Received: 24 February 2023 | Revised: 24 May 2023 | Accepted: 14 June 2023

DOI: 10.1111/acem.14768

ORIGINAL ARTICLE

AEM Academic Emergency Medicine

Accuracy of physician gestalt in prediction of significant abdominal and pelvic injury in adult blunt trauma patients

Nicole M. Rodriguez MS[1] | William R. Mower MD, PhD[2] | Ali S. Raja MD, MBA, MPH[3] | Malkeet Gupta MD, MS[2,4] | Juan Carlos Montoy MD, PhD[1] | Blair Parry[3] | Virginia Chan MPH[1] | Angela H. K. Wong[1] | James Wilcox[4] | Alexandra Quiñones[4] | Robert M. Rodriguez MD[1]

[1]University of California, San Francisco, California, USA

[2]University of California, Los Angeles, California, USA

[3]Massachusetts General Hospital, Boston, Massachusetts, USA

[4]Antelope Valley Medical Center, Lancaster, California, USA

Correspondence
Robert M. Rodriguez, MD, Department of Emergency Medicine, University of California, San Francisco, 1001 Potrero Ave, Bldg 5 Rm 6A, San Francisco, CA 94110 USA.
Email: robert.rodriguez@ucsf.edu

Abstract

Objective: Focusing on potential missed injury rates and sensitivity of low-risk of injury predictions, we sought to evaluate the accuracy of physician gestalt in predicting clinically significant injury (CSI) in the abdomen and pelvis among blunt trauma patients presenting to the emergency department (ED).

Methods: We collected gestalt data on physicians caring for adult blunt trauma patients who received abdominal/pelvic computed tomography (CT) at three Level I and one Level II trauma centers. The primary outcome of CSI was defined as injury on abdominal/pelvic CT requiring hospitalization or intervention. Physicians evaluating trauma patients estimated the likelihood of CSI prior to abdominal/pelvic CT review (response choices: <2%, 2%–10%, 11%–20%, 21%–40%, >40%). We evaluated potential missed injury rates (prevalence of CSI) and sensitivity for prediction categories, as well as calibration and area under the receiver operating characteristic (AUROC) curve for overall physician gestalt.

Results: Of 2030 patients, 402 (20%) had an injury on abdominal/pelvic CT and 270 (13%) had CSI. The <2% risk of CSI gestalt cutoff had a potential missed injury rate of 5.6% and a sensitivity of 95.2% (95% confidence interval [CI] 91.7%–97.3%). The 0%–10% cutoff of CSI gestalt had a potential missed injury rate of 6.3% (95% CI, 5.0%–7.9%) and a sensitivity of 75.2% (95% CI 69.5%–80.1%). With an overall AUROC of 0.699 (95% CI 0.679–0.719), physician gestalt was moderately accurate and calibrated for the midranges of predicted risk but poorly calibrated at the extremes.

Conclusions: Physician gestalt for the prediction of adult abdominal and pelvic CSI is moderately accurate and calibrated. However, the potential missed CSI rate and low sensitivity of the low perceived risk of injury cutoffs indicate that gestalt by itself is insufficient to direct selective abdominal/pelvic CT use in adult blunt trauma patient evaluation.

Supervising Editor: Christopher R. Carpenter

INTRODUCTION

The abdominal/pelvic CT is an integral component of the diagnostic evaluation of patients with blunt abdominal and pelvic trauma. In addition to providing detailed visualization of injuries to inform clinical management, abdominal/pelvic CT scans also function as a reliable screening tool, providing reassurance to both physicians and patients that those without clinically significant injuries (CSIs) can be safely discharged.[1] However, indiscriminate, reflexive abdominal/pelvic CT use in abdominal or pelvic blunt trauma may lead to CT overuse with attendant low diagnostic yields, high costs, and excessive radiation exposure in younger trauma patient populations.[2–4] The problematic downstream effects of excess CT (and other advanced imaging) use as part of a growing culture of overdiagnosis in the emergency department (ED) have been highlighted in several academic consensus conference meetings.[5,6]

In response to the need for objective tools to avoid CT overuse, several clinical decision rules (CDRs) have been derived and validated to guide selective head,[7,8] neck,[9–11] and chest[12–14] CT use in adult patients with blunt trauma. However, the only rule that has been developed for adult abdominal trauma included criteria that would not be readily available when decisions for CT are made (hematuria level greater than or equal to 25 red blood cells/high-powered field) and has not been adopted widely for use.[15] Recognizing the lack of robust, validated instruments for selective abdominal/pelvic CT in adult blunt trauma,[16] a consensus conference on CDRs cited this indication as one of the highest priority scenarios for need of a CDR.[5]

Whether a rule for selective abdominal/pelvic CT use in blunt trauma would improve upon current practice depends on the screening performance characteristics, especially missed injury rate and sensitivity, of baseline clinician gestalt.[17] If trauma physician gestalt based on history and physical examination adequately screens for blunt trauma injuries seen on CT, then there may be no need for the development of a CDR to guide selective imaging for that anatomic region.[5] With a focus on the potential missed injury rate and sensitivity in cases that physicians predicted to have low risk of CSI, we sought to evaluate the accuracy of physician gestalt in predicting clinically significant abdominal and pelvic injuries among trauma patients presenting to the ED by comparing physician predictions to abdominal/pelvic CT reports and patient outcomes.

METHODS

Study design and setting

From January 2017 to March 2020, we conducted this study adjunctive to a parent, observational, multicenter, prospective study seeking to derive criteria that may guide selective CT use in acute blunt abdominal trauma evaluation. Our study setting included three urban, Level I trauma centers (Zuckerberg San Francisco General Hospital [San Francisco], Ronald Reagan University of California Los Angeles Medical Center [Los Angeles], and Massachusetts General Hospital [Boston]) and one community hospital Level II trauma center (Antelope Valley Medical Center [Lancaster, CA]). We obtained institutional review board approval from the participating institutions with a waiver of informed consent.

We collected gestalt data from physicians (emergency medicine residents and attendings and general surgery residents) who were directing blunt trauma patient evaluations in which an abdominal/pelvis CT was ordered. Cases were excluded if patients had penetrating trauma, presented to the ED more than 72 h after trauma, or were aged <15 years or if the physician had viewed radiographic images or radiographic reports prior to completion of the study survey instrument.

Data collection procedures

Research staff screened ED dashboards and trauma alerts for eligible patients. After an abdominal/pelvic CT was ordered but before the CT scan was complete, research staff asked the primary physician who was directing the trauma patient evaluation to complete the abdominal/pelvis CT clinical questionnaire (one assessment per patient). These instruments included patient demographic characteristics and whichever trauma presentation elements were available at the time of CT ordering (mechanism of injury, systolic blood pressure, alertness, physical examination, lactate levels, chest x-ray, and focused assessment with sonography for trauma [FAST]). In reference to the abdominal/pelvic CT, physicians were asked the following question: "Please estimate the likelihood of significant injury requiring either intervention or hospital observation/admission," with response options including "<2%, 2%–10%, 11%–20%, 21%–40%, and >40%" (Appendix S1.). We revised the initial data collection instrument at approximately the midpoint of study enrollment to add training-level data on physician participants.

Measurements

Prior to study initiation, we implemented a modified Delphi process[18] with an expert panel consisting of three trauma surgeons and three emergency medicine physicians to define four study outcome classifications of injury seen on CT: *major injury* (injury within the abdomen, pelvis, or lumbar spine requiring procedural intervention[s], defined as surgical and interventional radiology procedures or blood transfusions); *minor injury* (injury within the abdomen, pelvis, or lumbar spine requiring hospitalization without procedural interventions); *insignificant injury* (injury within the abdomen, pelvis, or lumbar spine requiring neither intervention nor hospitalization, e.g., lumbar spine transverse process fracture); and *no injury* (no injury, chronic injuries [e.g., healing lumbar fracture] or incidental findings unrelated to recent trauma). We determined injury classifications using CT radiology reports from radiologists who were unaware of the study's implementation, along with electronic health record (EHR) review (Appendix S2.). Prior to EHR abstraction procedures, the lead investigators conducted multiple training sessions and implemented quality control measures with research staff (clinical research coordinators and student research volunteers) on an approximately quarterly basis.[19] Chart abstractors were unaware of the physician assessment and gestalt responses. Using a random sample of 50 cases, we performed an interrater reliability assessment (Pearson's kappa) of injury classification (two chart abstractors independently reviewed CT reports and EHR to classify each of the 50 cases). We also conducted a verification bias assessment (3-month phone and EHR follow-up) on a sample of patients who did not receive CT to evaluate this group for missed injuries.

Our primary outcome was CSI, defined as injury requiring procedural intervention or hospital admission, i.e., either major or minor injury per the above classification scheme. Our primary predictor was physician gestalt, defined as physicians' estimates of likelihood of CSI seen on abdominal/pelvic CT.

We also asked physicians to indicate reasons for ordering a CT scan if their suspicion for a CSI was low (in the first two prediction categories of <2% and 2%–10%). We provided the following response options and varied their order throughout the study: (1) need to diagnose all injuries even if they are not clinically significant, (2) severe mechanism of injury, (3) medicolegal concern, (4) this level of risk is still important to diagnose, (5) trauma protocol, (6) physical examination unreliable because of intoxication or altered mental status, and (7) other (with the request to explain through a narrative response).

Data analyses

Toward our primary objective of determining the adequacy of physician gestalt to rule out abdominal and pelvic CSI, we set an a priori cutoff missed CSI rate of <2%; i.e., if the missed CSI rate was ≥2% for those predicted to have a <2% likelihood of CSI, then gestalt would not have adequate screening function. Understanding that there is a great degree of variation in acceptable miss rates, we based this cutoff choice on literature review of missed rate and sensitivity thresholds in trauma research and via our expert panel consensus.[14,18,20–22]

To assess the overall screening accuracy of gestalt, we evaluated sensitivity, specificity, calibration, and the area under the receiver operating curve (AUROC) at four cutoff points of gestalt predictions (2%, 10%, 20%, 40%), using standard formulae (with the 2% threshold, for example, true positive = cases predicted to have ≥2% risk of CSI and found to have CSI; true negative = cases predicted to have <2% risk of CSI and found to not have CSI; false positive = cases predicted to have ≥2% risk of CSI and found to not have CSI; false negative = cases predicted to have <2% risk of CSI and found to have CSI).

Because our data are multilevel (five categories of prediction ranges), we calculated interval likelihood ratios (ILR) instead of positive and negative values and likelihood ratios; ILR is the probability that patients within an interval, e.g., 2%–10% likelihood of injury, would have the disease of interest divided by the probability that participants within the same interval would not have the disease of interest. Transforming multilevel categories into dichotomous variables of cutoff percentages would result in loss of information associated with the original prediction ranges and create aggregated categories that dilute the accuracy of the data and have limited clinical utility.[23–26]

All data were deidentified in terms of patient and provider. We managed data using REDCap tools hosted by the University of California, San Francisco, and the University of California, Los Angeles, and used Stata, Version 16.1 (StataCorp LLC) for data analysis.[27] We report findings in alignment with Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies.[28]

RESULTS

Patient population

We collected data from 2030 trauma patient cases. Their median (IQR) age was 48 (31–65) years; most patients were male (66%), White (57%), and non-Hispanic (67%; Table 1). The most common mechanisms of injury included motor vehicle collisions (25%), falls (19%), motorcycle collisions (14%), and pedestrian vehicle accidents (13%). Six percent of patients were clinically unstable, 4% had hypotension, and 3% had positive FAST exams. Approximately 23% displayed abnormal levels of alertness and 30% were deemed to have distracting injuries.

Among the 2030 patients, 402 (20%) had an abdominal/pelvic injury on CT with 270 (13%) having a CSI. Of patients with CSI, 145 (54%) had major injuries and 125 (46%) had minor injuries. The interrater agreement of CSI classifications was 0.97. Our verification bias assessment showed that there were no cases of missed injuries among enrolled participants. The proportion of abdominal or pelvic CSI within each risk group increased with estimated risk of injury (Table 2); 5.6% of those estimated to have <2% likelihood of injury had CSI and 29.9% of those estimated to have >40% likelihood of injury had CSI.

We further categorized major, minor, and insignificant injuries into injury subtypes based on organ system. Among those with major injury, the three most common injuries were spinal injuries requiring intervention (24%), pelvic fractures requiring intervention (16%), and hip fractures requiring intervention (16%). Among those with minor injury, the most common injuries included spinal injuries requiring observation (34%), splenic injuries requiring observation (14%), and pelvic injuries requiring observation (11%). Among those with non-CSI injury, the most common injuries included stable spinal injuries requiring neither intervention nor observation (67%), pelvic fractures requiring neither intervention nor observation (11%), and minimal intraabdominal contusions requiring neither intervention nor observation (5%).

TABLE 1 Patient characteristics.

Characteristic	N = 2030
Age (years)	48 (31–65)
Sex	
Male	1336 (65.8)
Female	692 (34.1)
Unknown	2 (0.1)
Race	
White	1157 (57.0)
Black	247 (12.2)
Asian	149 (7.3)
Native American	9 (0.4)
Middle Eastern	36 (1.8)
Other	301 (14.8)
Unknown	131 (6.5)
Ethnicity	
Hispanic	515 (25.4)
Non-Hispanic	1360 (67)
Unknown	155 (7.6)
Mechanism of injury	
Motor vehicle collision	506 (24.9)
Motorcycle collision	286 (14.1)
Pedestrian vehicle accident	254 (12.5)
Bicycle accident	149 (7.3)
Fall	381 (18.8)
Assault	39 (1.9)
Other	76 (3.8)
Unknown	339 (16.7)
Patient Status	
Stable	1490 (73.4)
Unstable	126 (6.2)
Unknown	414 (20.4)
Systolic blood pressure (mm Hg)	
<90	88 (4.3)
>90	1495 (73.7)
Unknown	447 (22.0)
Abnormal level of alertness	
Yes	470 (23.2)
No	1245 (61.3)
Unknown	315 (15.5)
Positive FAST	
Yes	65 (3.2)
No	1374 (67.7)
Unknown	591 (29.1)
Distracting painful injury	
Yes	601 (29.6)
No	1043 (51.4)
Unknown	386 (19)

Note: Data are reported as median (IQR) or *n* (%).
Abbreviation: FAST, focused assessment with sonography for trauma.

TABLE 2 Physician abdominal/pelvic injury predictions of likelihood of CSI.

Category of likelihood of CSI (physician predictions)	No Injury	Any injury on CT	CSI	Major CSI
All: 2030	1628	402	270	145
<2%: 233	210 (90.1)	23 (9.9)	13 (5.6)	5 (2.1)
2%–10%: 837	733 (87.6)	104 (12.4)	54 (6.5)	18 (2.2)
11%–20%: 411	321 (78.1)	90 (21.9)	58 (14.1)	34 (8.3)
21%–40%: 184	137 (74.5)	47 (25.5)	36 (19.6)	14 (7.6)
>40%: 365	227 (62.2)	138 (37.8)	109 (29.9)	74 (20.3)

Note: Data are reported as *n* (%). Any injury on CT = all cases with an injury on CT, including those with and those without CSI. CSI = cases with injuries on CT requiring either procedural intervention or hospital admission. Major CSI = CSI cases requiring procedural intervention, defined as a surgical procedure or blood transfusion.
Abbreviation: CSI, clinically significant injury.

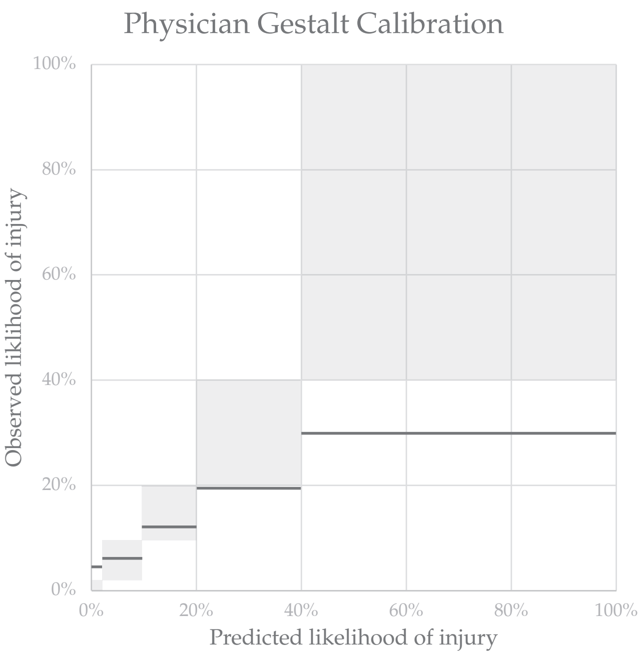

FIGURE 1 Calibration plot describing the accuracy of physician gestalt in estimating the likelihood of CSI. Five shaded boxes reflect each of the prediction categories for likelihood of CSI (<2%, 2%–10%, 11%–20%, 21%–40%, >40%). The observed likelihood of injury is displayed as a horizontal line crossing one point on the y-axis, reflecting the observed proportion of patients with CSI in this study. If the horizontal line falls within the shaded box, this indicates good calibration. If the horizontal line falls outside of the shaded box, this indicates poor calibration. CSI, clinically significant injury.

Physician population

Among the 1087 (54%) physicians for which we had physician training-level data, 859 (79%) of the evaluations were completed by resident physicians and 228 (21%) were completed by attending physicians. Of the resident physicians, 76 (9%) were Postgraduate Year (PGY)-4+, 130 (15%) were PGY-4, 236 (27%) were PGY-3, 318 (37%) were PGY-2, and 99 (12%) were PGY-1.

TABLE 3 Sensitivity and specificity of physician gestalt in predicting CSI.

Prediction threshold range	Sensitivity (95% CI)	Specificity (95% CI)
All physicians		
<2%	95.2 (91.7–97.3)	12.5 (11.0–14.2)
≤10%	75.2 (69.5–80.1)	57.0 (54.6–59.3)
≤20%	53.7 (47.6–59.7)	77.0 (75.0–79.0)
≤40%	40.4 (34.5–46.5)	85.5 (83.7–87.0)
Attending physicians		
<2%	95.7 (76.0–99.8)	17.1 (12.3–23.1)
≤10%	78.3 (55.8–91.7)	60.0 (52.9–66.7)
≤20%	47.8 (27.4–68.9)	83.9 (77.8–88.5)
≤40%	39.1 (20.5–61.2)	87.8 (82.3–91.8)
Resident physicians		
<2%	95.4 (89.9–98.1)	10.2 (8.1–12.7)
≤10%	72.5 (63.9–79.8)	55.1 (51.4–58.7)
≤20%	50.0 (41.4–58.7)	74.3 (70.9–77.4)
≤40%	40.5 (32.1–49.4)	84.1 (81.2–86.6)

Note: Responses regarding likelihood of CSI according to the following scheme: the cutoff point <2% = only the prediction category of <2%; the cutoff point ≤10% = the categories of <2% and <10%; the cutoff point ≤20% = the categories of <2%, 2%–10%, and 11%–20%; and the cutoff point ≤40% = the categories of <2%, 2%–10%, 11%–20%, and 21%–40%. Sensitivity and specificity were calculated using standard formulae (with the 2% threshold, for example, true positive = cases predicted to have ≥2% risk of CSI and found to have CSI; true negative = cases predicted to have <2% risk of CSI and found to not have CSI; false positive = cases predicted to have ≥2% risk of CSI and found to not have CSI; false negative or missed cases = cases predicted to have <2% risk of CSI and found to have CSI).

Abbreviation: CSI, clinically significant injury.

Calibration analysis (Figure 1) shows that physician gestalt was moderately well calibrated in the midranges of prediction (2%–10%, 11%–20%, and 21%–40% likelihood of CSI), but poorly calibrated at the extremes of prediction (<2% and > 40%). For the <2% cutoff of likelihood of CSI, the total group gestalt miss rate (prevalence of CSI) was 5.6% (95% confidence interval [CI], 3.3%–9.3%), and the sensitivity was 95.2% (95% CI, 91.7%–97.3%). For the 0% to 10% cutoff of likelihood of CSI, the total group gestalt miss rate was 6.3% (95% CI, 5.0%–7.9%), and the sensitivity was 75.2% (95% CI 69.5%–80.1%; Table 3). Stratified by training level (Figure 2), resident and attending physician gestalt sensitivity was similar: for a 2% cutoff for low risk of injury, sensitivity was 95.7% (76.0%–99.8%) for attending physicians and 95.4% (89.9%–98.1%) for resident physicians; for a 10% cutoff, sensitivity was 78.3% (55.8%–91.7%) for attending physicians and 72.5% (63.9%–79.8%) for resident physicians.

The gestalt AUROC was 0.699 (0.679–0.719) for all physicians, 0.732 (0.670–0.789) for attending physicians, and 0.668 (0.636–0.699) for resident physicians (Figure 3). When comparing physician predictions across estimates, the ILR associated with CSI increased with increasing suspicion of injury, with the ILR for <2% likelihood of injury at 0.37, the ILR for 2%–10% at 0.44, the ILR for 11%–20% at 1.1, the ILR for 21%–40% at 1.6, and the ILR for >40% at 2.9.

When physicians were asked to select their rationale for proceeding with CTs despite low estimated risk of injury (defined as <10% likelihood of CSI), the most common justifications were severe mechanism of injury (40.1%), physical examination unreliable because of intoxication or altered mental status (19%), and trauma protocol (14.7%; Table 4), When compared to patients with a 2%–10% estimated risk of CSI, a higher proportion of those with a <2% risk of CSI received CTs due to trauma protocol (22.2% vs. 12.6% [difference 9.8%, 95% CI 4.3%–15.9%]) and medicolegal concerns (10.9% vs. 6.3% [difference 4.6%, 95% CI 0.8%–9.6%]).

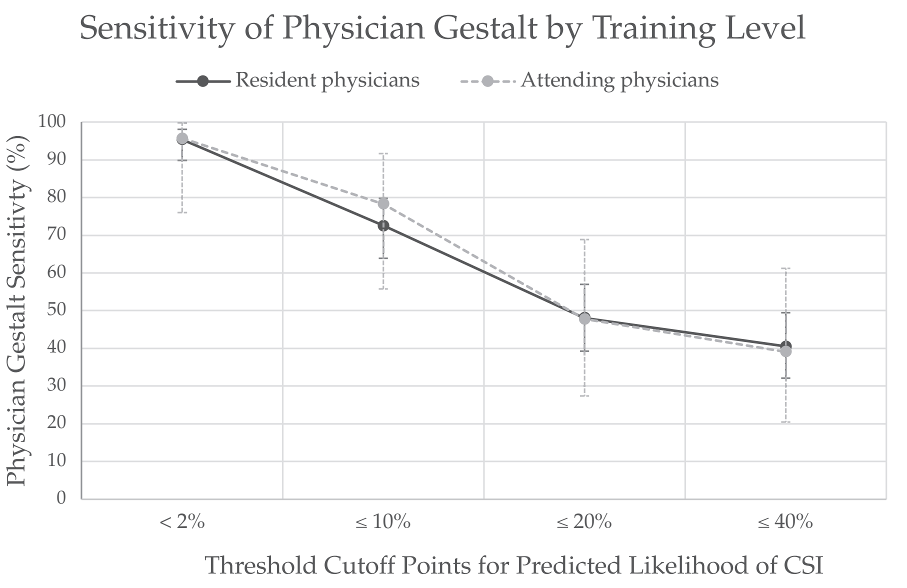

FIGURE 2 Changes in physician gestalt sensitivity as a function of distinct cutoff points of likelihood of clinically significant injury according to the following scheme: the cutoff point <2% = only the prediction category of <2%; the cutoff point ≤10% = the categories of <2% and <10%; the cutoff point ≤20% = the categories of <2%, 2%–10%, and 11%–20%; and the cutoff point ≤40% = the categories of <2%, 2%–10%, 11%–20%, and 21%–40%.

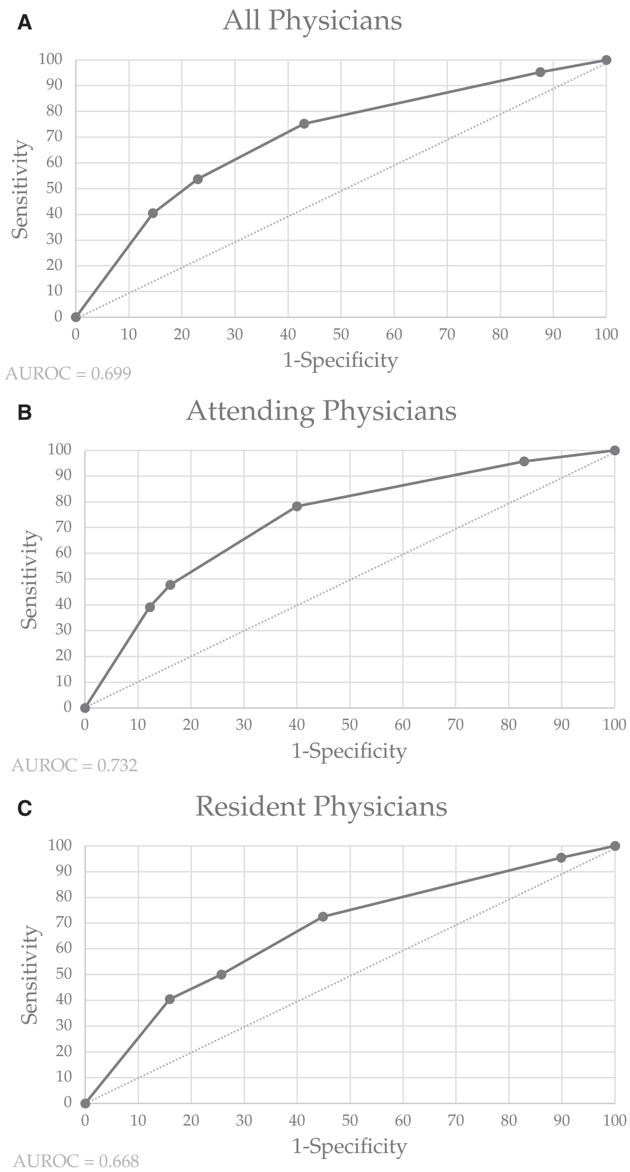

FIGURE 3 Receiver operating characteristic (ROC) curve for overall clinician accuracy in detecting CSI by training level. The gestalt area under the receiver operating characteristic (AUROC) for all physicians (A) is 0.699 (0.679–0.719), while the AUROC estimates stratified by training level are 0.732 (0.670–0.789) for attending physicians (B) and 0.668 (0.636–0.699) for resident physicians (C).

TABLE 4 Physician reasons for CT use in trauma patients perceived to be at low risk for injury.

Reason(s) for ordering CT in low-risk cases ($n = 1058$)	<2% estimated risk	2%–10% estimated risk
Severe mechanism of injury, 426 (40.3%)	81 (35.2)	345 (41.7)
Physical examination unreliable, 200 (18.9%)	35 (15.2)	165 (19.9)
Trauma protocol, 155 (14.6%)	51 (22.2)	104 (12.6)
Other, 120 (11.3%)	25 (10.9)	95 (11.5)
Medicolegal concern, 77 (7.3%)	25 (10.9)	52 (6.3)
Level of risk still important to diagnose, 62 (5.9%)	7 (3.0)	55 (6.6)
Need to diagnose all injuries even if not significant, 18 (1.7%)	6 (2.6)	12 (1.4)

Note: Data are reported as *n* (%). Physicians were asked to select the reason they ordered an abdominal/pelvic CT scan when they believed patients had a <2% or a 2%–10% likelihood of having a CSI. This table shows their reasons for AP CT use in these low-risk cases stratified by risk group.

Abbreviation: CSI, clinically significant injury.

DISCUSSION

We found a physician gestalt AUROC of 0.7, indicating moderate physician accuracy in discriminating between patients who had CSI and did not have CSI. The proportion of patients with abdominal or pelvic CSI on CT increased as physician suspicion for risk of CSI increased, suggesting a generally adequate calibration trend. However, while physician gestalt was well calibrated for the midranges of predicted risk of CSI, it was poorly calibrated at the extremes. This suggests that physicians overestimated the risk of CSI among patients predicted to be at high risk and underestimated the risk of CSI among patients predicted to be at very low risk.

While overestimation of the risk of CSI is unlikely to lead to missed injury (all patients with a high estimated risk would receive CT), underestimation of the risk of CSI in the very low risk prediction group could potentially lead to missed injury if those patients did not receive CT. We found a CSI prevalence rate of 5.6% in the <2% risk group that is likely well above the threshold that the majority of physicians would consider an acceptable miss rate. Similarly, the 95.2% sensitivity (with a lower margin of 91.7%) of the <2% gestalt cutoff is too low of a sensitivity for widespread acceptance.

Most of our evaluators were emergency medicine resident physicians and no trauma attending surgeons were included. Although more experienced physicians may be more accurate predictors, attending physician predictions at the <2% cutoff had similar sensitivity (95.7% vs. 95.4%) and an unacceptably high miss rate of 4.3%. Overall, our findings suggest that even attending physician gestalt fails to demonstrate sufficient sensitivity in detecting abdominal or pelvic CSI and should not be relied upon as a sole criterion to determine cases in which it is safe to forego order abdominal/pelvic CT in adult blunt trauma patients.

Physicians provided multiple reasons for ordering CT in perceived low-risk cases. Trauma protocols (reflexive scanning) played a significant role in these decisions, as did severe mechanism of injury, unreliable physical examination, and medical–legal concerns. Our research should not be viewed as justification for an even greater amount of reflexive whole-body CT (WBCT) in blunt trauma evaluation. Current trauma CT is already, for the most part, a low-yield, high-cost and high-radiation-exposure practice.[29–31] Our findings instead argue for the development of a high-sensitivity, low-missed-injury-rate CDR to safely guide selective abdominal/pelvic CT in adult blunt trauma patients. When added to current adult trauma CDRs for selective CT of the head, neck, and chest, a CDR that provides guidance for selective CT in the abdomen and pelvis anatomic region may fulfill the quest for selective head-to-pelvis CT and reduce reflexive WBCT in adult blunt trauma patients.

Acceptable miss rates in trauma evaluation vary considerably, especially when comparing the views of different specialties.[20] Some physicians may even espouse the view that no misses are acceptable and that CT should be universally utilized in adult blunt trauma evaluation. However, multiple investigators have shown that this overdiagnosis approach for a zero-miss rate is associated with great costs, low yields, and potential for cancer induction in young trauma populations.[29–31]

A number of other factors, especially patient age and sex, may influence decisions to order (or not order) CT in trauma. For example, providers may consider that the radiation risk in a 90-year-old is negligible and so that liberal, nonselective CT use in this patient is justified. Patient race and ethnicity have also been shown to influence imaging decisions. Considering these other rationale in future research may provide additional context regarding provider gestalt and application of an abdominal trauma CT decision rule.

LIMITATIONS

Although our verification bias assessment on a set of patients who did not receive CT revealed no CSI, we did not obtain physician gestalt assessment on these patients. This population would likely have had more true-negative gestalt assessments and resultant increased specificity. Similarly, this study does not capture cases in which patients were too unstable to receive abdominal/pelvic CT and likely have higher rates of CSI (more true positives) and greater physician gestalt sensitivity. Nevertheless, this spectrum bias would not likely affect the most important gestalt screening performance subgroup characteristics—sensitivity and missed injury rates in the <2% likelihood of injury group. Furthermore, our study method of including only those who received CT reflects real-world capturing of the true population that might benefit from efforts to decrease overuse of CT in blunt trauma, i.e., those that are currently getting CT.

Our use of categorical variables (ranges of physician gestalt) over continuous (0%–100% risk) may have induced menu effects regarding gestalt, with resultant loss of precision in AUROC and calibration estimates. Realizing that we were studying CDR development and CT ordering in low-risk cases, physicians may have been subject to a Hawthorne effect and altered their choices on the survey instrument, instead of giving their true gestalt prediction.

A major limitation is the study setting consisting primarily of urban, academic trauma centers. Our findings may not apply to dissimilar EDs and trauma centers. Furthermore, we had incomplete data on training level with overrepresentation of less experienced resident physicians. However, point estimates of AUROC and gestalt sensitivity at the <2% and 2%–10% intervals were similar between resident and attending physicians.

Surgeries for nonabdominal/pelvic injuries were not considered CSI, nor were admissions when there was no abdominal or pelvis injury. Nevertheless, some patients with abdominal injury on CT may have been admitted for other reasons than those particular injuries.

CONCLUSIONS

Although physician gestalt predictions regarding abdominal and pelvic clinically significant injuries in adult blunt trauma patients show good overall correlation with true clinically significant injury, the potential missed injury rate and sensitivity at the extreme end of the low-risk prediction category is insufficient to rule out injury, suggesting that physician gestalt may be inadequate to guide selective abdominal/pelvic computed tomography. We have thus established one criterion supporting the need for the development of a clinical decision rule to guide selective abdominal/pelvic computed tomography use in adult blunt trauma—the need for a tool that has a lower miss rate and higher sensitivity than gestalt.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Juan Carlos Montoy https://orcid.org/0000-0001-7438-0243
Robert M. Rodriguez https://orcid.org/0000-0003-1354-1773

REFERENCES

1. Pal JD, Victorino GP. Defining the role of computed tomography in blunt abdominal trauma. *Arch Surg*. 2002;137(9):1029-1033.
2. Rodriguez RM, Baumann BM, Raja AS, et al. Diagnostic yields, charges, and radiation dose of chest imaging in blunt trauma evaluation. *Acad Emer Med*. 2014;6:644-650.
3. Gupta M, Schriger DL, Hiatt JR, et al. Selective use of computer tomography compared with routine whole-body imaging in patients with blunt trauma. *Ann Emerg Med*. 2011;58(5):407-416.
4. Murphy SP, Hawthorne N, Rodriguez RM. Low yield of clinically significant injury with head-to-pelvis CT in blunt trauma evaluation. *J Emerg Med*. 2017;53(6):865-870.
5. Finnerty NM, Rodriguez RM, Carpenter CR, et al. Clinical decision rules for diagnostic imaging in the emergency department: a research agenda. *Acad Emerg Med*. 2015;22(12):1406-1416.
6. Carpenter CR, Raja AS, Brown MD. Overtesting and the downstream consequences of overtreatment: implications of "preventing Overdiagnosis" for emergency medicine. *Acad Emerg Med*. 2015;22(12):1484-1492. doi:10.1111/acem.12820
7. Mower W, Gupta M, Rodriguez RM, Hendey GW. Validation of the sensitivity of the National Emergency X-Radiography Utilization Study (NEXUS) head computed tomographic (CT) decision instrument for selective imaging of blunt head injury patients: an observational study. *PLoS Med*. 2017;14(7):e1002313.
8. Stiell IG, Clement CM, Rowe BH, et al. Comparison of the Canadian CT head rule and the New Orleans criteria in patients with minor head injury. *Jama*. 2005;294(12):511-518.

9. Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. *Jama*. 2001;286(15):1841-1848.
10. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med*. 1999;33(4):437-447.
11. Hoffman JR, Mower WR, Wolfson AB, Todd KH, Zucker MI. Validity of a set of clinical criteria to rule out injury to the cervical spine in patients with blunt trauma. National Emergency X-radiography utilization study group. *N Engl J Med*. 2000;343(2):94-99.
12. Rodriguez RM, Anglin D, Langdorf MI, et al. NEXUS chest: validation of a decision instrument for selective chest imaging in blunt trauma. *J Trauma*. 2013;71(3):549-553.
13. Rodriguez RM, Langdorf MI, Nishijima D, et al. Derivation and validation of two decision instruments for selective chest CT in blunt trauma: a multicenter prospective observational study (NEXUS chest CT). *PLoS Med*. 2015;12(10):e1001883.
14. Rodriguez RM, Hendey GW, Mower W, et al. Derivation of a decision instrument for selective chest radiography in blunt trauma. *J Trauma – Inj Infect Crit Care*. 2011;71(3):549-553.
15. Holmes JF, Wisner DH, McGahan JP, Mower WR, Kuppermann N. Clinical prediction rules for identifying adults at very low risk for intra-abdominal injuries after blunt trauma. *Ann Emerg Med*. 2009;54(4):575-584.
16. Sharples A, Brohi K. Can clinical prediction tools predict the need for computed tomography in blunt abdominal trauma? A systematic review. *Inj Int J Care Injured*. 2016;47:1811-1818.
17. Schriger DL, Elder JW, Cooper RJ. Structured clinical decision aids are seldom compared with subjective physician judgment, and are seldom superior. *Ann Emerg Med*. 2017;70(3):338-344.
18. Raja AS, Rodriguez RM, Gupta M, et al. Developing a decision instrument to guide abdominal-pelvic imaging of blunt trauma patients: methodology and protocol of the NEXUS abdominal-pelvic imaging study. *PLoS One*. 2022;17(7):e0271070.
19. Gilbert EH, Lowenstein SE, Koziol-McLain J, Barta DC, Steiner J. Chart reviews In emergency medicine research: where are the methods? *Ann Emerg Med*. 1996;27:305-308.
20. Gupta M, Schriger DL, Hiatt JR, et al. Selective use of computed tomography compared with routine whole body imaging in patients with blunt trauma. *Ann Emerg Med*. 2011;58:407-416.
21. Mower WR, Hoffman JR, Herbert M, et al. Developing a decision instrument to guide computed tomographic imaging of blunt head injury patients. *J Trauma*. 2005;59(4):954-959.
22. Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT head rule for patients with minor head injury. *Lancet*. 2001;357(9266):1391-1396.
23. Kohn MA. Key concepts in clinical epidemiology: reporting on the accuracy of continuous tests. *J Clin Epidemiol*. 2022;141:157-160.
24. Newman TB, Kohn MA. *Chapter 3: Multi-Level and Continuous Tests, Evidence-Based Diagnosis: an Introduction to Clinical Epidemiology*. Cambridge University Press; 2020:47-74.
25. Brown MD, Reeves MJ. Evidence-based emergency medicine/skills for evidence-based emergency care: interval likelihood ratios: another advantage for the evidence-based diagnostician. *Ann Emerg Med*. 2003;42:292-297.
26. Worster A, Carpenter C. A brief note about likelihood ratios. *Can J Emerg Med*. 2008;10(5):441-442.
27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
28. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-349.
29. Murphy SP, Hawthorne N, Haase D, Chiku C, Wen J, Rodriguez RM. Low yield of clinically significant injury with head-to-pelvis computed tomography in blunt trauma evaluation. *J Emerg Med*. 2017;53(6):865-870.
30. Rodriguez RM, Hawthorne N, Murphy SP, et al. Blunt trauma abdominal and pelvic computed tomography has low yield for injuries in more than one anatomic region. *West J Emerg Med*. 2018;19(5):768-773.
31. Rodriguez RM, Baumann BM, Raja AS, et al. Diagnostic yields, charges, and radiation dose of chest imaging in blunt trauma evaluations. *Acad Emerg Med*. 2014;21(6):644-650.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rodriguez NM, Mower WR, Raja AS, et al. Accuracy of physician gestalt in prediction of significant abdominal and pelvic injury in adult blunt trauma patients. *Acad Emerg Med*. 2023;00:1-8. doi:10.1111/acem.14768