To the Editor:

I read with great interest the article of McDonough et al. that has been recently published in The Journal of Trauma. I congratulate the authors for their excellent study, which raises a few points worthy of discussion.

Ethanol in high doses may cause hypotension through direct effect on peripheral blood vessels that may progress to a shock status. The present study addresses the issue of combining hemorrhagic shock and ethanol intoxication at the levels similar to those found in trauma patients. It is important to notice that many shock mediators are lipid soluble. A clear example for that is platelet-activating factor (PAF). PAF activity stays for weeks in ethanol compared with less than a day in distilled-water. The free active form of PAF is actually ethanol-extractable. The effects of PAF on cardiopulmonary system are much more potent when dissolved in alcohol. The increased potency of the lipid mediators when dissolved in alcohol may contribute to the shock status as shown by McDonough et al.

The interest of McDonough et al. was focused on the hemodynamic and metabolic variables. The mortality was different between the two groups but not significant (5/12 compared with 6/26; p = 0.27 Fisher’s exact test). Their study was not randomized. It is possible that the authors started with the same number of animals in the two groups, but because many animals of the experiment group died, they had to increase the number in the experiment group so that at after 24 hours they can get reasonable numbers to analyze. Combining ethanol and bleeding had a high mortality (20/26, 76%). In reality, the animals with a better hemodynamic status in the experiment group survived; this lead the authors to conclude that blood pressure recovery, although incomplete, was similar in the two groups. This conclusion is biased. If the authors had used more animals in the control group, they would have ended with same problem (unbalanced data). I think the results would have been more reliable if the authors had a shorter follow up between two balanced groups with more repeated measures. It is difficult from the graphs to know the number of the animals that survived in each group at each time point between 30 minutes and 24 hours so as to define the best follow up time for this model if used in the future by other researchers.

Fikri M. Abu-Zidan, MD, FRCS, PhD, DipApplStat
Associate Professor, Head Trauma Group,
Department of Surgery,
Faculty of Medicine and Health Sciences,
Al Ain, United Arab Emirates.
E-mail: fabuzidan@uaeu.ac.ae

REFERENCES


To the Editor:
The Utility of Physical Examination to Detect Hemopneumothorax in Patients with Blunt Chest Trauma

I read, with interest, the study by Bokhari et al. that evaluates the sensitivity of physical examination in chest trauma. The aim of the study was to investigate which signs and symptoms of chest trauma are sensitive for hemopneumothorax (HPX). The conclusion is that blunt chest trauma patients who are hemodynamically stable with a normal physical examination do not require chest radiograph. This conclusion was drawn from sensitivity, specificity, and predictive values of physical examination of a population of blunt chest trauma patients in which only 7 out of 523 patients (1.3%) were true positives with HPX. I doubt the validity of drawing such a conclusion from this study.

In this study, the inclusion criteria for blunt chest trauma was defined as “patients with suspected blunt trauma to the chest.” It is not unlikely that a significant proportion of the patients included might have only minor and negligible trauma. According to the methodology described and physical signs tested, patients who had no “pain or tenderness” were included. I cannot envisage a patient to be classified as a victim of blunt chest trauma if he or she is without pain or tenderness—unless the “trauma” is really very mild.

For interpreting the results, it would be immensely helpful if the number of patients included and found to have no pain or tenderness was also supplied. Similar data describing the other two physical features evaluated (auscultation and tachypnea) will also be useful. Are there further data (e.g., Injury Severity Scores, multiple vs. single injury) describing the population studied? The authors stated that the patients collected were not consecutive cases, implying that selection bias was possible, and this limitation should be acknowledged.

Furthermore, the sensitivity of “pain or tenderness” as a sign is only 57.1 percent. It implies that a fair proportion of patients was without pain and/or tenderness and, yet, found to have HPX (false negative). This is surprising to me on correlating with experience in clinical practice, therefore I
attempts to calculate the exact number of false negatives. Since Sensitivity = TP/(TP + FN), where TP denotes True Positive and FN denotes False Negative, and we are given that sensitivity is 57.1 percent and TP is 7, it worked out that FN is 5.2. I cannot understand how and why FN is not a whole number.

Finally, the negative predictive value of “pain or tenderness” is high (over 99%). This just means that the number of true negatives far exceeded the number of false negatives. A large number of true negatives may just mean that many of the patients included were without pain or tenderness and were negative for HPX, that is, most likely, cases of minimal trauma. Again, more data from the authors will help put the study population in context. Making prediction rules on the basis of seven patients is misguided.

Stewart S. Chan, MBBS (Syd), FRCSEd, FHKAM
( Emergency Medicine)
Adjunct Assistant Professor, Accident & Emergency
Department,
Prince of Wales Hospital,
The Chinese University of Hong Kong
Hong Kong
E-mail: sauakau@netvigator.com

REFERENCE

The Author’s Reply:
We would like to thank Dr. Chan for his helpful comments about our paper. The paper evaluates the sensitivity of physical examination for detecting hemo-pneumothorax (HPX) in penetrating and blunt chest trauma. Dr. Chan implies that “pain or tenderness” is a prerequisite for trauma that causes HPX. The evidence he cites is experience in clinical practice. It is our contention that while this argument appeals to common sense, it is not supported by scientific evidence. Indeed, the sensitivity of “pain or tenderness” for HPX was only 57.1 percent. There were 4 true positives (TP) and 3 false negatives (FN). Sensitivity = TP/(TP + FN) = 4/7 = 57.1 percent. Dr. Chan erroneously used 7 for TP, which is the number of total number of HPXs and not the subset of HPXs associated with “pain or tenderness” (TP = 4). In the blunt chest trauma category, approximately 20 percent of the patients had “pain or tenderness.”

The objective of the paper was to evaluate the need for chest radiography in hemodynamically stable patients with chest trauma. Clearly, the high negative predictive value (NPV) of “pain or tenderness” results from a lot more true negatives compared with false negatives (NPV = TN/TN + FP). Dr. Chan states that the large number of true negatives could be due to minor trauma in the majority of cases. However, to translate that description of the mathematical definition of NPV into an argument supporting minimal trauma is erroneous. As mentioned above, “pain or tenderness” is not a prerequisite for HPX. Scientific evidence for the actual force required to create a HPX is unknown. Furthermore, the explanation of HPX in blunt chest trauma is not limited to rib fracture and subsequent puncture of the lung, as mentioned in the manuscript.

A large force may produce chest wall pain and not cause HPX depending on where and how on the chest it is applied. An example would be force concentrated on a small area of a bony chest structure. This may not translate into increased intrathoracic pressure or fracture of the bone and not cause HPX. On the other hand, a smaller force applied to a larger area of the chest with a closed glottis may cause a HPX without causing significant pain or tenderness.

The study evaluated the utility of physical examination in a typical urban Level I trauma center in the United States. It is unknown if the principles elucidated by this study would hold up in injuries caused by radically different mechanisms, which produce a very different prevalence of the disease.

Lastly, Dr. Chan points out correctly that since the patients used in the study were not consecutive, selection bias is possible. The patients were not consecutive, yet there was no selection bias. The patients were not consecutive because the recorder was occasionally not present on some days and some weekends. While this might induce selection bias, the schedule of the recorder was not made with the study in mind. In addition, when the recorder was present, all patients were recorded. Thus, we do not believe that there was any systematic selection bias introduced.

Faran Bokhari, MD, FACS, Scott C. Brakenridge, BS, Kimberly K. Nagy, MD, FACS
Department of Trauma, Cook County Hospital, Chicago Illinois Rush-Presbyterian-St. Luke’s Medical Center Chicago, Illinois

REFERENCE

To the Editor: Statistical Analyses of Trauma Outcome: When is More too Much?

The methodological aspects of scoring algorithms that predict mortality have rightly gained much attention in the field of traumatology. Meredith et al. recently compared injury scales using advanced statistical procedures in a large database. Professor Champion suggests in his Editorial Comment that the statistical methodology from this study may serve as a tutorial to others. We doubt, however, whether all analyses are recommendable.

First, we support the concept of deriving calibration equations from the dataset by logistic regression. The logistic model implies that predicted probabilities can be calculated from each of the scores by an intercept b0 and a slope b1. However, since only two parameters are esti-
ated (b0 and b1), overfitting or overoptimism are not of concern in this procedure.\textsuperscript{2,3} Therefore, all performance measures could have been estimated on the full database without any danger, instead of the more complex process of taking the mean over 10 cross-validations.\textsuperscript{1} The only advantage of the cross-validations seems to be the ability to do head-to-head comparisons. For this purpose, bootstrap resampling would, however, be preferable, since this may provide more stable comparisons, e.g., by calculating 100 or more comparisons rather than 10.\textsuperscript{4,5}

Second, the authors studied the variance of the predicted probabilities, referring to studies on logistic calibration.\textsuperscript{6,7} The variance is conceptually related to discrimination: the more spread in predictions, the better survivors are separated from deaths.\textsuperscript{3} The refinement parameter b1 is, however, primarily, a calibration measure that needs to be close to one.\textsuperscript{3,7} The authors seem not to have realized that they already calibrated the models on their data, such that b1 = 1 by definition, however they found values between 6 and 10 are indicated. What went wrong? The authors state (page 625) that the mortality outcome was “modeled as a function of the predicted probability of death.”\textsuperscript{1} Loosely speaking, this is correct, but the technical calculation is that the mortality should be related to the log of the odds (or “logit”) of the predicted probability of death: y = b0 + b1 logit (predicted probability).\textsuperscript{3,6,7} If we do not apply the logit transformation, we lose any interpretation of the refinement parameter b1. For illustration, we calculated predicted probabilities with the calibrated ISS and NISS in a database of 1102 trauma patients seen in Rotterdam, the Netherlands. If we relate mortality to the predicted probability, b1 was 6.0 for the ISS and 6.8 for the NISS. This suggests a worse performance of the NISS (value further from 1). However, the ROC areas were 0.90 and 0.94, and Hosmer-Lemeshow (H-L) statistics 48 and 5 for ISS and NISS, respectively, clearly indicate a better performance of the NISS over the ISS in this relatively small database.

We like to challenge the authors to calculate the refinement parameter in the correct way, i.e., with the logit transformation. We predict that b1 will be 1 for all (calibrated) injury scales. Therefore, the authors’ Tables 9 and 10, in addition to the text on pages 626 and 627, should be discarded.\textsuperscript{4} This leaves the mean results on discrimination (authors’ Table 5, ROC areas) and calibration (authors’ Table 7, H-L statistics) as the central message from this important database. This message does not require unnecessary statistical analyses such as cross-validation and refinement parameter calculations.

* Ewout W. Steyerberg, † Sander P.G. Frankema, ‡ Frank E. Harrell, Jr.
† Department of Public Health, Erasmus MC, Rotterdam, the Netherlands;
‡ Department of General Surgery and Traumatology, Erasmus MC, Rotterdam, the Netherlands;
‡ Department of Health Evaluation Sciences, University of Virginia, Charlottesville, VA, U.S.A.

**REFERENCES**


**The Author’s Reply:**

We would like to thank Drs. Steyerberg, Frankema, and Harrell for their erudite discussion and review of the statistical methods used in our recent paper. Their main points of contention are that two unnecessary statistical procedures are applied to the data and that one of them is applied in error.

The editorialists note that the performance measures could safely have been estimated on the full database without spending the effort to calculate the measures in ten different pieces via a ten-fold cross-validation design. The decision to use the cross-validation design was made for several reasons. Initially, formulating an NTDB-derived ICISS and new NTDB-derived weights for an APS score were part of the study design. Without a cross-validation design, the estimates of the SRRs (for ICISS) and the logistically calibrated weights (for APS) would have been applied to the same patient population from which they were derived, an estimation faux pas. For example, in the case of ICISS, patients whose data contributed to the SRRs that ICISS is dependent upon would then be used to validate ICISS, giving it an inherent “home court advantage.” In the end, we decided not to report on the NTDB versions ICISS and APS, and we instead focused on existing scores and saved the new scores for future research. We believe that use of the cross-validation design serves to facilitate a basic understanding of the variability and distribution of the performance measures. The design also samples each observation a fixed number of times, rather than relying upon the more computationally intensive bootstrap to randomly sample observations (which essentially is a different vehicle to get to the same place). Rather than summarizing the results of 100s or 1000s of samples, we were able to paint the picture in a series of easily comprehensible tables of head-to-head matches. However, we agree with the authors that this level of detail need not be applied to every study, though it is advantageous.

The second criticism of the paper surrounds our interpretation of the b1
parameter obtained when fitting predicted probabilities from each model against survival. Drs. Steyerberg, Frankema, and Harrell suggest an alternative approach that fits the logit of the predicted probabilities versus mortality and predict that the b1 coefficients obtained in this manner will all be very close to 1. It is correct that if we had used the logit of the predicted probability, the resulting b1 coefficients would have been close to 1 for all models. However, our use of the predicted probabilities was predicated on an interest in the variance of the predicted probabilities obtained from each model. We believe the results presented in the manuscript are informative in this regard. In retrospect, we may have confused the issue by equating this approach to refinement. An unreported box and stem plot also confirmed our interpretation of the analysis, which is that logistically modeling mortality against the predicted probabilities gives us a measure of overall spread.

Again, we would like to thank Drs. Steyerberg, Frankema, and Harrell for their letter. They raise important issues that need to be considered when doing research of this sort. However, their findings do not change the results of our study. Their main point—that a study need not have this level of statistical detail to be successful—is well taken.

J. Wayne Meredith, MD,
Gregory W. Evans, MS,
Patrick D. Kilgo, MS
Department of General Surgery,
Wake Forest University School of Medicine,
Winston-Salem, North Carolina