical_reports/TR462. Accessed May 23, 2010.
8. Dunn EJ, Mills PD, Neily J, Crittenden MD, Carmack AL, Bagian JP. Medical team training: applying crew resource management in the Veterans Health Administration. *Jt Comm J Qual Patient Saf*. 2007;33(6):317-325.
9. Paull DE, Mazzia LM, Izu BS, Neily J, Mills PD, Bagian JP. Predictors of successful implementation of preoperative briefings and postoperative debriefings after medical team training. *Am J Surg*. 2009;198(5):675-678.
10. Musson D, Helmreich RL. Team training and resource management in health care: current issues and future directions. Harvard Health Policy Review. 2004;5(1):25-35. http://www1.va.gov/vhapublications/ViewPublication.asp?pub_ID=1398. Accessed August 2, 2010.
11. Neily J, Mills PD, Lee P, et al. Medical Team Training and coaching in the Veterans Health Administration: assessment and impact on the first 32 facilities in the programme. *Qual Saf Health Care*. 2010;19(4):360-364.
12. Khuri SF, Henderson WG, Daley J, et al; Principal Site Investigators of the Patient Safety in Surgery Study. The patient safety in surgery study: background, study design, and patient populations. *J Am Coll Surg*. 2007;204(6):1089-1102.
13. Khuri SF, Daley J, Henderson W, et al. The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *J Am Coll Surg*. 1995;180(5):519-531.
14. Khuri SF, Daley J, Henderson W, et al; National VA Surgical Quality Improvement Program. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. *Ann Surg*. 1998;228(4):491-507.
15. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005;138(5):837-843.
16. The American College of Surgeons, National Surgical Quality Improvement Program. About ACS NSQIP: history of the ACS NSQIP. https://acsnsqip.org/main/about_history.asp. Accessed December 23, 2009.
17. Daley J, Khuri SF, Henderson W, et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care:

results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg*. 1997;185(4):328-340.
18. Szabo CR; Veterans Healthcare Administration National Leadership Board Human Resources Committee 2005 Facility Complexity Model 2005; August 9, 2005. http://www1.va.gov/vhapublications/ViewPublication.asp?pub_ID=1398. Accessed May 23, 2010.
19. West AN, Lee RE, Shambaugh-Miller MD, et al. Defining "rural" for veterans' healthcare planning [published online ahead of print June 17, 2010]. *J Rural Health*. http://onlinelibrary.wiley.com/doi/10.1111/j.1748-0361.2010.00298.x/abstract. Accessed September 27, 2010.
20. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford, England: Clarendon Press; 1996.
21. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. London, UK: Chapman & Hall; 1989.
22. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
23. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med*. 1997;127(8 pt 2):757-763.
24. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984;79:516-524.
25. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985;39:33-38.
26. Morey JC, Simon R, Jay GD, et al. Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res*. 2002;37(6):1553-1581.
27. Williams KA, Rose WD, Simon R. Teamwork in emergency medical services. *Air Med J*. 1999;18(4):149-153.
28. Grogan EL, Stiles RA, France DJ, et al. The impact of aviation-based teamwork training on the attitudes of health-care professionals. *J Am Coll Surg*. 2004;199(6):843-848.
29. DeFontes J, Subida S. Preoperative safety briefing project. *Permanente Journal*. 2004;8(2):21-27. http://xnet.kp.org/permanentejournal/spring04/awardwin.html. Accessed September 29, 2010.
30. Makary MA, Mukherjee A, Sexton JB, et al. Operating room briefings and wrong-site surgery. *J Am Coll Surg*. 2007;204(2):236-243.
31. Pratt SD, Mann S, Salisbury M, et al. Impact of CRM-based training on obstetric outcomes and clinicians' patient safety attitudes. *Jt Comm J Qual Patient Saf*. 2007;33(12):720-725.
32. Neily J, Mills PD, Eldridge N, et al. Incorrect surgical procedures within and outside of the operating room. *Arch Surg*. 2009;144(11):1028-1034.