0022-5282/92/3302-0292$03.00/0
THE JOURNAL OF TRAUMA

Vol. 33, No. 2
Printed in U.S.A.

INTER-RATER RELIABILITY OF PREVENTABLE DEATH JUDGMENTS

Ellen J. MacKenzie, PhD, Donald M. Steinwachs, PhD, Lee R. Bone, RN, MPH, Douglas J. Floccare, MD, MPH, Ameen I. Ramzy, MD, and The Preventable Death Study Group*

This study examined the inter-rater reliability of preventable death judgments for trauma. A total of 130 deaths were reviewed for potential preventability by multiple panels of nationally chosen experts. Deaths involving a central nervous system (CNS) injury were reviewed by three panels, each consisting of a trauma surgeon, a neurosurgeon, and an emergency physician. Deaths not involving the CNS were reviewed by three panels, each consisting of two trauma surgeons and an emergency physician. Cases for review were sampled from all hospital trauma deaths occurring in Maryland during 1986. Panels were given prehospital and hospital records, medical examiner reports, and autopsy reports, and asked to independently classify deaths as not preventable (NP), possibly preventable (POSS), probably preventable (PROB), or definitely preventable (DEF). Cases in which there was disagreement about preventability were discussed by the panel as a group (via conference call). Results indicated that overall reliability was low. All three panels reviewing non-CNS deaths agreed in only 36% of the cases ($\hat{\kappa} = 0.21$). Agreement among panels reviewing CNS deaths was somewhat higher at 56% ($\hat{\kappa} = 0.40$). Most of the disagreements, however, were in judging whether deaths were NP or POSS. Agreement was higher for early deaths and less severely injured patients. For non-CNS deaths agreement was also higher for younger patients. When both autopsy results and prehospital care reports were available reliability increased across panels. A variety of approaches have been used to elicit judgments of preventability. This study provides information to guide recommendations for future studies involving implicit judgments of preventable death.

PREVENTABLE DEATH studies have been widely used to evaluate outcomes of trauma care and performance of trauma systems.[1,2] The preventable death rate (PDR), which is generally defined as the proportion of all deaths judged to have been preventable if optimal care had been delivered, has become a standard tool for measuring quality of trauma care.

In recent years, however, several investigators have questioned the reliability and validity of the method.[2-5] Of particular concern is the extent to which implicit judgments of preventability are reproducible when they are made by different experts or groups of experts. A variety of approaches have been used to elicit preventable death judgments; however, few studies have systematically examined the reliability of these alternative methods. The purpose of this study was to examine the inter-rater reliability of preventable death judgments for trauma and to identify characteristics of the method that are associated with lower or higher levels of agreement.

METHODS

The reliability of preventable death judgments was assessed as part of a larger effort to examine the relationship between the rate of preventable trauma deaths and the appropriateness of the level of hospital care received (i.e., trauma center vs. non-trauma center hospital).

For the larger study, a total of 256 trauma deaths were sampled from all trauma deaths occurring in Maryland's 52 acute care hospitals during 1986. A trauma death was identified as any death with underlying cause recorded on the death certificate in the ICD E-code range 800–999 excluding drownings, poisonings, inhalation and ingestion of food or foreign bodies, medical misadventures, and late effects of the incident injury. All patients who were declared dead in the hospital but who were either pulseless (blunt trauma) or had no electrical activity (penetrating trauma) on hospital arrival were excluded. The sample for the main study was stratified to achieve approximately equal numbers of deaths in four categories defined by (1) presence or absence of a significant (AIS ≥ 2) CNS injury, and (2) appropriateness of the level of hospital care

* See Appendix.

From The Johns Hopkins Health Services Research and Development Center, Johns Hopkins School of Hygiene and Public Health, and The Maryland Institute for Emergency Medical Services Systems, University of Maryland, Baltimore.

Presented at the Fifth Annual Scientific Assembly of the Eastern Association for the Surgery of Trauma, January 16–18, 1992, Hamilton, Bermuda.

Supported in part from Grant HS 05630 from the Agency for Health Care Policy and Research, U.S. Department of Health and Human Services.

Address for reprints: Ellen J. MacKenzie, PhD, The Johns Hopkins HSR&D Center, 624 N. Broadway, Room 639, Baltimore, MD 21205.

received (i.e., level of care appropriate vs. level of care lower than appropriate).[6]

Sample Selection for Reliability

All deaths in the larger study were reviewed for preventability by at least one of six expert panels. To examine reliability, a subsample of deaths was reexamined by an additional two panels. The subsample was chosen to include 100% of deaths judged in the initial review as preventable by at least one member of the expert panel and 20% of all deaths judged not preventable by all three panel members. This sampling strategy provided a total of 130 deaths for the reliability study: 67 deaths in which there was evidence of a CNS injury (CNS cases) and 63 in which there was no significant CNS injury (non-CNS cases).

Review Panels and Review Process

Six panels of nationally chosen experts were formed to judge the preventability of trauma deaths. Three of the panels consisted of two trauma surgeons and one emergency physician; each of these panels reviewed all 63 non-CNS cases. The remaining three panels included one trauma surgeon, one neurosurgeon, and one emergency physician; each of these panels reviewed all 67 CNS cases. (For the names and affiliations of panel members see the *Appendix*.) All panel members were board certified in their respective specialties and had at least 5 years of experience in treating multiple trauma patients. In addition, all practiced outside the study area (the State of Maryland).

The review of cases proceeded in two stages. First, panel members were sent the appropriate records for review and asked to independently classify each death according to its potential preventability as either (1) not preventable (NP); (2) possibly preventable (POSS); (3) probably preventable (PROB); or (4) definitely preventable (DEF). Panels were instructed to define a death as preventable (POSS, PROB, or DEF) only if the patient received inadequate care and the inadequate care was judged to have contributed to the death. These initial judgments were sent to the coordinating center at Johns Hopkins and summarized for each panel. A conference call was then scheduled for the panel to discuss all cases in which one or more panel members initially judged the death as potentially preventable. In addition to classifying deaths as to preventability, panels were asked to indicate the nature of the inadequacy of care and the immediate cause of each death (e.g., exsanguination, multiple organ failure/sepsis, CNS injury etc.).

Information Provided for Review

To the extent possible, panel members were provided with the following records for review: the prehospital case report, the complete emergency department and hospital record, the medical examiner's and autopsy reports, and the death certificate. Before distribution among panels, all records had specific dates as well as the names of hospitals, clinicians, and patients removed.

The completeness of the records available for review varied by case. A medical examiner's report and death certificate were available for each case in the study. Hospital records were also obtained for all but eight of the 130 cases. Autopsies, on the other hand, were performed in only 70 deaths (54%). Prehospital case reports were also not uniformly available even though all but two patients were transported by ambulance or helicopter. Prehospital reports were obtained for 71 (55%) of the patients who died who were included in the reliability subsample.

Autopsies were more often performed on younger patients (73% of dead patients < 65 years old had autopsies vs. 18% of those >65), patients who died in the first 24 hours versus one or more days postinjury (66% vs. 24%), and on patients with ISS scores of 25 and above versus below 25 (70% vs. 31%). The increased availability of prehospital records was similarly correlated with younger age, early deaths, and higher injury severity, although the differences were not as large. Given the large proportion of cases in which there were no autopsy or prehospital records, panelists were asked to indicate whether their judgment of preventability was affected by the inadequacy of the information provided.

Together with the records, panels were provided with an abstract summary of the case including demographics, transfer status, mode of arrival, cause of injury, Glasgow Coma Scale score on admission (if available), and key time variables (i.e., estimated time of injury, time until arrival at the scene by prehospital care providers, time until arrival at hospital, time until transfusion, time until general surgical evaluation, time until disposition to OR, ICU, or ward, and time to death). To assist in their review, each panel member was also provided with a structured evaluation form in which he or she could check the components of the prehospital, emergency department, operating room, ICU, and ward care felt to be inadequate or not timely. The use of the worksheet, however, was optional.

Analysis

The analysis of reliability included the following steps. First, agreement among physicians within panels regarding their independent classification of deaths was examined (intra-panel agreement). To determine whether the extent of agreement varied by specialty, agreement in independent assessments was also determined within each of the three specialties represented on the panels: trauma surgeons (TS), neurosurgeons (NS), and emergency physicians (EP) (intra-specialty agreement).

The final panel judgments of preventability were then compared across panels (inter-panel agreement). Inter-panel agreement was further examined by characteristics of the victims and their injuries (i.e., age, Injury Severity Score, time to death, and place of death) and by the quality of the information available to the physicians (i.e., availability of autopsy (yes or no); availability of prehospital report (yes or no); and whether, in the judgment of the panel, the absence or legibility of the documentation made it difficult to evaluate the case). The relationship between these characteristics and the extent of agreement was examined using contingency table analysis and the appropriate chi-square statistics.

Agreement was defined at three levels. Perfect three-way agreement occurred when physicians and panels agreed in their classification of deaths into one of three categories of preventability: NP vs. POSS vs. PROB/DEF. Cases in which death was judged definitely preventable were grouped together with those in which it was judged probably preventable because of the very small number of deaths classified as DEF. To better understand the nature of disagreements, two additional levels of agreement were defined: (1) agreement as to whether a death was NP vs. POSS or PROB/DEF, and (2) agreement as to whether a death was NP or POSS vs. PROB/DEF. In each case, the percentage of cases in which all physicians (or panels) agreed was computed.

To adjust these percentages for the amount of agreement expected by chance, Kappa-type statistics ($\hat{\kappa}$) were computed using the methodology described by Fleiss.[7] The κ statistic can range from -1 to $+1$. It is negative when expected agreement is greater than observed, 0 when expected agreement is equal to observed agreement, and positive when observed agreement is greater than expected. If there is perfect agreement, $\hat{\kappa}$ is equal

to +1. Landis proposed guidelines to judge the relative degree of importance associated with different Kappa statistics. In general, Kappa statistics in the range 0.41 to 0.60 represent moderate agreement, with those above 0.60 representing substantial to almost perfect agreement. Kappas between 0.21 and 0.40 reflect fair agreement and those below 0.20 only negligible agreement beyond chance. All measures of agreement were determined separately for non-CNS cases and CNS cases.

Finally, to determine the magnitude of the impact of disagreements on estimates of the preventable death rate, the number of deaths that would have been judged preventable using three different methods of review were calculated. These methods were chosen to reflect those commonly used in previously published preventable death studies. They include (1) independent review of deaths with the final judgment taken as the majority opinion: if there is no majority option, the case is discussed and a consensus developed (modified majority rule); (2) independent review of deaths with assignment of preventability only to cases in which all reviewers agree the death is preventable (unanimous decision rule); and (3) independent review of deaths followed by panel discussion of all cases in which at least one physician judged death as POSS or PROB/DEF preventable. The classification of preventability reflects the consensus of the panel or the majority opinion if no consensus is reached (panel consensus rule).

RESULTS

In four of the 130 sampled cases, one or more of the panels determined that the information provided to them was not sufficiently adequate to judge preventability. Two of these cases were missing the prehospital record, one was missing the hospital record, and one contained complete records but the hospital record was not legible because it was xerographically copied from microfiche. These cases were excluded from further analysis, leaving 64 non-CNS and 62 CNS deaths for the assessment of reliability.

Intra-Panel and Intra-Specialty Agreement

Agreement among panel members in their independent assessments of preventable death were uniformly low (Table 1). The percentage of perfect agreement ranged from 34% to 47% ($\hat{\kappa}$: 0.17–0.25) for non-CNS deaths and from 29% to 48% ($\hat{\kappa}$: 0.17–0.18) for CNS deaths. Agreement as to whether the death was NP vs. POSS or PROB/DEF was somewhat higher, although the percentage of agreement still averaged only 47% (range, 42%–56%) for non-CNS and 51% (range, 47%–56%) for CNS. Chance-corrected Kappa statistics were between 0.19 and 0.39, indicating only fair agreement. The majority (>70%) of the physicians, however, agreed in distinguishing between deaths that were NP or POSS vs. PROB/DEF. These levels of agreement should be interpreted with caution, however, given the relatively small number of deaths judged PROB/DEF. Even though the sample was specifically chosen to over-represent potentially preventable deaths, the resulting numbers were small. On average, a physician judged only 12 deaths (9%) as PROB/DEF. For this reason the chance-corrected agreement statistics for NP vs. POSS or PROB/DEF comparisons are generally low, indicating that observed agreement is not substantially higher than one would expect to obtain by chance alone.

Agreement of independent judgments was also examined by specialty. Intra-specialty agreement statistics provided no systematic evidence to suggest that reliability of preventable death judgments is any better or worse for any particular discipline.

Table 1
Intra-panel and intra-specialty agreement

	Percent Agreement ($\hat{\kappa}$)		
	NP vs. POSS vs. PROB/DEF	NP vs. POSS/ PROB/DEF	NP/POSS vs. PROB/DEF
Non-CNS Cases			
Panel A	39 (0.17)	42 (0.19)	84 (0.28)
Panel B	47 (0.25)	56 (0.36)	81 (0.22)
Panel C	34 (0.18)	45 (0.27)	70 (0.12)
Trauma surgeons—Gp1*	41 (0.20)	48 (0.26)	88 (0.42)
Trauma surgeons—Gp2*	28 (0.11)	41 (0.20)	66 (0.03)
Emergency physicians	34 (0.16)	45 (0.24)	80 (0.11)
CNS Cases			
Panel D	42 (0.17)	50 (0.26)	74 (0.18)
Panel E	29 (0.18)	47 (0.29)	71 (0.26)
Panel F	48 (0.18)	56 (0.28)	85 (0.05)
Neurosurgeons	47 (0.26)	56 (0.40)	74 (0.28)
Trauma surgeons	44 (0.27)	56 (0.38)	72 (0.09)
Emergency physicians	55 (0.32)	60 (0.36)	89 (0.27)

* Gp1 and Gp2 refer to arbitrary selection of trauma surgeons from panels A, B, and C.

Inter-Panel Agreement

Agreement among consensus panel judgments is summarized in Table 2. Extent of agreement among the panel judgments versus independent physician judgments within each panel was generally the same for non-CNS deaths and only slightly higher for CNS deaths. For CNS deaths, chance-corrected agreement statistics were high enough to indicate a moderate degree of reliability among panel judgments of preventability. To further characterize level of agreement, pairwise agreement statistics (i.e., agreement between two as opposed to all three panels)

Table 2
Inter-panel agreement

	Percent Agreement ($\hat{\kappa}$)		
	NP vs. POSS vs. PROB/DEF	NP vs. POSS/ PROB/DEF	NP/POSS vs. PROB/DEF
Non-CNS Cases			
Among all 3 panels	36 (0.21)	39 (0.24)	86 (0.47)
Panel A vs. Panel B	58 (0.21)	61 (0.20)	95 (0.62)
Panel B vs. Panel C	56 (0.25)	62 (0.27)	89 (0.48)
Panel A vs. Panel C	56 (0.26)	65 (0.30)	86 (0.37)
CNS Cases			
Among all 3 panels	56 (0.40)	66 (0.54)	79 (0.30)
Panel C vs. Panel D	65 (0.37)	79 (0.57)	81 (0.54)
Panel D vs. Panel E	64 (0.35)	72 (0.44)	82 (0.20)
Panel C vs. Panel E	76 (0.51)	84 (0.64)	92 (0.41)

were calculated. Pairwise percentage of perfect agreement averaged 57% (average $\hat{\kappa} = 0.24$) for non-CNS deaths and 68% (average $\hat{\kappa} = 0.41$) for CNS deaths.

To examine whether any one physician tended to dominate the panel discussion and disproportionately influence the final judgment, percentage of agreement between each physician's initial classification and the final panel judgment was calculated. As can be seen in Table 3, there was no consistent evidence to support this hypothesis.

Correlates of Agreement

When the percentage of perfect agreement among the three panels was examined by characteristics of the victim and the injury, several trends were noted (Table 4). Agreement was significantly higher for injuries of higher severity and deaths that occurred soon after the injury ($p < 0.05$). In addition, for non-CNS cases, agreement was higher for younger versus older patients (<65 vs. >65 years), although the difference was not statistically significant, possibly because of the small sample size.

Inter-panel agreement was also higher for cases in which autopsy and prehospital care reports were available ($p < 0.10$). Agreement was also higher but not statistically significant for non-CNS deaths in which panels indicated that judgments were difficult to make because of lack or illegibility of the information provided to them.

Given that older, less severely injured, and late deaths were less likely to have autopsies and prehospital care reports, stratified analyses were performed to determine whether the associations between agreement and characteristics of the deaths were confounded by the quality of information available for review. Based on the appropriate Mantel-Haenszel statistics, there is little evidence to suggest that this is occurring ($p > 0.10$). Percentage of agreement is higher for these subgroups of deaths regardless of whether the autopsy and prehospital report were available.

Table 3
Independent physician* judgment vs. final panel judgment

	Percent Agreement
Non-CNS Cases	
Trauma surg. (Panel A-1)	74
Trauma surg. (Panel A-2)	73
Trauma surg. (Panel B-1)	78
Trauma surg. (Panel B-2)	73
Trauma surg. (Panel C-1)	72
Trauma surg. (Panel C-2)	73
Emerg phys. (Panel A)	70
Emerg phys. (Panel B)	76
Emerg phys. (Panel C)	55
CNS Cases	
Trauma surg. (Panel D)	72
Trauma surg. (Panel E)	68
Trauma surg. (Panel F)	84
Neurosurg. (Panel D)	70
Neurosurg. (Panel E)	61
Neurosurg. (Panel F)	73
Emerg phys. (Panel D)	71
Emerg phys. (Panel E)	72
Emerg phys. (Panel F)	75

* TS = trauma surgeon; NS = neurosurgeon; EP = emergency physician.

Table 4
Percent perfect agreement among panels by characteristics of patients and quality of information

Characteristic	Percent Perfect Agreement Among Panels		
	All Deaths	Non-CNS	CNS
Patient Age (years)			
<65	51	42	58
65+	38	29	53
Time to Death (hours)			
<1	70*	57	80
2–6	44	35	56
7–24	55	42	63
25+	30	27	33
Place of Death			
Trauma Center	50	39	57
Nontrauma Center	38	33	56
ISS			
<16	26*	26	—
16–24	30	36	25*
25–34	60	46	68
35+	59	50	62
Overall Quality Affected Judgment?			
Yes	46	25	67
No	47	38	55
Autopsy Available?			
No	38†	28	50
Yes	53	44	62
Prehospital Record Available?			
No	30*	17*	44†
Yes	58	51	66

* $p < 0.05$.
† $p < 0.10$.

Table 5
Total number of POSS or PROB/DEF preventable deaths (number of *PROB*/DEF preventable deaths) by decision rule

	Modified Majority Rule	Unanimous Decision Rule	Panel Consensus Rule
Non-CNS			
Panel A	24 (4)	7 (1)	32 (6)
Panel B	23 (3)	7 (1)	21 (5)
Panel C	30 (8)	15 (1)	35 (10)
CNS			
Panel D	20 (5)	8 (1)	23 (6)
Panel E	29 (8)	14 (2)	29 (12)
Panel F	13 (2)	6 (0)	19 (3)

Impact on Preventable Death Rate

The number of deaths judged preventable varied according to the method of review and rules used to declare

a death preventable (Table 5). The unanimous decision rule is clearly the most conservative method for defining preventable deaths. In general, the number of deaths judged preventable using this rule was less than half the number of deaths judged preventable using either the modified majority rule of the panel consensus rule. On average, the panel consensus rule yielded three to four more potentially preventable deaths (and two more PROB/DEF preventable deaths) than the majority decision rule. None of the approaches, however, appeared to stand out as substantially more reliable, as evidenced by the large discrepancies in numbers across panels.

DISCUSSION

Since the early 1960s more than 30 preventable death studies have been published in the trauma literature.[5,8–41] For the most part, these studies have been based on implicit judgments made by physician panels of varying size and composition. In few of these studies, however, has the issue of reviewer reliability been addressed. The present study represents an attempt to systematically examine the reliability of preventable death judgments for trauma and the factors affecting levels of reliability.

Results indicate that in general, reliability of implicit preventable death judgments for trauma is low. All three panels reviewing non-CNS deaths agreed in only 36% of the cases ($\hat{\kappa} = 0.21$), whereas agreement among panels reviewing CNS deaths was 56% ($\hat{\kappa} = 0.40$). These results are similar to those of Dubois and Brook, who investigated the reliability of preventable death judgments for 182 patients with one of three medical conditions.[42] In their study, preventability was judged independently by three physicians; reported Kappa statistics were 0.4, 0.3, and 0.2, respectively, for patients with a diagnosis of a cerebrovascular event, myocardial infarction, and pneumonia. Dubois and Brook commented that although these reliability coefficients are low, they do not differ considerably from those found in other studies examining the consistency of physician judgments regarding the occurrence of an event (e.g., presence or absence of a cardiac murmur or peripheral pulses).[43]

The only study we found that examined reliability of preventable death judgments for trauma was by Reines and Duffy.[27] They examined reliability of preventable death judgments independently determined by five physicians, including one neurosurgeon, one general surgeon practicing in the study area, one general surgeon practicing outside the study area, one trauma surgeon practicing at the level I trauma center in the study area, and one military surgeon. These authors found that at least three of the five reviewers agreed on all cases; all five physicians agreed on approximately 64% of the cases. These results are not directly comparable with the present study, however, since preventability of death in the majority (81%) of the cases was judged using autopsy and medical examiner data only; hospital records were not available for review. In addition, the lack of chance-corrected agreement statistics for the Reines study make comparisons difficult.

Not unexpectedly, reliability in the present study was significantly higher for early deaths and more severe injuries. These results confirm the observation by Cayten et al. that determination of preventability is more difficult the longer the patient lives.[41] For non-CNS deaths, reliability was also related to the age of the patient. The lower reliability of preventable death judgments for older patients likely reflects general disagreement among physicians regarding the relative benefits of aggressive trauma care for the elderly. During the conference calls considerable discussion ensued regarding the potential preventability of death among elderly trauma patients. There was significant disagreement among panelists as to whether underlying disease and general frailty would in many cases have led to the demise of the patient regardless of whether care was judged to be optimal. In addition, questions of longer term survival and quality of life were continuously raised. Despite the magnitude of this problem of injuries in the elderly, we know comparatively little about outcomes of trauma in this age group or what factors contribute to outcomes that are poorer than expected.[44–49]

The low reliability of implicit preventable death judgments found in this study underscores the need to standardize the methodology. A review of preventable death studies conducted since 1975 indicates a general lack of consensus regarding the size and composition of review panels, the information needed to judge preventability, and the methods used to elicit and summarize judgments of preventability (Table 6). Although the present study was not designed to explicitly compare various approaches, the results provide insight into some of the methodologic concerns.

Review Process and Decision Rule

Most of the preventable death studies published since 1975 used either a majority decision rule or a panel consensus rule to translate judgments of multiple physicians into a final classification of preventability. Only a few studies have used the unanimous decision rule. Because the true rate of preventable deaths is not known, the accuracy of these alternative approaches cannot be determined. It appears, however, that the unanimous decision rule provides an estimate of the lower bound of the true preventable death rate and the panel consensus approach (using a multidisciplinary panel), an estimate of the upper bound. Depending on the objective of the preventable death review one might choose one method over the other. When results are reported, however, it is important that the method used be clearly delineated and appropriately interpreted.

The use of panel discussions following independent

reviews as opposed to using a majority or modified majority rule based on independent review only provides a marginal increase in reliability. This increase is most apparent for deaths involving a CNS injury. In general, the panel discussions resulted in more deaths judged potentially preventable. Convening panels can be costly, however. The results of this study suggest that one can approximate the upper bound of the preventable death rate reasonably well using a modified majority rule in which only the cases where there is no majority opinion are discussed.

Composition of Panels

The composition of physician review panels also varied among the studies reviewed. Some of the panels included trauma surgeons only, whereas other panels were multidisciplinary, often including neurosurgeons, emergency physicians, and occasionally pathologists and anesthesiologists. Studies also varied with respect to level of experience of reviewers in treating trauma patients as indicated by years in practice or practice at a level I trauma center vs. a community hospital. Although the present study did not explicitly compare judgments based on panels of varying composition, it was evident from the panel discussion that there is substantial value in using a multidisciplinary review panel. Lowe et al. cautioned against the use of multidisciplinary panels and recommended the use of general trauma surgeons only in judging preventability.[26] In their study, preventable deaths were identified using a six-member panel including three general surgeons, one neurosurgeon, one emergency physician, and one anesthesiologist. The attribution of preventability required agreement of only two or more panel members. All reviews were performed independently, however, with no discussion by the panel as a group. They found a "broad range of responses" by the individual panel members, with the neurosurgeon focusing for the most part on the management of neurological injuries as opposed to general surgery problems. Because of the substantial lack of agreement among the independent reviews, Lowe and colleagues indicated they had to rely for the most part on the judgments of the general surgeons. Their results are not surprising but their conclusions are perhaps misleading. Even though multidisciplinary panels contribute to disagreement, they provide the range of clinical expertise necessary to evaluate all aspects of patient care. For a multidisciplinary panel to function effectively, however, there must be an opportunity for discussion of perspectives among its members.

It is important to emphasize that, with few exceptions, physicians agreed about the nature of the inadequacy of care (e.g., delay to surgical intervention, inadequate ICP monitoring, delay in diagnosis). The disagreement in preventable death judgments instead concerned the extent to which the physicians believed the inadequacy of care contributed to the death. Discussions among panel members typically focused on the severity of injury, the status of the patient on arrival at the hospital, and the extent to which age and pre-existing chronic disease influenced probability of survival.

Classification of Deaths

Multiple approaches have been used for classifying deaths. Most studies have used either a two-category (preventable or not preventable) or a three-category classification (preventable, possibly preventable, or not preventable). We have found the distinction between not preventable and possibly preventable to be problematic and to contribute significantly to lower reliability. Based on this finding, we recommend the use of at least a three-point scale (i.e., not preventable, possibly preventable, and probably or definitely preventable). Furthermore, a distinction should be made between preventable death rates based on possibly and probably/definitely preventable deaths and probably/definitely preventable deaths alone. The distinction between probably and definitely preventable was generally difficult for physician judges to make and therefore may not be a useful category. In only two non-CNS cases and one CNS case did panels ultimately conclude that death was definitely as opposed to probably preventable.

It is likely that reliability in this study could have been increased if panels had been provided with more explicit guidelines for distinguishing among the four categories of preventability. Shackford et al. have developed explicit guidelines for judging preventability.[34] These guidelines are currently used by San Diego County in their prospective audit of all trauma deaths.[35] The guidelines include the classification of deaths according to the probability of survival (Ps) calculated on the basis of the Revised Trauma Score, age, ISS, and TRISS analysis.[50] In general, reviewers are instructed to consider cases with $Ps < 0.25$ as not preventable, those with $0.25 < Ps < 0.50$ as potentially preventable, and those with $Ps > 0.50$ as frankly preventable.

Information Available for Review

The type and completeness of records available for review by physician panels has varied across published studies. Approximately two thirds of the preventable death studies published since 1975 have used the hospital record together with autopsy results and the medical examiner's report as a basis for review (referred to as the clinical method). The remaining one third used autopsy results only (autopsy method). The general consensus is that wherever possible, both clinical and autopsy records should be made available for review. This consensus is derived from early work by West, who showed that review of clinical records and autopsy results identifies more deaths as preventable than does review of autopsy data alone.[20] The discrepancy in results was particularly pronounced for CNS deaths. For this reason, more recent

Table 6
Comparison of preventable death study methodology: 1975–1991

Authors	Injuries[a]	Number of Deaths[b]	Panel Size	Panel Composition[c]	Panel Outside Study Area?	Categories in Definition of Preventability	Blind Review?	Information Available	Decision Rules
Moylan et al.[16]	Major trauma; CNS and nCNS	57	4	4 GS	No	2	Yes	HOSP	NS
Foley et al.[17]	Vehicular abdom. injuries	43	3	NS	No	2	NS	HOSP; AUT	NS
West et al.[19]	Vehicular CNS and nCNS	40 92	3	3 TS	No	3	No	AUT for Orange Co. HOSP; AUT for SF Co.	Panel consensus
Baker et al.[21]	All causes CNS and nCNS	136	5	5 TS	No	2	No	HOSP; AUT	NS
Neuman et al.[23]	All causes CNS and nCNS	177	6	NS	No	2	Yes	AUT	Modified unanimous decision (5/6 had to agree)
West *et* al.[20]	Vehicular nCNS	64	4	2 GS; 1 EP; 1 NEU	Yes	2	NS	PRE; HOSP; AUT	Modified unanimous decision (For CNS cases, judgment of preventable by NEU sufficed)
West et al.[22]	Vehicular nCNS	29	3	2 TS; 1 EP	No	2	No	ABSTR based on AUT/ME	Panel consensus
Lowe et al.[26]	Severe vehicular CNS and nCNS	135	6	3 GS; 1 NEU; 1 EP; 1 ANES	Yes	2	Yes	ABSTR based on PRE; HOSP; AUT	Independent reviews—2 or more had to agree
Reines and Duffy[27]	All causes CNS and nCNS	97	5	2 GS; 1 TS; 1 NEU; 1 MS	Both	3	Yes	AUT for 79 cases HOSP; AUT for 18 cases	Majority
McKoy and Bell[25]	Pediatric; All causes; CNS and nCNS	100	2	2 PTS	No	2	NS	ME/AUT	NS
Certo et al.[24]	Vehicular nCNS	45	NS	TS	No	2	No	HOSP; AUT	NS
Cales[30]	Vehicular CNS and nCNS	58 60	4	2 GS; 1 NEU; 1 EP	No	2	No	PRE; HOSP; AUT	Modified majority
Krob[32]	Vehicular CNS and nCNS	100	4	3 GS; 1 NEU	NS	3	Yes	ABSTR based on HOSP; AUT (when available)	Panel consensus
Ramenofsky et al.[29]	Pediatric; All causes; CNS and nCNS	100	4	2 TS; 1 PTS; 1 NEU	NS	2	NS	ABSTR based on PRE; HOSP; AUT	Unanimous decision
Spain et al.[31]	Unintentional injuries CNS and nCNS	421	5	1 TS; 1 NEU; 1 PATH; 1 CARD; 1 THOR	Both	3	Yes	AUT	Initial screen by PATH identified 30 cases as POSS preventable: these were reviewed using panel consensus rule.

Baker et al.[33]	Adult admissions to SICU; CNS and nCNS	31 20	2	1 TS; 1 T Nurse	No	3	No	NS	Modified majority
Kreis et al.[36]	All causes, nCNS	246	4	4 TS	No	3	NS	ABSTR based on PRE; HOSP; AUT/ME	Panel consensus
Rivara et al.[5]	Ped. and bicycle collisions with motor vehicles; CNS and nCNS	84	6	5 TS; 1 EP	No	3	Yes	ABSTR based on PRE; HOSP; AUT	Modified unanimous decision (5/6 had to agree)
Shackford et al.[34] [Also Shackford et al. 1986[35] and Davis et al. 1991[39]]	All causes CNS and nCNS	90 112	NS	TS; NEU; EP; PATH; ANES	No*	3	No	ABSTR based on PRE; HOSP; AUT	Prospective audit with panel consensus
Campbell et al.[37]	All causes, nCNS	62	3	3 TS	No	2	NS	ABSTR based on AUT/ME	Panel consensus
Guss et al.[38]	All causes; CNS and nCNS; Excluded deaths within 1 hour of arrival	211	6	NS	No	2	Yes	AUT	Modified unanimous decision (5/6 had to agree)
Webb et al.[40]	All causes; CNS and nCNS, ED deaths only	186	3	NS	No	2	NS	PRE; HOSP; AUT	NS
Cayten et al.[41]	All causes; CNS and nCNS; Excluded distal fxs and dislocations; TRISS $P_s \geq 0.50$	233	3	3 TS	No	2	Yes	ABSTR based on PRE; HOSP; AUT	Panel consensus with independent review

[a] All causes generally excluded drownings, electrocutions, suffocations, and often burns without other trauma.

[b] Numbers connected by "&" indicate size of individual groups compared in study.

[c] Abbreviations: nCNS = non central nervous system; TS = trauma surgeon; GS = general surgeon; EP = emergency physician; NEU = neurosurgeon; PATH = pathologist; ANES = anesthesiologist; PTS = pediatric trauma surgeon; PRE = prehospital record; HOSP = hospital record; AUT = autospy; ABSTR = record abstract; NS = not specified.

* Included annual extramural review by medical audit committee.

studies using the autopsy method have included only non-CNS deaths in their review.

Few of the studies that have used the clinical method, however, discuss the extent to which actual information provided to reviewers was adequate enough to judge preventability. In some studies, autopsies were not uniformly performed in all cases of death. Even less clear is the extent to which prehospital reports were included. We found only seven studies that explicitly stated that prehospital records were included for review, although in some cases prehospital reports may have been available as part of the hospital record or medical examiner's report. Of equal if not greater concern is the extent to which prehospital reports provided complete data on critical issues such as scene and transport times and treatment provided. Lowe et al., for instance, reported that in 14% of the cases in their study, preventability could not be determined because of lack of autopsy data or transport and field times.[26]

We found that the completeness of the record substantially affects the reliability of the review. The percentage of perfect agreement among the three panels increased from 38% to 53% when autopsy reports were available and from 30% to 50% when prehospital records were available. These results argue strongly for making available, where possible, complete information including the prehospital care report, hospital record, and autopsy report. When complete records are not available, this should be noted and an attempt made to assess the potential impact of the inadequacy of the information on the results.

Several studies have provided physician panels with an abstract summary of the case in lieu of a complete set of records for review. Summaries have varied from a one-page narrative to a 14-page abstract of critical data elements. Although this approach uses the time of the expert panel most efficiently, it has not been widely tested for reliability and accuracy. Dubois and Brook provide some data on a small number of cases (n = 18) to suggest that physicians agree in their judgments whether provided with a summary or a complete series of photocopied medical records.[42] Cayten et al., on the other hand, indicate that, especially for deaths from low falls, review of the entire medical record provides a better basis for deciding preventability.[41] More work is needed to determine the adequacy of the abstract record approach.

SUMMARY

Preventable death studies have provided one important method for measuring the effectiveness of trauma care and trauma systems. For this reason, it is imperative that more attention be paid to improving the reliability of the approach. The results of the present analysis suggest several recommendations.

First, although inter-rater reliability of preventable death judgments was found to be generally low, there is considerable room for improvement if the information made available for review is complete, including the medical record, prehospital care report, and autopsy results. The results underscore the importance of autopsy reports in judging preventability and reinforce the need for legislation that would mandate autopsies for all trauma deaths.[51] In addition, it is important that mechanisms be established to ensure that both the prehospital record and the autopsy report be routinely included in the medical record. In the current study, prehospital, hospital, and autopsy records were obtained from several different sources. This process was time consuming and costly, making routine system-wide assessments difficult.

Based on the results of this and other studies, we recommend that in analyzing and reporting the results of a preventable death review, cases be stratified by patient age and injury severity, since reliability of judgments was highly correlated with these characteristics. Further, as indicated, preventability should be determined and reported as a three-way classification allowing for a distinction between deaths judged possibly vs. probably or definitely preventable.

Perhaps most important, serious consideration should be given to including an evaluation of reliability in all preventable death studies. To the extent possible, reviewers should be trained and standardized before the study. The relatively poor reliability of preventable death judgments also needs to be taken into account in determining the sample size used in preventable death evaluations, especially when comparing systems of care. As reliability declines, the sample size required to demonstrate a significant difference increases substantially.

Last, the present study emphasizes the need to use a broader range of indicators than just preventable death for assessing the effectiveness of trauma care systems. These should include measures of the appropriateness of care received as well as statistically based estimates of lives saved.

As noted, there were very few instances in which physician judges disagreed about the adequacy or appropriateness of care. Rather, the discussions concerning preventability centered on the extent to which errors in management contributed to the death. Yet very few preventable death studies published to date have systematically and uniformly documented inadequacies of care independent of preventability. A few notable exceptions are studies conducted by Moylan et al.,[16] Shackford and colleagues,[34,35,39] and Rivara et al.[5] Given the low reliability of implicit preventable death judgments, more emphasis should be placed on developing and using explicit criteria for measuring appropriateness of care, linking appropriateness to the actual outcomes, and measuring the potential preventability of adverse outcomes.

In addition to documenting appropriateness, preventable death reviews should incorporate more objective statistical measures of the probability of survival (e.g.,

TRISS probabilities) either as additional information available to judges at the time of review or in the analysis and interpretation of results. Estimates of survival probability derived from the actual experience of large numbers of trauma patients treated in more than 100 institutions across the country (as represented by the Major Trauma Outcome Study) provide a firmer base on which to judge potential salvageability.[52] Unfortunately, the information required for computing TRISS probabilities is not uniformly recorded in medical records, especially those maintained by non-trauma center hospitals. In the present study, Revised Trauma Scores were retrieved for only half the patients. The lack of adequate documentation of critical variables required to characterize the severity of injuries remains an impediment to effective evaluation of trauma care.

Careful attention to these issues should enhance the reliability and hence the validity of the preventable death review as a method for evaluating trauma care and trauma systems.

Appendix

The Preventable Death Study Group:

Carl R. Boyd, M.D.
Memorial Medical Center
Savannah, Georgia

G. Richard Braen, M.D.
Buffalo General Hospital
Buffalo, New York

Charles G. Brown, M.D.
Ohio State University
Columbus, Ohio

Sylvia D. Campbell, M.D.
217 S. Matanzas Avenue
Tampa, Florida

James W. Carter, M.D.
University of Saskatchewan
Regina, Saskatchewan

C. Gene Cayten, M.D.
Our Lady of Mercy Medical Center
Bronx, New York

William F. Fallon, JR., M.D.
University Hospital of Jacksonville
Jacksonville, Florida

Douglas A. Gentile, M.D.
Vanderbilt University
Nashville, Tennessee

James M. Hassett, Jr., M.D.
Buffalo General Hospital
Buffalo, New York

Robert C. Mackersie, M.D.
University of California at San Diego
San Diego, California

Wendy J. Marshall, M.D.
Loyola University
Maywood, Illinois

Ricardo Martinez, M.D.
Standford University Hospital
Stanford, California

Raj K. Narayen, M.D.
Baylor College of Medicine
Houston, Texas

Loren D. Nelson, M.D.
Vanderbilt University Medical Center
Nashville, Tennessee

Charles E. Saunders, M.D.
University of California at San Francisco
San Francisco, California

Donald S. Soloniuk, M.D.
Buffalo General Hospital
Buffalo, New York

Ronald D. Stewart, M.D.
University of Toronto
Toronto, Ontario

Philip A. Villanueva, M.D.
University of Miami
Miami, Florida

REFERENCES

1. Cales RH, Trunkey DD: Preventable deaths: A review of trauma care systems development. *JAMA* 254:1059, 1985
2. Mendeloff JM, Cayten GC: Trauma systems and public policy. *Ann Rev Publ Health* 12:402, 1991
3. Salmi LR, Williams JI, Guibert R, et al: Preventable deaths and the evaluation of trauma programs: Flawed concepts and methods. *30th Proceedings of American Association of Automotive Medicine*, 1986, pp 179–195
4. Wesson DE, Williams JI, Salmi R, et al: Evaluating a pediatric trauma program: Effectiveness versus preventable death rate. *J Trauma* 28:1226, 1988
5. Rivara FP, Maier RV, Mueller BA, et al: Evaluation of potentially preventable deaths among pedestrian and bicyclist fatalities. *JAMA* 261:566, 1989
6. MacKenzie EJ, Steinwachs DM, Ramzy AI: Evaluating performance of statewide regionalized systems of trauma care. *J Trauma* 30:681, 1990
7. Fleiss JL: *Statistical Methods for Rates and Proportions*, ed 2. New York, John Wiley & Sons, 1981
8. Van Wagoner FH: Died in hospital: A three-year study of deaths following trauma. *J Trauma* 1:401, 1961
9. Fitts WJ, Lehr HB, Bitner RL, et al: An analysis of 950 fatal injuries. *Surgery* 56:663, 1964
10. Waller JA, Curran R, Noyes F: Traffic deaths: A preliminary study of urban and rural fatalities in California. *Calif Med* 101:272, 1964
11. Bruser DM: Emergency care of auto crash victims. *Proceedings of American Association of Automotive Medicine*, 1970, pp 232–239
12. Perine MW, Waller JA, Harris LS: Alcohol and highway safety: Behavioral and medical aspects. Washington, DC, US Department of Transport, 1971
13. Gertner HR, Baker SP, Rutherford RB, et al: Evaluation of the management of vehicular fatalities secondary to abdominal injury. *J Trauma* 12:425, 1972
14. Waters JM, Wells CH: The effects of a modern emergency medical care sytem in reducing automobile crash deaths. *J Trauma* 13:645, 1973
15. Trunkey DD, Lim RC: Analysis of 425 consecutive trauma fatalities: An autopsy study. *J Am Coll Emerg Phys* 3:368, 1974

16. Moylan JA, Detmer DE, Rose J, et al: Evaluation of the quality of hospital care for major trauma. *J Trauma* 16:517, 1976
17. Foley RW, Harris LS, Pilcher DB: Abdominal injuries in automobile accidents: Review of care of fatally injured patients. *J Trauma* 17:611, 1977
18. Detmer DE, Moylan JA, Rose J, et al: Regional categorization and quality of care in major trauma. *J Trauma* 17:592, 1977
19. West JG, Trunkey DD, Lim RC: Systems of trauma care—A study of two counties. *Arch Surg* 114:455, 1979
20. West JG, Yurko AA: Validation of autopsy method for evaluating trauma care. *Arch Surg* 117:1033, 1982
21. Baker CC, Oppenheimer L, Stephens B, et al: Epidemiology of trauma deaths. *Am J Surg* 140:144, 1980
22. West JG, Cales RH, Gazzaniga AB: Impact of regionalization: The Orange County experience. *Arch Surg* 118:740, 1983
23. Neuman TS, Bockman MA, Moody P, et al: An autopsy study of traumatic deaths: San Diego County, 1979. *Am J Surg* 144:722, 1982
24. Certo TF, Rogers FB, Pilcher DB: Review of care of fatally injured patients in a rural state: 5-year followup. *J Trauma* 23:559, 1983
25. McCoy C, Bell MJ: Preventable traumatic deaths in children. *J Pediatr Surg* 18:505, 1983
26. Lowe DK, Gately HL, Goss JR, et al: Patterns of death, complication, and error in the management of motor vehicle accident victims: Implications for a regional system of trauma care. *J Trauma* 23:503, 1983
27. Reines HD, Duffy KH: Survivability of victims of traumatic injuries in Eastern South Carolina. *Am Surg* 49:203, 1983
28. Ottosson A, Krantz P: Traffic fatalities in a system with decentralized trauma care: A study with special reference to potentially salvageable casualties. *JAMA* 251:2668, 1984
29. Ramenofsky ML, Luterman A, Quindlen E, et al: Maximum survival in pediatric trauma: The ideal system. *J Trauma* 24:818, 1984
30. Cales RH: Trauma mortality in Orange County: The effect of implementation of a regional trauma system. *Ann Emerg Med* 13:1, 1984
31. Spain DM, Fox RI, Marcus A: Evaluation of hospital care in one trauma care system. *Am J Publ Health* 74:1122, 1984
32. Krob MJ, Cram AE, Vargish T, et al: Rural trauma care: A study of trauma care in a rural emergency medical services region. *Ann Emerg Med* 13:891, 1984
33. Baker CC, Deglutis LC, DeSantis J, et al: Impact of a trauma service on trauma care in a university hospital. *Am J Surg* 149:453, 1985
34. Shackford SR, Hollingsworth-Fridlund P, Cooper GF, et al: The effect of regionalization upon the quality of trauma care as assessed by concurrent audit before and after institution of a trauma system: A preliminary report. *J Trauma* 26:812, 1986
35. Shackford SR, Hollingsworth-Fridlund P, McArdle M, et al: Assuring quality in a trauma system—The medical unit committee: Composition, cost and results. *J Trauma* 27, 866, 1987
36. Kreis D, Plascencia G, Augenstein D, et al: Preventable trauma deaths: Dade County, Florida. *J Trauma* 26:649, 1986
37. Campbell S, Watkins G, Kreis D: Preventable deaths in a self-designated trauma system. *Am Surg* 55:478, 1989
38. Guss DA, Meyer FT, Neuman TS, et al: The impact of a regionalized trauma system of trauma care in San Diego County. *Ann Emerg Med* 18:1141, 1989
39. Davis JW, Hoyt DB, McArdle MS, et al: The significance of critical care errors in causing preventable death in trauma patients in a trauma system. *J Trauma* 31:813, 1991
40. Webb GL, McSwain NE, Webb WR, et al: Emergency department deaths. *Am J Surg* 159:377, 1990
41. Cayten CG, Stahl WM, Agarwal N, et al: Analyses of preventable deaths by mechanism of injury among 13,500 trauma admissions. *Ann Surg* 214:510, 1991
42. Dubois RW, Brook RH: Preventable deaths: Who, how often and why? *Ann Intern Med* 109:582, 1988
43. Koran LM: The reliability of clinical methods, data and judgments. *N Engl J Med* 293:642, 1975
44. Champion HR, Copes WS, Buyer D, et al: Major trauma in geriatric patients. *Am J Publ Health* 7:1278, 1989
45. Finelli FC, Jonsson J, Champion HR, et al: A case control study for major trauma in geriatric patients. *J Trauma* 29:541, 1989
46. Oreskovich MR, Howard JD, Copass MK, et al: Geriatric trauma: Injury patterns and outcome. *J Trauma* 24:565, 1984
47. DeMaria EJ, Kenney PR, Merriam M, et al: Aggressive trauma care benefits the elderly. *J Trauma* 27:1200, 1987
48. Scalea TM, Simon HM, Duncan AO, et al: Geriatric blunt trauma: Improved survival with early invasive monitoring. *J Trauma* 30:129, 1990
49. Morris JA, MacKenzie EJ, Damiano AM, et al: Mortality in trauma patients: The interaction between host factors and severity. *J Trauma* 30:1476, 1990
50. Boyd CR, Tolson MA, Copes WS: Evaluating trauma care: The TRISS method. *J Trauma* 27:370, 1987
51. Stothert JC, Gbaanador GBM, Herndon DN: The role of autopsy in death resulting from trauma. *J Trauma* 30:1021, 1990
52. Champion HR, Copes WS, Sacco WJ, et al: The Major Trauma Outcome Study: Establishing national norms for trauma care. *J Trauma* 30:1356, 1990

DISCUSSION

Dr. Michael Rhodes (Allentown, Pennsylvania): Dr. MacKenzie and her colleagues have presented a thoughtful and important study. We now have scientific evidence to support our suspicion that surgeons cannot agree on much of anything. Certainly, surgery department chairmen throughout the country would concur with that conclusion.

I agree with the authors that the study represents one of the first attempts to systematically examine the reliability of preventable death judgments for trauma and the factors affecting levels of reliability. The manuscript presents an excellent review of published studies of preventable death highlighted by a table comparing preventable death study methodologies. The statistical analysis is thorough, but as the authors point out, the small sample size is limiting.

The authors note that the availability of autopsy results and prehospital reports increases the reliability in judging preventability and recommend legislation promoting mandatory autopsy for trauma deaths. However, agreement was higher for early deaths in younger, less severely injured patients and they found that the autopsy was most helpful in this group.

The authors suggest that objective criteria such as the TRISS probability of survival of >50%, which has been used by the San Diego group to define "frankly preventable deaths," may better objectify the review process. However, the data to compute this were not uniformly available to them in this study.

One of the most important aspects of the study is the demonstration that even genuine experts disagree on the preventability of deaths. This paper will undoubtedly be quoted in courtrooms to counter the pseudo-experts who frequently display reasonable medical certainty about preventability of death in very complex and confounding trauma cases.

I have three questions for the authors: (1) Your study, as well as multiple published autopsy studies, suggest that the autopsy becomes less useful the longer the patient is hospitalized. Might it not be more prudent in today's cost-conscious healthcare environment to promote mandatory autopsies for early trauma deaths rather than all trauma deaths? (2) Several recently published studies have demonstrated that 85% of trauma deaths with a TRISS probability of >0.50 are deemed nonpreventable when subject to peer review. Therefore, shouldn't we use objectification such as TRISS as a screening tool rather than as a definition of preventability? (3) We have been using the "modified consensus rule" for many years in our trauma mortality conferences to determine preventability, but I did not know what it was called until I had the pleasure of reading this manuscript. We, as the authors have suggested, are using a three-point system of preventability. Would you think that substituting the word "potentially" for "possibly" might be helpful, since the word, "potential," seems to fall somewhere between possible and probable by dictionary defi-

nition and since this was an area that was problematic in your study.

This paper is well written and will be useful to all of us who are involved in mortality and morbidity conferences, as well as other quality assurance efforts.

Dr. Lenworth Jacobs (Hartford, Connecticut): I enjoyed this paper and I would like to ask what is the precision of the definitions used. How important is experience in this evaluation, since the evaluators had at least five years of experience? Have you thought of giving any weighting to "experience"? What is the trade-off between experience versus using a template? This is becoming increasingly important, since the use of these data is changing. Initially these studies were used for quality assurance, but now they are being used in designation and de-designation of trauma centers. It is also becoming really quite important in legal issues. I think that some thought needs to be given to whether a template should be developed which doesn't require experience or, if experience is important, should it be quantified? Thank you.

Dr. Kimball I. Maull (Knoxville, Tennessee): Dr. MacKenzie, I noticed that you included emergency physicians in these teams. Recognizing that emergency medicine is a horizontal specialty and that their window of expertise in dealing with trauma patients is limited to the emergency department, have you looked at whether there was disparity in judgment between the emergency physicians and the surgeons and, if so, whether that played a role in some of the inconsistencies that you saw in the study?

Dr. John A. Morris, Jr. (Nashville, Tennessee): I would like to thank Dr. MacKenzie for sending me a copy of this manuscript. Anyone who is considering either writing a grant or doing quality assurance in this area needs to spend the requisite hour or two to get the wealth of information that the manuscript contains—it is well worth the effort.

Dr. Steven R. Shackford (Burlington, Vermont): I would agree with what Mike Rhodes just said about the concise and nice review of the literature. It is very well done. I rise to make two comments. The first is on the "bad news." I think, as Dr. MacKenzie pointed out, that the reliability certainly was not what we would have anticipated, but nevertheless, if you really examine these data, when surgeons are given the information they need (i.e., autopsy information, prehospital records) they don't do so badly. She points out in the discussion that had she given the surgeons an explicit format to make the judgments I think they would have done better and I think that reliability would have been improved. Inter-rater reliability is not really the thrust of this paper and I hope that everyone here doesn't focus on the reliability information, but rather on what Dr. MacKenzie has done for us. She has provided us with the essentials in terms of the elements required for a preventable death study. She has given us some very nice guidelines and I think that preventable death studies in the future will be made a lot better by the efforts of this group. The second point I would like to make concerns the quality of trauma care. Since all of us are now dealing with total quality management, we no longer focus on a "bad apples" approach to quality assessment. The point that Dr. MacKenzie made in her discussion was that surgeons didn't disagree on where care deviated from the standard, and that is important. When there is agreement about the deviation from a standard, one can focus on the process that brought about the adverse outcome and subsequently improve the process. This is more important than determining whether or not a death was preventable or not preventable. We must agree on where the ball is dropped and we have to go back and attempt to prevent the fumble in the future.

Dr. Ellen MacKenzie (Closing): Thank you very much for all your comments. Dr. Rhodes, your first question concerned the use of autopsies in looking at early versus late deaths. We looked at this to the extent that our small sample allowed us to look at it. We found that the availability of autopsy data really helped reliability in both instances. So I think that it would be important to have autopsy information in both. Your second comment had to do with using TRISS as a screening tool as opposed to using it as the sole criterion for identifying preventable deaths or potentially preventable deaths. I think I would agree. We still don't have precise enough case mix tools to objectively identify deaths that are potentially preventable. I think we need to use TRISS analysis as an adjunct to a preventable death review. And finally, your questions regarding the use of the word, "potentially," instead of "possibly." I think that is an excellent idea because that really captures the essence of this first group. It really is a group of *potentially* preventable deaths; there is considerable uncertainty with regard to how preventable these deaths might be. In response to Dr. Jacobs' question regarding the experience of the panel members, unfortunately all the panel members that we included did have at least five years of experience. So we weren't really able to look at that issue. Dr. Maull, your question was whether or not there was more disagreement because we included the emergency physicians in the reviews. We did look at this in a couple of different ways. First of all, we compared the initial judgments made by the individual physicians within each specialty. That is, we looked at agreement among the emergency physicians, among the trauma surgeons, and among the neurosurgeons. We found there weren't big discrepancies, that agreement levels were pretty much the same across the board. The other thing we did was to look at the agreement of each physician with his or her overall panel judgment to see if there was any one particular physician or whether emergency physicians versus trauma surgeons or neurosurgeons seemed to control the consensus-building process in the reviews. The data show that, in general, it didn't seem that any one individual was dominating the consensus review. Having sat in on 18 conference calls and listening to these discussions, I found that the multidisciplinary panel approach was extremely helpful and there were numerous cases in which an emergency physician, for instance, would identify problems that occurred either early in the resuscitation phase or in the prehospital phase that either the trauma surgeon or the neurosurgeon did not identify. There were other instances where the neurosurgeon and the trauma surgeon identified problems that the other two did not. I found that the multidisciplinary approach was very very helpful and I think that people who participated in this study would generally agree. Thank you very much.