**Review of Evidence Regarding Trauma System Effectiveness Resulting from Panel Studies**

Ellen J. MacKenzie, PhD

Johns Hopkins Center for Injury Research and Policy, and The Johns Hopkins School of Public Health, Baltimore, Maryland.

THE JOURNAL OF TRAUMA: INJURY, INFECTION, AND CRITICAL CARE 1999;47:S34-S41

---

**Objective:** The objectives this study are to elucidate the advantages and disadvantages of the panel method for evaluating the quality of trauma care and to review the individual and collective evidence in support of regionalized trauma care derived from these studies.

**Method:** A review of the published literature was conducted and identified 10 panel studies that compared appropriateness of care and/or the preventability of deaths occurring either across hospitals in a trauma system versus non-trauma system, in a defined region before and after implementation of a trauma system, between trauma center and non-trauma center hospitals within a defined region, or across hospitals of varying levels of care presystem.

**Results:** Panel studies vary widely in the approaches used to elicit judgments; low rates of inter-rater reliability have been reported. The strength of the evidence derived from panel studies in support of trauma system effectiveness must be reviewed in this context.

**Conclusion:** All panel studies are classified as providing weak, Class III evidence. Yet, collectively they provide some face validity in support of the hypothesis that treatment at a trauma versus non-trauma center may be associated with less inappropriate care and fewer preventable deaths among the seriously injured.

**Key Words:** panel studies; preventable death; trauma systems; trauma centers.

Panel studies have been widely used to evaluate outcomes of trauma care and the performance of trauma systems. Since the early 1960s with the publication of the study by Van Wagoner on potentially preventable deaths among noncombat military injuries, more than 50 studies have used the panel approach to examine quality of care and outcomes among trauma patients. The method has evolved substantially over time; early studies typically relied on subjective impressions (often by only one or two experts) based on implicit criteria, whereas more recent studies have involved multiple panels of experts making decisions based on explicit criteria by using standardized information abstracted on each patient. Collectively, these studies have been used to support the argument that a regionalized approach to trauma care is effective in saving lives. The purpose of this study is twofold. First, the advantages and disadvantages of the panel method for evaluating the quality of trauma care are discussed in light of the studies published in the past 25 years. Second, the
individual and collective evidence in support of trauma care regionalization derived from these studies is reviewed.

**THE PANEL APPROACH: A REVIEW AND CRITIQUE**

The panel studies that have been published in the trauma literature vary considerably in the definitions of preventability, case mix of the study population, the size and composition of the panel reviewing the cases, and the process and criteria used to judge preventability. The inter-rater reliability of preventable death judgments has been shown to vary as a function of these parameters.\(^3\)\(^-\)\(^8\) Overall reliability of preventable death judgments is low when adequate data are not available for review and guidelines for judging preventability not clearly defined.\(^3\)\(^,\)\(^4\)\(^,\)\(^10\) However, when attention is paid to the issues summarized below, significant gains in reliability are achieved.\(^3\)\(^,\)\(^5\)\(^,\)\(^7\)\(^,\)\(^8\) The recently published studies by Esposito et al.,\(^8\) Maio et al.,\(^7\) and McDermott et al.\(^5\) are examples of well-executed panel studies that have achieved acceptable rates of inter-rater reliability.

**Definition of Preventability**

A potentially preventable death has been defined as any death that might have been prevented if optimal care had been delivered. There has been a tendency toward the increasing use of a three (vs. two) category classification of preventability (not preventable, possibly preventable, and definitely or frankly preventable) to acknowledge the small number of cases (the frankly preventable deaths) in which an obvious and grievous error was made that most certainly influenced outcome; agreement as to the preventability of these deaths is typically high. Shackford et al. have developed explicit guidelines for judging preventability,\(^9\) and the guidelines are recommended for use to increase consistency in judgments across studies.

It is important to underscore that implicit in this definition of preventability are three criteria that need to be met before assigning a death as potentially preventable: the injury or sequelae of the injury must be survivable; the care delivered must be judged suboptimal; and identified errors in the delivery of care must be directly or indirectly implicated in the demise of the patient. This link between the appropriateness of care and outcome is particularly critical in judging the preventability of a death. Acknowledging this important link, recent studies have incorporated a two-stage process in judging preventability. First, the case is reviewed with regard to appropriateness of care. Then, a judgment is made regarding the extent to which an identified error significantly influenced the outcome of the patient (taking into account the severity of the injury, the age of the patient, and the presence of comorbidities). Studies have shown that, although expert clinicians can agree about the adequacy or appropriateness of care, they often disagree about the extent to which errors in management contribute to the outcome of the patient.\(^3\)\(^-\)\(^5\)\(^,\)\(^8\) For this reason, it is important to separate these two thought processes and interpret the data accordingly.

**Case Mix of the Study Population**

The preventable death rate (PDR) is defined as the number of potentially preventable deaths divided by all deaths within the defined limits of the study population. Thus, when comparing results across studies, it is critical that attention be paid to the composition or case mix of the population of deaths reviewed. Most panel studies have excluded prehospital deaths from consideration, because data on these deaths are often sparse and difficult to obtain. Doing so, however, limits the use of panel studies for an evaluation of the trauma system as distinct from an evaluation of hospital trauma care.
The majority of studies that include only deaths occurring in the hospital also vary with regard to specific inclusion and exclusion criteria. Some include all major trauma (defined with varying criteria), whereas others include only vehicular trauma. Several studies also exclude deaths involving injury to the central nervous system (CNS). These latter exclusions can significantly influence the magnitude of the overall preventable death because the PDR for CNS deaths is typically much lower than for non-CNS deaths.\textsuperscript{3}

**Size and Composition of the Review Panel**

As summarized in Table 1, studies vary considerably in the size and composition of the panels used to judge preventability. Some of the panels included trauma surgeons only, whereas others were multidisciplinary often including neurosurgeons, emergency physicians, and occasionally pathologists/medical examiners, anesthesiologists, and emergency medical technicians. Multidisciplinary panels are now accepted as critical to a thorough and informed decision regarding preventability. For a multidisciplinary panel to function effectively, however, there must be an opportunity for discussion of cases for which there is disagreement so that different perspectives can be articulated and an informed decision made regarding potential preventability. Multidisciplinary discussions have been shown to modify individual judgments and result in higher agreement across panels.\textsuperscript{3-5}

<table>
<thead>
<tr>
<th>TABLE 1. Evidence table illustrating attributes of reviewed panel studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://ipsapp003.lwwonline.com/content/getfile/2281/75/8/fulltext.htm">http://ipsapp003.lwwonline.com/content/getfile/2281/75/8/fulltext.htm</a></td>
</tr>
</tbody>
</table>

**Review Process**

Perhaps most critical to the reliability and validity of panel studies is the process used in making judgments and the quality of information on which those judgments are made. It is now generally recognized that preventable death judgments based on autopsy results alone are inadequate; the consensus is that judgments should be made on the basis of both the autopsy and clinical records. In general, reviews based on this more complete set of information will identify more deaths as preventable.\textsuperscript{3,10} Few of the studies that use the clinical method, however, discuss the extent to which actual information provided to the reviewers was adequate enough to judge preventability. In several studies, autopsies were not uniformly performed in all deaths. Even less clear is the extent to which prehospital reports were included. And, even when prehospital reports are available for review, the completeness of the information available from the report varies considerably. Studies have shown that the availability and completeness of both the autopsy and prehospital care report strongly influence the reliability (and outcome) of the preventable death judgment.\textsuperscript{3,11}

For practical reasons, many of the more recent panel studies have provided an abstract summary to the reviewers in lieu of a complete (often voluminous) set of records. These abstracts have varied in detail from a 1-page narrative to a detailed 14-page abstract of critical data. This approach, although cost-efficient has not been widely tested for reliability and accuracy. The few studies that have
compared the two methods differ in their conclusions.\textsuperscript{12-14} More work is needed to establish the validity of this approach.

Finally, panel studies vary regarding the actual process of review and the rules used to arrive at a final judgment. Most of the preventable studies published in the past 20 years have used either an independent review of the records with the final judgment taken as the majority opinion (the \textit{independent review}), or an independent review followed by panel discussion of all cases in which at least one expert judged the death as potentially preventable (the \textit{panel consensus review}). A \textit{modified independent review} allows for the discussion of those deaths where the independent review does not result in a majority opinion. Only a few studies have used a \textit{unanimous decision review}, in which there is an independent review of the records and an assignment of preventability is made to only those cases where all reviewers agree that the death was potentially preventable. The unanimous decision rule will typically provide an estimate of the lower bound of the PDR, whereas the panel consensus approach (by using a multidisciplinary panel) provides an estimate of the upper bound.\textsuperscript{3}

In summary, panel studies vary widely in the approaches used to elicit judgments about appropriateness of care and potential preventability of death. Different results are achieved, depending on the case mix of the study population, the composition of the panel, the information available for review, and the process used in arriving at a final judgment. These differences make a comparison across studies difficult, and in many cases, inappropriate. The strength of the evidence derived from panel studies in support of trauma system effectiveness must be reviewed in the context of these parameters.

**REVIEW OF THE EVIDENCE FROM PANEL STUDIES**

By using methods described by Mann et al.,\textsuperscript{15} a review of the published literature was conducted and resulted in the identification of 10 panel studies that compared the appropriateness of care and/or the preventability of deaths occurring either across hospitals in a system versus nonsystem, in a defined region before and after the implementation of a trauma system, between trauma center and non-trauma center hospitals within a defined region, or across hospitals of varying levels of care presystem. These 10 studies were all conducted in one of four regions of the country: the state of Wisconsin; Orange County, California; San Diego, California; and Florida. They are described below.

**Wisconsin Studies**

The first panel studies to address the relationship between level of hospital capability, appropriateness of care, and outcomes were published by Moylan et al.\textsuperscript{16} and Detmer et al.\textsuperscript{17} by using data from Wisconsin. These two studies compared the appropriateness of care received by patients treated at hospitals categorized according to criteria promulgated in the early 1970s by the American Medical Association (AMA).\textsuperscript{18} It should be noted that these criteria differ substantially from current guidelines of the Committee on Trauma of the American College of Surgeons (ACSCOT).\textsuperscript{19} Of particular note is that AMA criteria were developed to categorize hospitals according to their capabilities of managing all emergency patients and not for specific subgroups such as major trauma. The first study by Moylan et al. examined trauma care in South Central Wisconsin and compared treatment of major trauma patients across one Level II (university) hospital, one Level III (general teaching) hospital, two Level IV (general) hospitals, and one rural hospital that was not classifiable according to AMA criteria (unclassifiable).\textsuperscript{16} A total of 823 cases (both deaths
and survivors) were identified as "major" trauma (loosely defined as trauma involving head, chest, and abdominal injuries, plus extremity injury). A total of 237 cases (57 deaths and 180 survivors) were selected for review by a panel of general surgeons for appropriateness of care. The percentage of cases with errors in treatment/diagnosis increases from 7% in category II, to 12% in category III, to 29% in category IV, and 57% in the rural (unclassifiable) hospital. Of those judged to have received inappropriate care, 18 died and 20 survived to hospital discharge. Although an overall PDR of 21% was reported, the distribution of preventable deaths by level of hospital care was not presented.

Although the study by Moylan et al. was limited because of small sample sizes, a questionable strategy for selecting cases for review and the lack of detail regarding the process of review and decision rules used in classifying cases, it did lay the foundation for a larger, more carefully designed study of major trauma treated at a stratified random sample of 28 hospitals located throughout the state of Wisconsin. This study examined care delivered at 8 area or AMA category II/III hospitals, 10 community or Level IV hospitals, and 10 rural or unclassifiable hospitals. A stratified random sample of cases across hospitals was selected and 2,500 medical records were abstracted. The complete charts of all deaths and one third of survivors (a total of 556 cases) were subsequently photocopied and subjected to a review by five panels each with four to five physician members. Of the 556 cases reviewed, 167 (30%) were judged to have received unacceptable care; the percentage of unacceptable care increased with decreasing level of hospital emergency capability (19% at Level II/III; 35% at Level IV; and 55% at unclassified hospitals) \( p < 0.05 \). There was no correlation between level of hospital capability and mortality rate; this finding may have been explained by the small number of severely injured patients treated in the unclassified hospitals. In 36% of the 151 deaths reviewed, care was judged unacceptable and was felt to have contributed to the adverse outcome. As was true for the previous Wisconsin-based study, however, no attempt was made to correlate PDR by level of hospital care. Among three types of specific management problems identified by the reviewers, failure to act (e.g., delay in surgery) was the most common. Sixty-nine percent of the unacceptable cases were so characterized, compared with 41% of cases characterized as involving inappropriate acts and 33% inappropriate diagnosis (Table 2). It is interesting to note that, although the percentage of all unacceptable cases characterized as involving a failure to act was similar across the levels of hospitals, both the percentage characterized as involving inappropriate action and inappropriate diagnoses were significantly lower in the Level II/III versus Level IV or unclassified hospitals. Although not strictly a preventable death study comparing care received at trauma centers versus non-trauma centers (as defined by using current ACSCOT criteria), this study provides some early evidence to suggest that unacceptable care may be more prevalent when delivered in smaller hospitals with minimal capability for treating major trauma patients. The extent to which unacceptable care impacted outcome, however, was less clear.

**Orange County Studies**

In the late 1970s, West and colleagues took advantage of a natural experiment to examine whether regionalization of trauma care lowered the proportion of potentially preventable trauma deaths. They first compared the PDR in San Francisco County (where all trauma victims were brought to a single, centrally located trauma center) with that of Orange County (where patients were transported to the
nearest of 31 different hospitals). The results are summarized in Table 3. Across all trauma, the PDR was reported to be 1% in San Francisco compared with 43% in Orange County; 73% of the 30 non-CNS deaths in Orange County were judged potentially preventable compared with only 6% of the 16 non-CNS deaths in San Francisco.

This study, although representing the first systematic study of trauma care in two communities (with and without an organized system of trauma care), was limited in several ways: the number of deaths reviewed was small (especially of non-CNS trauma); judgments regarding potential preventability of deaths in Orange County were based on autopsy records only, whereas both clinical and autopsy records were available in reviewing deaths in San Francisco; injury severity was not carefully derived (the Injury Severity Score was based on autopsy data for Orange County deaths and no physiologic scores were available for comparison); and the review was not blinded so that full knowledge of where the case had been treated was known to the reviewers when making judgments. In addition, the two communities being compared were vastly different in geography and available resources. At the time of the study, San Francisco covered only 49 square miles with only one hospital, whereas Orange County covered 782 square miles and supported a total of 39 hospitals (31 of which had emergency departments). These vast differences make the interpretation and generalizability of the study results difficult.

In 1980, Orange County implemented a regionalized system of trauma care and used ACSCOT criteria to designate trauma centers at five hospitals in the county (one Level I center and four Level II centers). Implementation of the system provided the opportunity for a before-after comparison of quality of care and outcomes. Two such comparisons were made and reported in the literature. The first was published by West et al who reviewed a total of 29 non-CNS deaths occurring in Orange County and concluded that only 6 of the 29 deaths (21%) were potentially preventable. Although this PDR represented a substantial reduction in PDR when compared with PDRs of 73% (based on a review of 30 deaths occurring in 1974) and 71% (based on a review of 21 deaths occurring in 1978-1979), it is still relatively high. This particular study was based on very small numbers of non-CNS deaths; however, and the reviews were non-blinded and relied on autopsy records only.

In a second before-after comparison of the PDR in Orange County, Cales reviewed 58 deaths occurring in 1977 to 1978 (preimplementation) with 60 deaths occurring in 1980 to 1981 (postimplementation). The PDR for non-CNS trauma decreased from 86 to 40% (p < 0.05), whereas the PDR for CNS trauma was low overall with little change before and after implementation (5% vs. 3%). For all cases combined, the PDR decreased significantly from 34 to 15% (Fig. 1; Table 4). Of particular note is the observation that seven of the nine preventable deaths occurring after system implementation occurred in non-trauma center hospitals; the PDR among trauma center cases was only 4% compared with 54% among non-trauma center cases. The reviewers also recorded fewer cases of inappropriate care among deaths occurring after versus before implementation of the system (Table 5). Specifically, the percentage of cases with appropriate use of intravenous fluids increased from 38 to 90%; the percentage of cases with appropriate surgical intervention increased from 6 to 76%. The overall conclusion of the study was that although an overall improvement in PDR was
observed after the system was implemented, the number of potentially preventable deaths occurring outside the trauma centers continued unabated. This study used similar methodology in judging the preventability of deaths before and after implementation, and the review was based on all available information from autopsy reports, prehospital records, and hospital charts. As with the previous study, however, the review was not blinded and a total of six different panels were used with no discussion of inter-rater reliability or the potential for biased judgments.

San Diego Studies

Perhaps the most rigorously designed preventable death studies are those published by Shackford and colleagues that report on the experience of San Diego before and after implementation of a regionalized trauma system. In the first study, the percentage of potentially preventable deaths (among both CNS and non-CNS related deaths) dropped from 22% (19 of 88 deaths occurring in 1982) to 10% (11 of 112 deaths occurring in 1984) (p < 0.05); only 12 of the deaths before system implementation were frankly preventable compared with only 3 after implementation (Table 6; Fig. 1). A major advantage of this study over previous preventable death studies is the careful and independent assessment of errors made in the management of the cases subjected to review. Table 7 summarizes the results and shows a dramatic decrease in the percentage of errors made before and after implementation of the system.

Although the preliminary study by Shackford provides some of the strongest evidence in support of the impact of regionalized care for major trauma patients, it is not without its limitations. The methods used to review cases before and after implementation were not identical and, as in previous...
studies, the reviewers were not blinded, which may introduce a serious bias in the results. Perhaps most problematic, however, is that the presystem review is based on a sample of deaths occurring at all 30 hospitals in the area, whereas the postsystem review only included deaths occurring at the 5 designated trauma centers.

In a follow-up study, Shackford and his colleagues reviewed deaths occurring at both the non-trauma center and trauma center hospitals during a 22-month period (1984-1986) and found that the PDR in non-trauma centers (15%) was indeed significantly higher than in trauma centers (10% [Table 6]). It is interesting to note that this difference in the overall PDR was accounted for by a difference in the percent of frankly preventable deaths (8% and 2% in non-trauma centers and trauma centers, respectively) rather than in those that were possibly preventable (7% and 8%, respectively). The most common error implicated in preventable deaths occurring at non-trauma centers was an error in diagnosis, whereas preventable deaths occurring at trauma centers were most likely attributable to errors in technique. The overall PDR for this time period (among deaths occurring at both non-trauma and trauma center hospitals) was 11%, however, which is still significantly lower than the presystem PDR of 22%. A major contribution of these studies by Shackford and his colleagues was the publication of guidelines for judging the preventability of trauma deaths.

Mention should also be made of an additional study conducted by Guss and Neuman that used data from San Diego. In two separate publications, they report PDRs for deaths occurring in San Diego before implementation of the system (1979) and after implementation (1986). Of 211 deaths reviewed from 1986, 2 deaths (1%) were judged as preventable compared with 20 of 177 deaths (11%) occurring in 1979 (p < 0.01). Similar approaches were used in both reviews; the same six reviewers were used, and they were blinded to the treating physician and the hospital. However, only autopsy records were available, which most likely explains the low PDRs both before and after implementation of the system. Given that only autopsy records were available for review, a systematic assessment of the quality of care was not possible. It is also important to note that this study excluded not only prehospital deaths but also deaths occurring within 1 hour of arrival at the hospital (30% of all deaths).

Florida Studies

A final set of studies designed to examine the impact of trauma system and trauma center care on quality of care and preventable deaths are based on data from Florida. The first study was published in 1986 by Kreis et al. and reviewed 246 non-CNS trauma deaths occurring in Dade County (at a time when a trauma system was being organized but not yet implemented). Of these 246 cases, 91 were transported to the only hospital that had the capabilities of an ACSCOT verified Level I trauma center; the remaining 155 cases were treated across 6 hospitals that were planned Level II hospitals and 16 other hospitals that would not meet the criteria of either a Level I or II center. A panel of four trauma surgeons judged 21% of all deaths as potentially preventable with the PDR for the functional Level I trauma center (12%) significantly lower than the other 22 hospitals (21% in the planned Level II and 30% in the noncategorized hospitals [p < 0.01; Fig. 2]). Although this study is only a comparison across different levels of "presumed" hospital capability before actual system implementation, it was well designed with reasonable sample sizes.

![FIG 2. Percentage of deaths judged potentially preventable (±2 SD) by presystem hospital characteristics for Dade County, Florida. Adapted from Kreis et al.26](http://ipsapp003.lwwonline.com/content/getfile/2281/75/8/fulltext.htm)
Campbell and colleagues conducted a similar presystem study in Hillsborough County, Florida. A three-member physician panel reviewed the autopsies of 62 non-CNS deaths occurring in one of six hospitals in the county and found 23% of the deaths potentially preventable. Although they found a lower percentage of potentially preventable deaths in the area’s self-designated trauma center hospitals (20%) versus outlying or non-trauma center hospitals (36%), the difference was not statistically significant because of the small number of deaths in the outlying hospitals (n = 11). This difference is difficult to interpret, however, because it is unclear what the capabilities of the self-designated hospitals were. In addition, Injury Severity Scores were markedly different in the self-designated versus outlying hospitals, making direct comparisons difficult.

In late 1988, Hillsborough County organized a County Trauma Agency to coordinate trauma care among prehospital providers and state-designated trauma centers (one Level I and two Level II). Thoburn and colleagues subsequently reviewed all 232 non-CNS trauma deaths transported to the hospital during an 18-month period (October of 1989 to April of 1991) by using a well-structured panel approach and concluded that 10 of these deaths (4%) were potentially preventable. They further argued that if 143 patients who arrived at the emergency department undergoing cardiopulmonary resuscitation were removed from the denominator, the PDR would increase to 11%. In either case, the PDR was significantly lower than the rate of 23% previously reported by Campbell et al. Although these two studies are not directly comparable because of very different review methodologies, the relatively large differences suggest that implementation of the system may have had an impact.

DISCUSSION

Of the 10 studies reviewed here, only two provide valid comparisons of the PDR before and after implementation of a trauma system. Although not without their limitations, both studies provide evidence in support of a significant decline in PDR postimplementation. It is interesting to note that although baseline estimates differed between Orange County and San Diego, both studies showed a 50% reduction in the PDR. Perhaps even more compelling than differences in PDRs, however, are the consistent patterns of inappropriate care noted across all studies. The study by Shackford and colleagues, for instance, documented a decrease in delays to disposition from 54 to 7% before and after implementation of the system in San Diego and a decrease in the percentage of cases with suboptimal care from 32 to 3%. Although assessments regarding appropriateness of care were not blinded, they were made independent of the judgment regarding the preventability of death and are generally associated with high inter-rater agreement.

As indicated by Mann and colleagues in this issue, all of the panel studies in support of trauma system effectiveness are classified as providing weak, Class III evidence. Yet, collectively they provide some face validity in support of the hypothesis that care delivered at trauma versus non-trauma center hospitals may be associated with less inappropriate care and fewer preventable deaths.
among the very seriously injured. It is important to remember, however, that because prehospital deaths are routinely excluded from preventable death reviews, these studies do not adequately address the broader issue of "system" versus "trauma center" effectiveness. Perhaps more importantly, these reviews provide little, if any, information about the effectiveness of trauma center or trauma system care for the much larger number of trauma patients who survive their injury. Although a couple of the studies\(^9,16,17\) provide estimates of inappropriate care for both deaths and survivors, no assessment is made regarding the impact of inappropriate care on nonfatal outcomes such as functional status or quality of life. Clearly, an important next step in garnering evidence in support of regionalized trauma care is the development of estimates of cost-effectiveness where effectiveness is defined in terms of functional outcomes and quality of life.

Nonetheless, when properly designed, panel studies can generate important information regarding compliance with generally accepted principles of trauma care and often provide the foundation on which to effect change. This is especially true when key stakeholders are integrally involved in the review process. Thus, although panel reviews may not provide the class I or class II evidence one would like in support of trauma systems, they should not be abandoned as a valid approach to system evaluation and continuous quality improvement. They remain an effective tool for educating providers and pointing to deficiencies in the system that could be addressed to increase both the effectiveness and efficiency of trauma care.

REFERENCES


This review was supported in part by grant R49/CCR302486 from the National Center for Injury Prevention and Control of the Centers for Disease Control and Prevention.

Address for reprints: Ellen J. MacKenzie, PhD, The Johns Hopkins School of Hygiene and Public Health, 615 North Wolfe Street, Room W1025, Baltimore, MD 21205; email: emackenz@jhsph.jhu.edu.