Validating a vignette-based instrument to study physician decision making in trauma triage

Deepika Mohan, MD MPH[*,§], **Baruch Fischhoff, PhD**[¶], **Coreen Farris, PhD**[‡‡], **Galen E. Switzer, PhD**[†,**,§§,‡], **Matthew R. Rosengart, MD MPH**[*,§], **Donald M. Yealy, MD**[‖], **Melissa Saul, MS**[††], **Derek C. Angus, MD MPH FRCP**[*], and **Amber E. Barnato, MD MPH MS**[**]

[*]The CRISMA (Clinical Research, Investigation, and Systems Modeling of Acute Illness) Center, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[§]Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[¶]Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, USA

[†]VA Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, University of Pittsburgh, Pittsburgh, PA USA

[**]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[‡]Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, USA

[‡‡]RAND Corporation, Pittsburgh, PA, USA

[‖]Department of Emergency Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[§§]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[††]Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, PA USA

Abstract

Background—The evidence supporting the use of vignettes to study physician decision making comes primarily from the study of low-risk decisions and the demonstration of good agreement at the group level between vignettes and actual practice. The validity of using vignettes to predict decision making in more complex, high-risk contexts and at the individual level remains unknown.

Methods—We had previously developed a vignette-based instrument to study physician decision making in trauma triage. Here, we measured the re-test reliability, internal consistency, known-groups performance, and criterion validity of the instrument. Thirty-two emergency physicians, recruited at a national academic meeting, participated in reliability testing. Twenty-eight trauma surgeons, recruited using personal contacts, participated in known-groups testing. Twenty-eight emergency physicians, recruited from physicians working at hospitals for which we had access to medical records, participated in criterion validity testing. We measured rates of under-triage (the proportion of severely injured patients not transferred to trauma centers) and over-triage (the proportion of patients transferred with minor injuries) on the instrument. For physicians

Address for correspondence: Deepika Mohan, Room 637 Scaife Hall, The CRISMA Laboratory, Critical Care Medicine, University of Pittsburgh, 3550 Terrace Street, Pittsburgh, PA 15261, Tel: (412) 647 6783, Fax: (412) 647 8060, mohand@upmc.edu.
The work was performed at the University of Pittsburgh

participating in criterion validity testing, we compared rates of triage on the instrument with rates in practice, based on chart review.

Results—Physicians made similar transfer decisions for cases ($\kappa = 0.42$, p<0.01) on two administrations of the instrument. Responses were internally consistent (Kuder-Richardson 0.71–0.91). Surgeons had lower rates of under-triage than emergency physicians (13% v. 70%, p<0.01). No correlation existed between individual rates of under- or over-triage on the vignettes and in practice ($r = -0.17$, $p = 0.4$; $r = -0.03$, $p = 0.85$).

Conclusions—The instrument developed to assess trauma triage decision making performed reliably and detected known group differences. However it did not predict individual physician performance.

INTRODUCTION

Over 1000 papers in the last decade have used case vignettes to study physician decision making. Researchers have used vignettes to explore variation in care,[1] analyze cognitive biases,[2] and assess the effectiveness of quality improvement interventions.[3] Vignettes offer significant advantages in studying decision making, including low cost and ease of use. Moreover, they may allow quantification of physician performance in ways that avoid the limitations of other quality metrics, such as case-adjusted outcomes.[4, 5]

The best evidence supporting the validity of vignettes comes from studies of physician behavior in outpatient family practice clinics, in eye clinics, and in hospitals when prescribing antibiotics – all assessed common, low-risk conditions.[6–8] In these contexts, physicians' responses to case vignettes correlate with their decisions regarding actual clinical encounters at the *group*-level. Such correlations prompted many investigators to use case vignettes as a convenient proxy for real-world decision making, offering an efficient strategy to study factors influencing physician performance.[4, 9] However, it is unclear if the associations observed between vignette-based and real-world performance can be generalized to medical decisions in more complex, uncertain, and time-pressured environments such as trauma triage or treatment of hemorrhagic shock. External factors, such as time pressure and cognitive load, may play a more significant role in these encounters than in the low-acuity, high base rate conditions studied previously, where physicians may be better able to focus on details like those described in vignettes. Moreover, even in these settings, the evidence involves group-level correlations, with little evidence that vignettes predict the decision making of individual physicians.

Trauma triage is an archetype of medical decisions that must be made under conditions of uncertainty, time pressure, and affective and cognitive load. When a trauma patient arrives at a regional emergency department, the physician must evaluate the patient's condition, stabilize the patient, and then quickly transfer those with serious injuries to a Level I trauma center. Well-established clinical practice guidelines for trauma triage provide clear rules for transfer, which have been taught to nearly all emergency department (ED) physicians (e.g., multiple rib fractures should prompt immediate transfer to a Level I trauma center).[10] Nonetheless, rates of under-triage (the proportion of patients who are not transferred to a trauma center despite having moderate to severe injuries) and over-triage (the proportion of patients who are transferred to a trauma center, despite having only minor injuries) remain alarmingly high (70% and 50% respectively),[12] and contribute to significant patient morbidity and mortality.[13] Four decades of quality improvement interventions by federal, state, and professional organizations have failed to shift these rates significantly.[14, 15] Efforts to understand non-compliance with clinical practice guidelines have concentrated on patient-, hospital-, and regional-level determinants of variation.[14–17] A better understanding of triage decision making by physicians is, therefore, essential for quality improvement in

trauma triage. We had previously developed a vignette-based instrument to study physician decision making in trauma triage.[16] In the three studies described below, we evaluated the reliability and validity of this instrument, using classic psychometric analyses and signal detection theory. The University of Pittsburgh Institutional Review Board approved this project.

STUDY ONE

The first study in the series evaluated the test-retest and internal-consistency reliability of a vignette-based, trauma-triage instrument.

Methods

Participants—ED physicians were recruited at the national meeting of the American College of Emergency Physicians (ACEP) and were eligible to participate if they had completed residency and cared for adult ED patients at a non-trauma center or a Level III/IV trauma center in the United States. One hundred sixty-eight physicians completed the vignette instrument during the first assessment, the results of which are reported in an earlier manuscript.[16] Of these 168, a random sample of 50 were contacted one-year later to complete the instrument again, with 32 (64%) doing so. We expected a high correlation (>0.5) between time 1 and time 2 performance and calculated that power would reach 0.80 with a sample of 28 participants (α=0.05).[17]

Among the 32 participants, the average age was 43 years (SD = 8.7). On average, they were 12 years post residency (SD = 8.9); 94% were board certified in Emergency Medicine, and 63% were certified in Advanced Trauma Life Support (ATLS). There were no significant differences between the physicians in the reliability testing and the original study, on these four variables.

Trauma Triage Assessment Instrument (TTAI)—Using the ACS-COT guidelines for the transfer of trauma patients as our reference standard (see Appendix), we constructed 50 case vignettes, 30 of which described trauma patients.[16] All were based on case histories of individual patients admitted to the University of Pittsburgh Medical Center – Presbyterian Hospital trauma service. Each vignette included all the information that a physician would ordinarily obtain from a history, physical exam, chest and pelvis x-ray, including all the information the ACS-COT considers necessary to triage the patient.[18] We presented the information in the format of a completed trauma care flow sheet. [Figures 1 and 2] By design, one-half the trauma vignettes met ACS-COT criteria for transfer (mean Injury Severity Score [ISS] 21, range 9–48) and one-half did not (mean ISS 2.5, range 1– 4). Independent review by three trauma surgeons confirmed the transfer categories (κ = 0.85). We systematically varied the complexity of the cases to encompass the range of possible triage decisions. The vignette set was constructed to ensure that age, gender, and mechanism of injury were uncorrelated to injury severity or need for transfer.

After studying each vignette's care flow sheet, respondents used a free-response text box to answer the question, "What would you do to manage the patient?" They were prompted to include information about treatment, interventions, and disposition. Text response dispositions were scored as compliant or non-compliant with ACS-COT guidelines.[10] Each respondent's under-triage rate was calculated as the proportion of vignettes describing cases that meet ACS-COT guidelines for transfer that the respondent failed to transfer (number of patients with moderate to severe injuries not transferred/total number of patients with moderate to severe injuries). Each respondent's over-triage rate was calculated as the proportion of cases that the respondent chose to transfer that had minor injuries only (number of patients with minor injuries transferred/total number of patients transferred).[10]

Procedure and Design—Participating physicians accessed the TTAI via a secure website at their convenience. After completing informed consent procedures, they responded to a brief set of demographic questions and then read and responded to all TTAI case vignettes. All procedures for the 2nd assessment, one year after the first, were identical, and physicians received a $100 for participating. They received no feedback on their performance at either assessment.

We measured test-retest reliability by comparing individual physicians' triage decisions on the two administrations of the instrument using the kappa statistic,[19] and estimated the internal reliability of the two types of cases on the second administration using the Kuder-Richardson coefficient.[20]

Results

There was moderate agreement between physicians' transfer decisions ($\kappa = 0.42$, $p < 0.01$) on the first and second administration of the instrument. Internal consistency reliability was acceptable among cases that met ACS-COT criteria for transfer (KR-20 = 0.91) and among cases that did not require transfer (KR-20 = 0.71).

Discussion

The TTAI demonstrated acceptable reliability. It consistently measured individual differences in accuracy across time with physicians' choices at the second administration significantly associated with their performance a year prior. Moreover, the internal consistency of the instrument was good with physicians' selecting similar dispositions for similar cases.

Of course, reliability is a necessary but not sufficient condition for the validity of an instrument. Our assessment of validity began with a known-groups analysis, comparing the performance of trauma surgeons (a new sample) with that of the ED physicians recruited for study one. Trauma surgeons have extensive training and experience recognizing and treating subtle distinctions across trauma cases. They see such cases regularly, unlike ED physicians who may see only 1–4 trauma patients with moderate to severe injuries in a given year.[21] Given these differences in training and experience, a valid instrument of trauma triage decision making should reveal superior performance among trauma surgeons. Study two examines this hypothesis.

STUDY TWO

Methods

Participants—The ED physicians described in study one were compared to trauma surgeons in a new sample. We recruited trauma surgeons using personal contacts and snowball sampling. Eligible surgeons had completed general surgery residency and cared primarily for trauma patients. Among the 32 trauma surgeons whom we contacted, 28 (88%) completed the instrument. Our power analysis required 26 physicians in each group in order to detect the expected large effect-size difference in performance between the two groups ($\alpha = .05$, $\beta = 0.80$).[17]

Among the trauma surgeons who participated, average age was 44 years ($SD = 6.9$) and they were 11 years ($SD = 7.0$) post-residency. All surgeons worked primarily at Level I trauma centers; 26 (93%) were certified ATLS instructors.

Procedure and Design—Trauma surgeons accessed and completed the TTAI following the procedure in study one. They were instructed to respond to the instrument as if they were ED physicians working at a non-trauma center.

We evaluated the sensitivity of the instrument to known-group differences by comparing the under- and over-triage rates of ED physicians who participated in the first study with that the of trauma surgeons using Wilcoxon Mann-Whitney tests, expecting the latter to follow ACS-COT guidelines more consistently.

Results

On average, trauma surgeons under-triaged fewer cases with moderate to severe injuries (13%) than did ED physicians (70%; $z = 2.92$, $p < .01$). However, their rate of over-triage (29%) did not differ significantly from that of the ED physicians (26%; $z = -0.7$, $p = 0.48$).

In post-hoc analyses, we noted that five trauma surgeons failed to transfer any patients. We suspected that these surgeons may not have noted the instruction to behave as an ED physician (thus requiring disposition decisions) and conducted a sensitivity analysis that excluded these five participants. As would be expected, excluding participants with high under-triage rates pushed the under-triage rate among trauma surgeons (7%) even lower relative to ED physicians (70%; $z = 4.7$, $p<0.01$), but notably, left over-triage rates unchanged (29% v. 26%; $z = -0.01$, $p=0.81$). [Figure 3]

Discussion

The known-groups test supports the validity of the TTAI, corroborating the observation that vignettes can detect group-level differences in decision making.[6] Using our vignettes, trauma surgeons, with extensive training and experience in the variability of patient presentations after injury, adhered more closely to the ACS-COT guidelines than did ED physicians, with less training and experience. That difference emerged in rates of under-triage. The ED physicians transferred many fewer of the patients with moderate to severe injuries. The extent to which under-triage reflects ED physicians' inability to distinguish such cases or more conservative decision rules for transfer is addressed by the signal detection analysis performed in the final study.

We designed the TTAI to be a simple, efficient measure of physicians' triage decisions in their normal work environments. The first two studies provided evidence of its reliability and known-groups validity. The third study examined whether the instrument did in fact predict actual performance. Although, we anticipated that triage in emergency departments would be influenced by a variety of external factors such as time pressure, cognitive load, an institutional norms, we hypothesized that the characteristics of the patient injury would be the most salient factor guiding disposition, In other words, injury characteristics would explain a substantial portion of the variance in physicians' trauma triage decisions. As noted, the TTAI vignettes included complete descriptions of injury characteristics, presented in the same format as used in emergency departments. To test this hypothesis, we compared performance on the TTAI with physician's under- and over-triage rates calculated via record review.

STUDY THREE

Methods

Participants—We recruited a new sample of ED physicians from non-trauma centers run by a large healthcare system in western Pennsylvania for which we had electronic access to billing and clinical data. Physicians were eligible for participation if they had completed

residency and worked at these non-trauma centers between 2007 and 2010. Among the 50 ED physicians working at non-trauma centers for which we had access to medical records, 28 (56%) agreed to participate and completed the instrument. We estimated that a sample of 28 physicians was required to detect a large correlation (effect size = 0.5) between performance in response to the TTAI and in practice (α=.05, β=0.80).[17]

The ED physicians who participated had a mean age of 47 years (*SD* = 8.8); they had completed residency 16 years (*SD* = 9.8) previously. Many (71%) were board-certified in Emergency Medicine; 68% were ATLS certified.

Procedure and Design—Participants accessed and completed the TTAI as described in study one. Each physician's home ED trauma triage decisions were assessed via an extensive chart review.

Chart Review—We obtained de-identified medical records for patients managed by study participants between 2007 and 2010 and collected International Classification of Diseases, 9th revision, clinical modifications (ICD9-CM) codes, demographics, and discharge disposition status from the discharge abstract, and radiological procedures from radiology reports. We identified adult patients with a primary ICD9-CM code between 800 and 959. We excluded patients seen for late effects of injuries (ICD9-CM codes 905–909), foreign bodies (ICD9-CM codes 930–940), burns (ICD9-CM codes 940–950), or minor injuries, including isolated strains/sprains (ICD9-CM codes 840–849), superficial injuries (ICD9-CM codes 910–919), and contusions (ICD9-CM 920–924). The ACS-COT classifies patients as having a moderate to severe injury if they have an Injury Severity Score >15 or a "life-threatening/critical injury" (see Appendix). We used a validated computer program to translate ICD9-CM diagnostic codes into ISS.[12, 22] We identified patients with "life-threatening" or "critical" injuries with an algorithm that we created based upon ICD9-CM codes and which had strong agreement with one trauma surgeon's chart review (κ =0.8). We dichotomously scored disposition as compliant or non-compliant with ACS-COT guidelines, based on injury severity and whether the patient received a CT scan in the ED.[10, 18]

Disposition codes were collapsed across all trauma patients cared for by a given physician to calculate each physician's rate of under-triage (proportion of patients with moderate to severe injuries not transferred to trauma centers) and over-triage (proportion of transferred patients with minor injuries only).[5] Standard signal detection theory measures were calculated using the same data.

Analytic Strategy—We compared group rates of under- and over-triage on the instrument and in practice using Wilcoxon signed rank tests. We used the Spearman correlation coefficient to assess the individual-level relationship between triage decisions on the TTAI and in practice.

We also used signal detection theory (SDT) to characterize these decisions further. SDT parses decision making into two separate processes: *discrimination* between states of the world and the application of a decision rule or *threshold*.[23, 24] In the context of trauma triage, the SDT measure of discrimination quantifies a physician's perceptual ability to distinguish between patients with minor injuries and with moderate to severe ones. In our analysis, zero indicates no ability to distinguish between the two groups of patients; higher values indicate greater ability. The SDT measure of decisional threshold values quantifies a physician's tendency to err on the side of under- or over-triage. In our analysis, zero indicates equilibrium, with false positives (over-triage) treated as being as bad as false negatives (under-triage). Negative values indicate greater tolerance for false positives (hence over-triage); positive values indicate greater tolerance for false negatives (hence under-

triage). The small number of vignettes included in the TTAI precluded traditional SDT parameter estimation techniques. Therefore, we used a regression-based approach to estimate these measures for each participant. We fit a model for each physician, predicting the log-odds of his or her disposition decisions (dependent variable) using an intercept and a regression weight on the ACS-COT guidelines for transfer (independent variable). The model's intercept predicts disposition decisions in the absence of any information, and hence provides an estimate for the decisional threshold. The regression weight on the ACS-COT guideline variable represents perceptual sensitivity, showing the degree of reliance on guidelines (implicitly or explicitly) when making disposition decisions. We scaled our estimates by 1.8 to approximate the standard deviation units of signal detection values. [16, 25]

Results

In response to the TTAI case vignettes, physicians in study three transferred a median of 10 (IQR 5–13) of 15 patients with moderate to severe injuries, and a median of 4 (IQR 2–8) of 15 patients with minor injuries. One physician did not transfer any patients. The median under-triage rate was 33% (IQR 13–67). The median over-triage rate was 33% (IQR 22–50%).

In their home ED practices, these physicians evaluated a median of 2,423 patients (inter-quartile range [IQR] 1637–3030) per year during the four-year study period, including 148 (IQR 92–246) trauma patients meeting our inclusion criteria. These ED physicians saw a median of 3 patients (IQR 1–4) with moderate-to-severe injuries each year and transferred a median of 9 patients (IQR 4–14). All failed to transfer the vast majority of their patients with moderate to severe injuries (under-triage rate IQR 90–100%). Their over-triage rates—the proportion of patients who were transferred to a Level 1 trauma center with only minor injuries—was also very high (IQR 94–100%).

These physicians had much lower rates of under-triage on the vignettes than in their practices (medians: 33% v. 100%, $z = -4.2$, p<0.01) and lower median rates of over-triage (medians: 33% v. 100%, $z = -4.3$, p<0.01).

Among physicians who saw at least one patient with moderate to severe injuries, their trauma triage performance as measured by the TTAI and in their actual practices were unrelated (under-triage rate: $r = -0.17$, $p = 0.4$; over-triage rate: $r = -0.03$, $p = 0.85$). [Figure 4]

Signal detection analysis—Adherence to the guidelines on 69% of both minor and moderate to severe cases would produce a discrimination estimate of 1. The mean discrimination estimate in response to TTAI vignettes was 0.97 (SD = 1.2) and was 0.70 (*SD* = 1.0) for physicians' performance in practice (see [16] for procedure). Therefore physicians demonstrated moderate ability to discriminate between minor and moderate to severe injuries, which did not vary between the instrument and real-life ($z = -1.0$, p=0.3).

The mean decisional threshold estimate was 0.72 (*SD* = 1.2) on the TTAI and 2.07 (*SD* = 0.6) in practice. Both values indicated a tendency to err on the side of false negatives (failing to transfer patients with moderate to severe injuries), which was significantly greater in practice than in response to case vignettes ($z = -3.8$, $p < .01$). [Figure 5] Across these physicians, there was no correlation in perceptual discrimination scores on the TTAI and in practice ($r = -0.2$, $p = 0.3$); nor was there any correlation between decisional threshold scores ($r = -0.1$, $p = 0.71$).

Discussion

The TTAI showed both good reliability and known-groups validity. Nonetheless, it did not predict individual physicians' decision making, a high standard for assessing the validity of an instrument – but often the one implicitly assumed to be met by studies relying on vignettes. Physicians' decision-making performance on the TTAI case vignettes was unrelated to their performance in practice. For example, as shown in Figure 4, physician 151 performed almost perfectly on the TTAI, correctly triaging 14 of the 15 case vignette patients with moderate to severe injuries. However, that physician transferred zero (of 18 patients) with moderate to severe injuries treated in actual practice. What might explain this discrepancy?

Researchers in the traditions of Egon Brunswick and E.C. Poulton, have argued that the external validity of experiments studying psychological processes depends on how well its stimuli represent those of the world to which results are generalized.[26–28] When creating the TTAI, we recognized that the task presented by the "paper cases" would differ from actual clinical situations in three important ways. First, the vignettes provided unusually good conditions for evaluating clinical evidence, summarizing all relevant information (and only relevant information), in a standard synoptic format, with unlimited time to respond, and no competing demands. Studies finding external validity for vignette-based instruments have focused on tasks where clinical practice may approximate these conditions (e.g., common, low-acuity clinical situations, such as low back pain).[2–4] However, in trauma triage, as in other acute-care decisions, physicians must rapidly assemble and integrate the relevant clinical information, while discarding irrelevant evidence, under the time pressure and cognitive load imposed by other patients and competing responsibilities. Second, as with many other vignette studies, we included a disproportionate number of true positive cases (i.e patients with moderate to severe injuries), relative to actual practice, so as to be able to assess under-triage rates with a reasonably sized test set.[6, 9] As a result, the rate of injuries requiring transfer in the TTAI was much higher than in actual practice (0.5 vs 0.001).[21] Realizing that, our physician respondents might have been primed to pick up information that they might miss in their more mundane, normal practices.[30] Third, we could not replicate the incentives and environmental influences of actual practice, such as distance to a trauma center, institutional norms about adherence to guidelines, or the influence of social networks.[31–33]

We conducted a signal detection analysis to clarify which of the task properties described above might have driven differences between performance on the vignettes and in practice. Developed to improve the performance of radar operators in World War II, signal detection theory parses decisions into two dimensions: perceptual and decisional.[23, 24] Perceptual factors influence individuals' ability to distinguish between groups (called *discrimination*). Decisional factors shape individuals' willingness to tolerate different kinds of errors (determining their *decisional threshold*). Factors that make a task harder (e.g., time pressure, irrelevant information, interruptions) should affect discrimination.[34, 35] Factors that affect incentives (e.g., payment schemes, social pressure) should affect decisional thresholds. In theory, if variation occurred primarily because of task properties affecting either the perceptual or decisional dimension of decision making, we could expect to see a correlation in performance along the other dimension.

At the group-level, we found physicians had different decision thresholds in the two settings rather than different discrimination ability. As shown in Figure 5, relative to vignette performance, in real-life physicians had a much lower tolerance for transferring patients with minor injuries, explaining differences between rates of under- and over-triage seen on the vignettes and in practice. However, at the individual-level, no correlation existed

between performance on the instrument and performance in practice either for their discrimination or their threshold scores.

The lack of correlation between the performance of physicians on the instrument and in real-life affirms the claim (of Brunswik, Poulton, and others) that experimental results cannot be extrapolated to real-world setting unless the conditions match one another on key properties.[36] In designing the instrument, we assumed (as have others using vignettes to study physician practice patterns) that decision making would primarily reflect physicians' interpretation of the available clinical information. However, our results suggest that non-clinical elements of the task (e.g. cognitive load, institutional norms) may swamp the effect of the clinical cues. The TTAI was sensitive to differences between the performance of trauma surgeons and ED physicians. As a result, it might have potential for evaluating the knowledge that ED physicians acquire through their experience and training. However, we would need to modify the instrument to accomplish the original objective of understanding physician decision making in practice. Potential strategies would include changing the types of vignettes from "paper cases" to simulated patients or more interactive scenarios that incorporate important contextual factors, such as time-pressure and uncertainty.

We note the following potential limitations to our studies. First, the sample sizes for all three studies only allowed us to detect large effect sizes ($r=0.5$). We believed that an instrument that predicted actual practice any less robustly would lack clinical significance. Second, we measured reliability by comparing responses made after a one-year interval. That time period reduced the chance of participants remembering their initial response, while increasing the chances that their decision making strategies changed between administrations.[37] Third, the small number of patients with moderate-to-severe injuries and patients transferred by each physician reduced the precision of our evaluation of performance. Small case loads have typically hampered the use of administrative data as a measure of physician quality.[5] Nonetheless, no good surrogate exists to evaluate actual performance in the emergency setting. Along the same lines, we grouped all injuries in two large classes, potentially missing better correspondence among subclasses. As it took four years of records to produce these samples, any subclass analysis would require a very large database.[5]

In conclusion, we found that although the TTAI performed reliably and detected known group differences, it did not predict individual physician practice patterns in trauma triage. We believe that our observations provide a cautionary tale against assuming that performance on all vignettes will predict actual decision making. Researchers interested in using vignettes must carefully consider how contextual factors might influence performance. Future improvements in the representativeness of our case vignettes using dynamic simulation to capture environmental factors may improve the validity of the TTAI.

Acknowledgments

References

1. Sirovich B, Gallagher PM, Wennberg DE, Fischer ES. Discretionary decision making by primary care physicians and the cost of US health care. Health Aff. 2008; 27:813–823.

2. Kostopoulou O, Mousoulis C, Delaney B. Information search and information distortion in the diagnosis of an ambiguous presentation. Judgm and Decis Mak. 2009; 4:408–415.

3. Drexel C, Jacobson A, Hanania NA, Whitfield B, Katz J, Sullivan T. Measuring the impact of a live, case-based, multiformat, interactive continuing medical education program on improving clinician

knowledge and competency in evidence-based COPD care. Int J Chron Obstruct Pulmon Dis. 2011; 6:297–307. [PubMed: 21697994]

4. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. Ann Intern Med. 2004; 141:771–780. [PubMed: 15545677]

5. Nyweide DJ, Weeks WB, Gottlieb DJ, Casalino LP, Fisher ES. Relationship of primary care physicians' patient caseload with measurement of quality and cost performance. JAMA. 2009; 302:2444–2450. [PubMed: 19996399]

6. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA. 2000; 283(no. 13):1715–1722. [PubMed: 10755498]

7. Shah R, Edgar DF, Evans BJW. A comparison of standardised patients, record abstraction and clinical vignettes for the purprose of measuring clinical practice. Ophthal Physiol. 2010; 30:209–224.

8. Lucet J, Nicolas-Chanoine M, Lefort A, et al. Do vignettes accurately reflect antibiotic prescription? Infect Control Hosp Epidemiol. 2011; 32:1003–1009. [PubMed: 21931251]

9. Lee CY, Bernard AC, Fryman L, et al. Imaging may delay transfer of rural trauma victims: a survey of referring physicians. J Trauma. 2008; 65:1359–1363. [PubMed: 19077627]

10. Committee on Trauma American College of Surgeons. Resources For Optimal Care of the Injured Patient. Chicago, IL: American College of Surgeons; 2006.

11. Nathens AB, Jurkovich GJ, MacKenzie EJ, Rivara FP. A resource-based assessment of trauma care in the United States. J Trauma. 2004; 56:173–178. [PubMed: 14749585]

12. Mohan D, Rosengart MR, Farris C, Cohen E, Angus DC, Barnato AE. Assessing the feasibility of the American College of Surgeons0027 benchmarks for the triage of trauma patients. Arch Surg. 2011; 146:786–792. [PubMed: 21422329]

13. MacKenzie EJ, Rivara FP, Jurkovich GJ, et al. A national evaluation of the effect of traumacenter care on mortality. NEJM. 2006; 354:366–378. [PubMed: 16436768]

14. Chang DC, Bass RR, Cornwell EE, Mackenzie EJ. Undertriage of elderly trauma patients to state-designated trauma centers. Arch Surg. 2008; 143:776–781. [PubMed: 18711038]

15. Haas B, Stukel TA, Gomez D, et al. The mortality benefit of direct trauma center transport in a regional trauma system: a population based analysis. J Trauma. 2012; 72(6):1510–1517.

16. Gomez D, Haas B, de Mestral C, et al. Institutional and provider factors impeding access to trauma center care: an analysis of transfer practices in a regional trauma system. J Trauma. 2012; 73(5):1288–1293.

17. Mohan D, Rosengart MR, Farris C, Fischhoff B, Angus DC, Barnato AE. Sources of non-compliance with clinical practice guidelines in trauma triage: a decision science study. Implement Sci. 2012 Forthcoming.

18. Cohen J. A power primer. Psychol Bull. 1992; 112:155–159. [PubMed: 19565683]

19. American College of Surgeons Committee on Trauma. Advanced Trauma Life Support for Doctors. Chicago, IL: American College of Surgeons; 2008.

20. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG. Tips for learners of evidence-based medicine: measures of observer variability (kappa statistic). CMAJ. 2004; 171:1369–1373. [PubMed: 15557592]

21. Nunnally, J. Psychometric Theory. 2nd edition. New York: McGraw Hill; 1978.

22. Mohan D, Barnato AE, Rosengart MR, et al. Trauma triage in the Emergency Departments of non-trauma centers: an analysis of individual physician case-load on triage patterns. J Trauma. 2013 Forthcoming.

23. MacKenzie EJ, Steinwachs DM, Shankar B. Classifying trauma severity based on hospital discharge diagnoses: validation of an ICD9-CM to AIS-85 conversion table. Med Care. 1989; 27(4):412–422. [PubMed: 2649755]

24. Swets JA, Dawes RM, Monahan J. Psychological science can improve diagnostic decisions. Psychol Sci Public Interest. 2000; 1(1):1–26.

25. Macmillan, NA.; Creelman, CD. Detection Theory: A User's Guide. New York: Lawrence Erlbaum Associates; 2005.

26. DeCarlo TL. Signal detection theory and generalized linear models. Psychol Methods. 1998; 3(2): 186–205.

27. Hammond, KR., editor. The psychology of Egon Brunswik. New York: Holt, Rinehart and Winston; 1996.

28. Poulton, EC. Behavioral decision theory. Cambridge: Cambridge University Press; 1994.

29. Poulton, EC. Bias in quantifying judgment. Hove: Erlbaum; 1989.

30. Finkelstein, EA.; Corso, PS.; Miller, TR. Associates. The incidence and economic burden of injuries in the United States. New York: Oxford University Press; 2006.

31. Birnbaum MH. Base rates in bayesian inference: signal detection analysis of the cab problem. Am J Psychol. 1983; 96:85–94.

32. Doumaras AG, Haas B, Gomez D, et al. The impact of distance on triage to trauma center care in an urban trauma system. Prehosp Emerg Care. 2012 Epub ahead of print.

33. Bradley EH, Curry LA, Webster TA, et al. Achieving rapid door-to-balloon times: how top hospitals improve complex clinical systems. Circulation. 2006; 113:1079–1085. [PubMed: 16490818]

34. Landon BE, Keating NL, Barnett ML, et al. Variation in patient-sharing networks of physicians across the United States. JAMA. 2012; 308:265–273. [PubMed: 22797644]

35. Thompson C, Dalgleish L, Bucknall T, et al. The effects of time pressure and experience on nurses' risk assessment decisions. Nurs Res. 2008; 57(5):302–310. [PubMed: 18794714]

36. Gillard E, van Dooren W, Schaeken W, et al. Proportional reasoning as a heuristic-based process. Exp Psych. 2009; 56(2):92–99.

37. Dhami MK, Hertwig R, Hoffrage U. The role of representative design in an ecological approach to cognition. Psychol Bull. 2004; 130:959–988. [PubMed: 15535744]

38. Carmines, EG.; Zeller, RA. Reliability and validity assessment. Thousand Oaks, California: Sage Publications Inc; 1979.

Appendix. Algorithm used to select injuries described by the ACS-COT as indicating the need for transfer

Injuries the ACS-COT considers life-threatening or critical	ICD9-CM codes	Assumptions/Notes
Carotid or vertebral injuries	900	
Aorta or great vessel injuries	901	
Cardiac rupture	861.0 and 861.1	1. Unable to identify cardiac rupture from diagnosis codes so included all cardiac injuries.
Bilateral pulmonary contusions with PaO$_2$/FiO$_2$ ratio<200	-	1. Unable to identify based on discharge codes
Major abdominal vascular injury	902	
Grade IV or V liver injury with >6 units pRBC	864.04 or 864.14	1. Unable to identify RBC transfusion so included all Grade IV liver lacerations
Unstable pelvic fracture with >6 units pRBC	808.43 or 808.53	1. Unable to identify unstable fractures so used the surrogate of the disrupted pelvic circle.
Fracture or dislocation with loss of distal pulses	903.1 – 903.3 in conjunction with 812, 813, and 818 904.1–904.5 in conjunction with 821, 822, 823, 824, and 827	1. Unable to identify fractures with loss of pulses so used the surrogate of fracture with vascular injury. 2. Excluded patients with amputations, which would technically fit into this category, because of the problem of misclassification [*]

Injuries the ACS-COT considers life-threatening or critical	ICD9-CM codes	Assumptions/Notes
Open skull fracture or penetrating injury	800.5–800.9 801.5–801.9 803.5–803.9 804.5–804.9	1. Unable to identify penetrating injury so captured only the open skull fractures.
GCS<14 or lateralizing neurologic injury	800–804 850.2–850.5 851–854	1. Unable to calculate GCS from discharge diagnosis codes so used the fifth digit subclassification to identify patients who had either a moderate (1–24h) or prolonged (>24h) loss of consciousness (e.g. 852.03). We assumed these patients would appear clinically to have a GCS<14.
Spinal cord deficit or lateralizing neurologic sign	806 952–955.2 956–956.2	
Spinal column fractures	-	1. Excluded because of the problem of misclassification[*]
>2 unilateral rib fractures or bilateral rib fractures	807.03–807.1 807.4–807.6	1. Unable to identify patients with bilateral rib fractures so included only patients with >2 rib fractures.
Open long bone fracture	812.1, 812.3, 812.5 813.1, 813.3, 813.5 820.1, 820.3, 820.5 821.1, 821.3, 821.5 823.1, 823.3, 823.5 824.1, 824.3, 824.5	
Significant torso injury with advanced co-morbid disease	860.1, 860.3, 860.5 861.2–861.9 862–870	1. ICD9-CM codes indicated significant torso injury. We included patients who had these ICD9 codes as well as an Exlixhauser co-moribidity code.

[*] Amputations and spinal column fractures were poorly captured in ICD9-CM discharge codes. They frequently appeared in the medical record discharge abstract even when the injury was chronic, rather than acute. To avoid misclassification, we excluded patients with these two injuries from our analysis (see [21] for further details).

Figure 1. Example of case vignette depicting a patient with a moderate-severe injury
It is presented on a trauma care flow sheet, in compliance with ACS-COT suggestion for the capture of clinical information. This particular vignette describes a patient with an aortic transection after a motor vehicle collision (a severe injury).
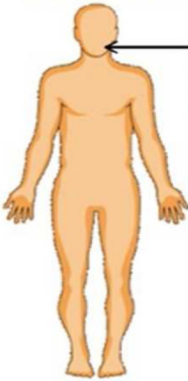
Decision Making in Trauma

TRAUMA CARE FLOW RECORD

Assumptions
- Patient on back board with c-collar in place.
- 2 18G IVs in place.
- IVF hanging.
- Exam normal except where noted on primary and secondary survey.

Mechanism of Injury:
- ☐ MVC   ☐ Stab   ☐ Fall _____ ft.
- ☐ ATV   ☐ Assault   ☐ Other
- ■ GSW   ☐ Motorcycle

Chief Complaint:
38 year male, shot while driving. Heard 1 shots. One struck him in the face. Patient drove himself to the ED. H/o recent cocaine use.

PRIMARY SURVEY

Respiratory

Breath Sounds	R	L
Clear	■	■
Dimished	☐	☐
Absent	☐	☐

Airway
- ■ Patent
- ☐ Compromised
- ☐ Obstructed

Cardiovascular: Pulse
- ■ Strong
- ☐ Palpable
- ☐ Weak

Vitals: Pulse 105   BP 154/92   RR 24
Rhythm Regular   Pulse OX 98%   O$_2$ 2L

GLASGOW COMA SCALE: GCS TOTAL: 15

Eye Opening		Verbal Response		Motor Response	
Spontaneous	④	Oriented	⑤	Obeys Commands	⑥
To Voice	3	Confused	4	Localizes (Pain)	5
To Pain	2	Inapprop. Words	3	Withdraws (Pain)	4
None	1	Incomprehensible	2	Flexion (Pain)	3
		None	1	Extension (Pain)	2
				None	1

Neurological:
Oriented to: ■ Person ■ Place ■ Time ■ Event
Pupils: R   L
Brisk ■ ■
Sluggish ☐ ☐
Non-Reactive ☐ ☐
Baseline: GCS 15

SECONDARY SURVEY
Entry wound L cheek. Patient complains of feeling projectile below L lower molar

Medical History: Denies

Medications: Denies

IMAGING
CXR: Pelvis:
Other:

Figure 2. Example of case vignette depicting a patient with a minor injury
This vignette describes a patient shot in the face, but with only minor injuries.
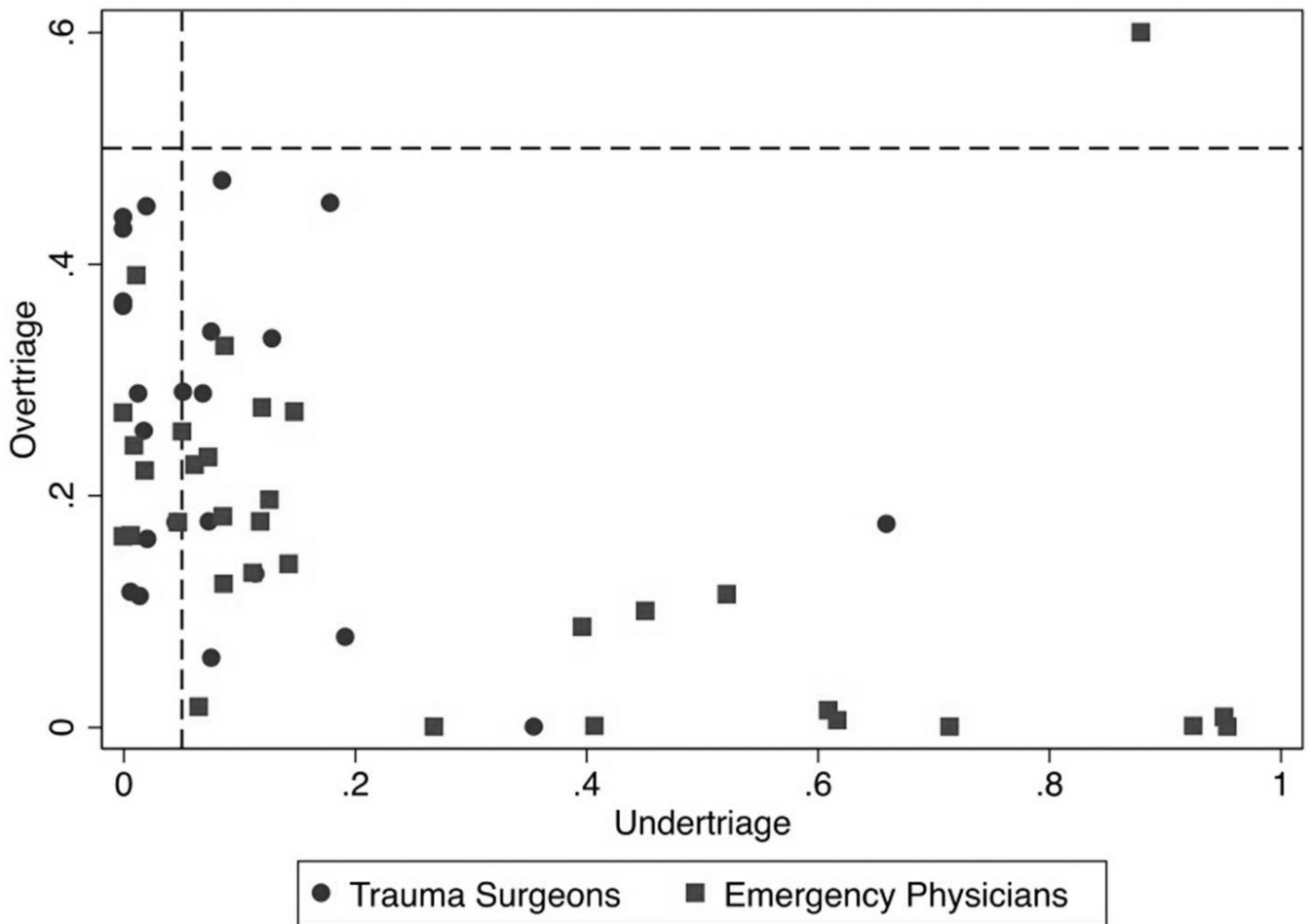
Figure 3. Known groups performance
Comparison of rates of under- and over-triage among trauma surgeons who transferred at least one patient and emergency physicians. To meet ACS-COT benchmarks for performance, physicians should under-triage<5% and overtriage< 50% of patients (shown with dotted lines).
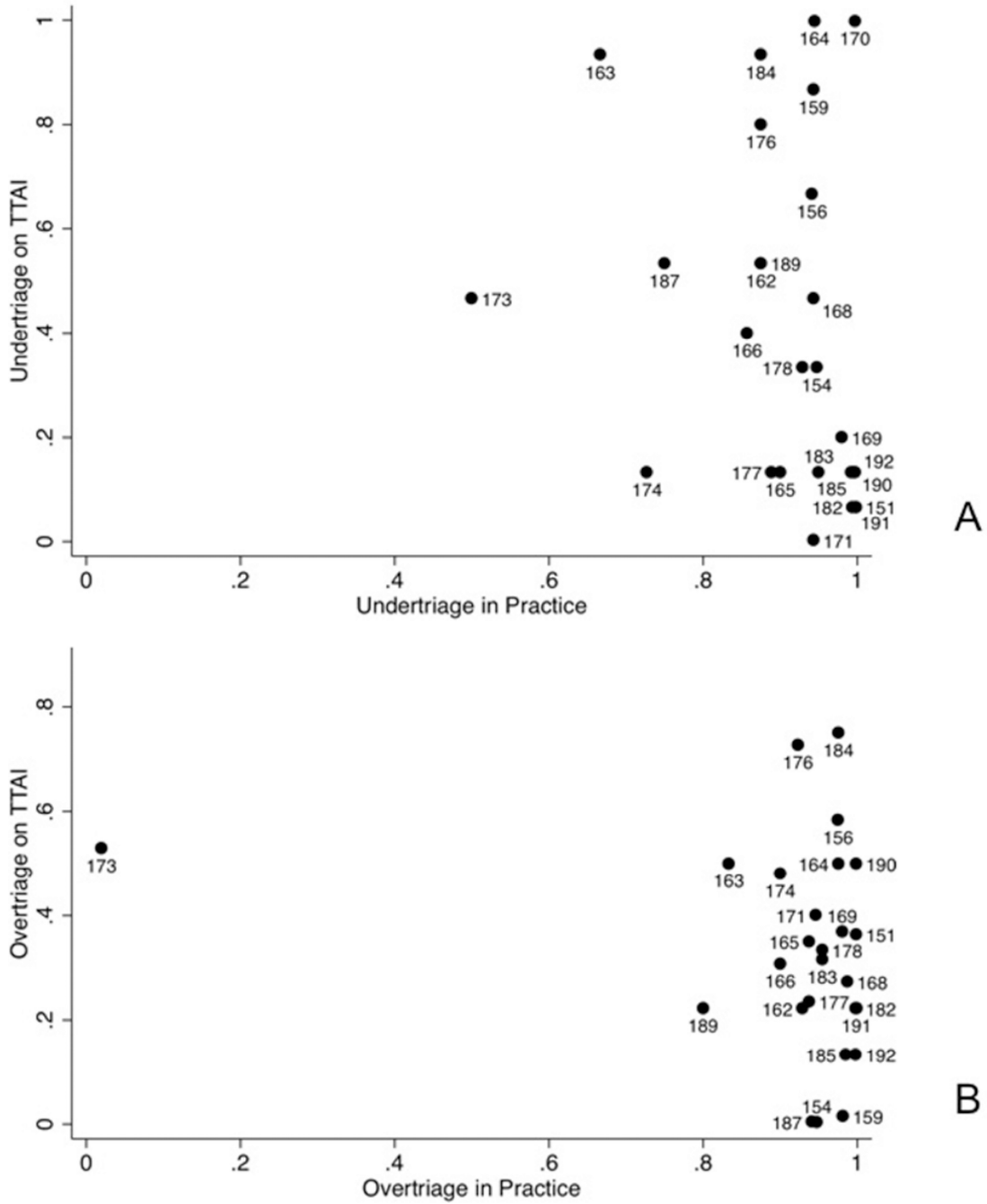
Figure 4. Criterion validity
Rates of under-triage (Panel A) and overtriage (Panel B) in practice and in response to case vignettes. There was no correlation between performance on the instrument and in practice for either under or over-triage ($r = -0.17$, $p = 0.4$; $r = -0.03$, $p = 0.85$).
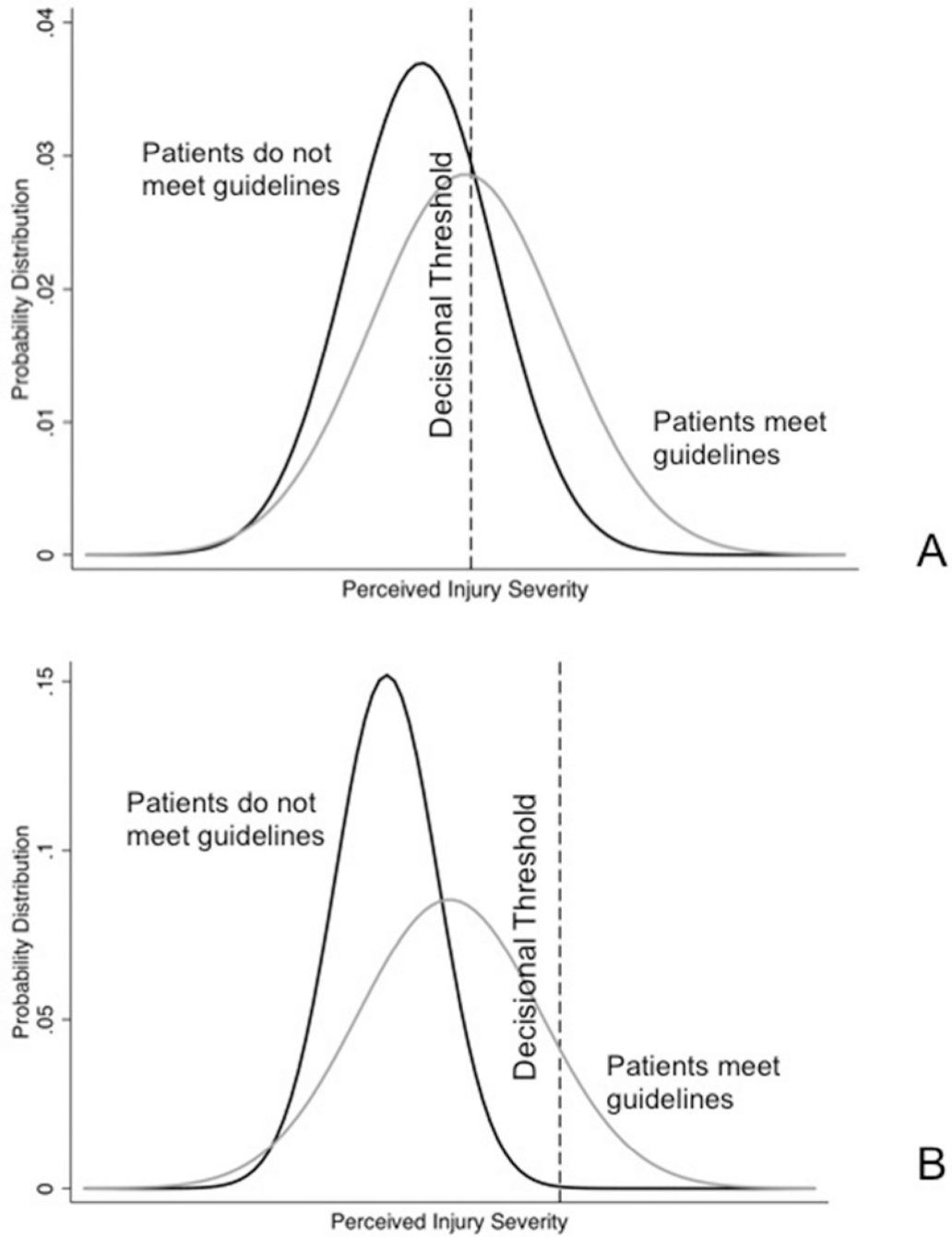
Figure 5.
Signal detection analysis of decision making on the TTAI (Panel A) and in practice (Panel B). Physicians perceive injuries on a continuum ranging from very minor injury to severe injury. *Discrimination* can be thought of as the standardized distance between the means of the two distributions. It reflects the ease with which a physician perceives the distinction between the two categories. The *decisional threshold* is the point on the continuum above which physicians transfer patients. It reflects the types of errors the physicians tolerate. Physicians demonstrated lower decisional thresholds for the transfer of patients on the vignettes than in practice (0.72 v. 2.07, p<0.01).