Physiologic Predictors of Severe Injury: Systematic Review

Evidence Summary

Background

Unintentional injury is the fourth leading cause of death in the United States, the leading cause for people 1 to 44 years of age, and the reason for millions of emergency department (ED) visits. Trauma is the primary reason emergency medical services (EMS) transport people to the hospital. Out-of-hospital care includes the early interventions and life support needed to prevent immediate deterioration and to secure vital functions after injury.

In the United States, out-of-hospital trauma care is delivered predominately by EMS personnel. EMS personnel can include individuals with different levels of training and certification, including emergency medical responder, emergency medical technician (EMT), advanced EMT, and paramedic. EMS personnel assess patients in environments that are often chaotic and sometimes dangerous. Some out-of-hospital care decisions can be made based on observable characteristics of the injury (e.g., a crush injury or amputation), but other injuries require additional assessment. Triage guidelines and protocols include the assessment of circulatory and respiratory compromise as essential components of the triage process to identify high-risk trauma patients and inform transport destination decisions.

In the current guidelines, field triage of injured patients consists of four steps:

Purpose of Review

To summarize evaluations of physiologic measures that can be used by emergency medical services personnel to identify patients at high risk of serious injury and inform decisions about the level of trauma care needed.

Key Messages

- Studies examined individual measures and combinations for trauma triage, including systolic blood pressure, heart rate, shock index, lactate, base deficit, respiratory rate, oxygen saturation, and need for airway support.
- Included measures have:
  - Low sensitivities, so normal values on the physiologic measure (a negative test) cannot be used with confidence to determine that patients are not seriously injured.
  - High specificities, meaning abnormal values on the physiologic measure (positive test) are unlikely in patients not seriously injured.
- Combinations of physiologic measures with measures of consciousness may perform better than physiologic measures alone, but feasibility and reliability of performance in the field are significant challenges.
- Measures perform less well in children and older people. Changes in cut-points for these age groups may improve performance but have not yet been rigorously evaluated.
The scope is limited to considering how well the physiologic measures predict serious injury in trauma patients evaluated by EMS personnel. The assumption is that being able to identify seriously injured patients will inform triage and transport decisions, and these decisions will impact care for the injury, which will affect outcomes. These assumed relationships and the ultimate patient outcomes are important, but not part of the review.

We included measures in the review that can be obtained by standard medical equipment or devices specially designed for field assessment or monitoring. We included ED studies and their data because conducting studies in the field is difficult and the body of evidence based on out-of-hospital data is limited for some measures. However, the data were analyzed and the results are presented separately for out-of-hospital and ED data.

The Key Questions for the review were:

**Key Question 1:** For patients with known or suspected trauma who are treated out-of-hospital by EMS personnel, what is the predictive utility of measures of circulatory compromise or derivative measures (e.g., the shock index) for predicting serious injury requiring transport to the highest level trauma center available?

1a. How does the predictive utility of the studied measures of circulatory compromise vary across age groups (e.g., children or the elderly)? Specifically, what values for the different age ranges are supported by the evidence?

**Key Question 2:** For patients with known or suspected trauma who are treated out-of-hospital by EMS personnel, what is the predictive utility of measures of respiratory compromise for predicting serious injury requiring transport to the highest level trauma center available?

2a. How does the predictive utility of the studied measures of respiratory compromise vary across age groups (e.g., children or the elderly)? Specifically, what values for the different age ranges are supported by the evidence?

**Key Question 3:** For patients with known or suspected trauma who are treated out-of-hospital by EMS personnel, what is the evidence that scales combining (a) measures of respiratory and circulatory compromise or (b) measures of respiratory and/or circulatory compromise together with measurement of altered levels of consciousness (as defined by Glasgow Coma Scale or its components) can predict the need for transport to a trauma center?

3a. How does the predictive utility of combinations of measures vary across age groups (e.g., children or the elderly)? Specifically, what values for the different age ranges are supported by the evidence?
The scope and Key Questions for this topic were initially developed by the Agency for Healthcare Research and Quality in conjunction with the sponsoring partner agency, the National Highway Transportation Safety Administration.

To identify studies we searched Ovid MEDLINE®, CINAHL®, and the Cochrane Databases from 1996 through August 2017. We included studies of individual measures and measures that combined circulatory, respiratory, and level of consciousness assessment. For studies that met inclusion criteria, the key characteristics and results were abstracted into evidence tables that provide the basis for the description and synthesis of this body of literature. Studies were evaluated for risk of bias using the Quality in Prognosis Studies (QUIPS) tool. The complete evidence tables and the risk of bias ratings for each included study are available in Appendixes D and F of the full report.

The end points or outcomes of interest were the predictive utility of the measures. We included different approaches to assessing predictive utility: (1) measures of diagnostic accuracy (e.g., sensitivity and specificity,) and (2) discrimination (e.g., area under the receiver operating characteristic curve [AUROC]). Studies that provided only descriptive information, unadjusted risk estimates, or assessments of continuous variables (e.g., correlations or tests of differences in means) were excluded unless data were provided that could be used to calculate included outcomes. For this review our focus was on the predictive utility for identifying patients at high risk of being seriously injured. We defined seriously injured broadly and used a range of indicators of serious injury, including resource use (e.g., blood transfusion, intensive care unit [ICU] admission, and life-saving interventions), measures of anatomic injury severity measures (e.g., the Injury Severity Score [ISS]), and mortality, or combinations of any of these).

We conducted quantitative synthesis (i.e., meta-analysis) of diagnostic measures when there were adequate data from included studies. In cases with few studies, lack of data, or when data were only available as adjusted risk estimates from multivariate analyses, the range of the results and qualitative summaries were provided.

For meta-analyses we used a bivariate logistic mixed effects model to analyze sensitivity and specificity, incorporating the correlation between sensitivity and specificity. We assumed random effects across studies for sensitivity and specificity, and heterogeneity among the studies was measured based on the random effect variance. We also assessed statistical heterogeneity using the standard χ² test and F statistic. We calculated positive likelihood ratio (LR+) and negative likelihood ratio (LR-) using the summarized sensitivity and specificity. Analyses were stratified by different cutoff points when necessary to generate clinically meaningful combined estimates. Similarly, we performed random effects meta-analysis to calculate the combined AUROC using the profile likelihood method, which incorporates the uncertainty related to estimating between-study heterogeneity into account.

All analyses were conducted using Stata/IC 13.1 (StataCorp LP, College Station, TX), except for the bivariate logistic mixed effects model, for which SAS 9.4 (SAS Institute Inc., Cary, NC) was used.

The review team and Technical Expert Panel included experts who have conducted and published research in this field. In order to avoid bias, or the appearance of bias, we took the following steps: (1) authors were not involved in any decisions about including or excluding their own work, (2) to the extent it was feasible, reviewers were blinded to authors during title and abstract review so that the other team members/reviewers were not biased in favor of colleagues, (3) for full-text review, no one was assigned to review research they contributed to, and (4) team members and experts did not rate the risk of bias or abstract data from studies to which they contributed.

Results

We identified and included 138 articles reporting results of 134 studies: 90 evaluated measures of circulatory compromise, 39 respiratory compromise, and 64 included combination measures. Over two-thirds (96) of the studies were retrospective and the remainder (42) were prospective. A total of 25 studies used data from multi-site registries, 65 studies used administrative or registry data from a single site, and 48 studies engaged in primary data collection for the study. Three-quarters (103) of the studies were assessed as moderate risk of bias and the others were rated low risk of bias (10) or high risk of bias (25 studies). The concerns about bias were primarily related to three domains: study participation (e.g., concerns about sampling or recruitment), attrition (e.g., lost to followup), and lack of control for confounding factors that were not adequately addressed in the study design.

Studies used different indicators for serious injury, and often assessed more than one indicator. We grouped the
indicators into three categories: (1) resource utilization, which includes lists of life-saving interventions, surgery, transfusion, ICU admission, or the published consensus-base criterion standard; (2) ratings of anatomic injury severity such as the ISS or types of injury or diagnosis such as traumatic brain injury; and (3) mortality. Resource utilization was the most common indicator (110 studies). A similar number of studies reported on the relationship between the measures and mortality (95 studies), while injury severity or type was less common (19 studies).

Tables A, B, and C consolidate the key estimates of the predictive utility of each identified measure. Sensitivity, specificity, and AUROC values for out-of-hospital and ED measurements are provided. When we were able to pool data, the pooled estimates are given in bold; when data were not pooled, the range of values from the included studies are given in italics. Additional information, such as the number of patients in the included studies and 95% confidence intervals for the estimates, are available in the figures and tables in the full report.

Table A. Key Question 1 results: overview of predictive utility of circulatory measures for serious injury* by setting

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (SOE)</td>
<td>Specificity (SOE)</td>
<td>AUROC (SOE)</td>
<td>AUROC (SOE)</td>
</tr>
<tr>
<td></td>
<td>Number of Studies</td>
<td></td>
<td>Number of Studies</td>
<td>Number of Studies</td>
</tr>
<tr>
<td>Systolic Blood Pressure</td>
<td>SBP &lt;90 mmHg</td>
<td>SBP &lt;90 mmHg</td>
<td>0.67 (SOE: Moderate)</td>
<td>0.64 (SOE: Moderate)</td>
</tr>
<tr>
<td></td>
<td>Sen: 19% (SOE: Low)</td>
<td>Sen: 18% (SOE: Low)</td>
<td>9 studies</td>
<td>12 studies (in 13 articles)</td>
</tr>
<tr>
<td></td>
<td>Sp: 95% (SOE: Moderate)</td>
<td>Sp: 97% (SOE: Moderate)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 studies</td>
<td>9 studies (in 10 articles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SBP higher thresholds (&lt;100, 110, or 120 mmHg)</td>
<td>SBP higher thresholds (&lt;100, 110, or 120 mmHg)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sen: 35% (SOE: Low)</td>
<td>Sen: 35% (SOE: Low)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sp: 88% (SOE: Low)</td>
<td>Sp: 89% (SOE: Moderate)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 studies</td>
<td>4 studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Rate</td>
<td>HR &gt;110 bpm</td>
<td>HR &gt;110 bpm</td>
<td>0.67 (SOE: Low)</td>
<td>0.66 (SOE: Moderate)</td>
</tr>
<tr>
<td></td>
<td>Sen: 28% (SOE: Low)</td>
<td>Sen: 29% (SOE: Moderate)</td>
<td>5 studies</td>
<td>9 studies (in 10 articles)</td>
</tr>
<tr>
<td></td>
<td>Sp: 85% (SOE: Low)</td>
<td>Sp: 93% (SOE: Moderate)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 studies</td>
<td>5 studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shock Index</td>
<td>SI &gt;0.9 or &gt;1</td>
<td>SI &gt;0.9 or &gt;1</td>
<td>0.72 (SOE: Low)</td>
<td>0.71 (SOE: Moderate)</td>
</tr>
<tr>
<td></td>
<td>Sen: 37% (SOE: Low)</td>
<td>Sen: 40% (SOE: Low)</td>
<td>7 studies</td>
<td>11 studies (in 12 articles)</td>
</tr>
<tr>
<td></td>
<td>Sp: 85% (SOE: Low)</td>
<td>Sp: 93% (SOE: Moderate)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 studies</td>
<td>11 studies (in 12 articles)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*SOE: Sensitivity, Specificity, and AUROC values for out-of-hospital and ED measurements are provided. When we were able to pool data, the pooled estimates are given in bold; when data were not pooled, the range of values from the included studies are given in italics. Additional information, such as the number of patients in the included studies and 95% confidence intervals for the estimates, are available in the figures and tables in the full report.
### Table A. Key Question 1 results: overview of predictive utility of circulatory measures for serious injury* by setting (continued)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Studies</td>
<td>Number of Studies</td>
<td></td>
<td>Number of Studies</td>
</tr>
</tbody>
</table>
| **Lactate**              | Lactate >2 or 2.5 mmol/L  
Sen: 74% (SOE: Low)  
Sp: 62% (SOE: Moderate)  
3 studies16,71,72  
Lactate >4 mmol/L  
Sen: 23% (SOE: Insufficient)  
Sp: 93% (SOE: Insufficient)  
1 study75 | Lactate >2 or 2.5 mmol/L  
Sen: 74% (SOE: Low)  
Sp: 52% (SOE: Low)  
9 studies (in 10 articles)35,36,43,73-79  
Lactate >4 mmol/L  
Sen: 50% (SOE: Low)  
Sp: 86% (SOE: Moderate)  
9 studies39,43,58,61,73,75,77,80,81 | **0.77** (SOE: Low)  
2 studies16,72 | **0.68** (SOE: Moderate)  
14 studies (in 15 articles)35,36,43,58,73,74,76,78-80,82-86 |
| **Base Deficit**         | None                                                 | Sen: 19 to 59% (SOE: Low)  
Sp: 59 to 98% (SOE: Moderate)  
9 studies (in 10 articles)32,35-37,61,75,80,83,87 | None | **0.67 to 0.90** (SOE: Moderate)  
12 studies (in 13 articles)32,35,36,49,52,65,73,77,80,82,83,85,86 |
| **Heart Rate Variability/Heart Rate Complexity** | Sen: 80 to 90% (SOE: Low)  
Sp: 67 to 100% (SOE: Low)  
2 studies44,88 | None | **0.60 to 0.95** (SOE: Low)  
7 studies44,46,88-92 | **0.67 to 0.68** (SOE: Insufficient)  
1 study93 |

AUROC = area under the receiver operating characteristic curve; bpm = beats per minute; HR = heart rate; SBP = systolic blood pressure; Sen = sensitivity; SI = shock index; SOE = strength of evidence; Sp = specificity

Note: Bold font = data from pooled estimates; italic font = range from unpooled studies

*Serious injury includes resource use (e.g., blood transfusion, intensive care unit [ICU] admission, and life-saving interventions) and injury severity measures (e.g., the Injury Severity Score [ISS], mortality, or combinations of any of these)
Table B. Key Question 2 results: overview of predictive utility of respiratory measures for serious injury* by setting

<table>
<thead>
<tr>
<th>Measure</th>
<th>Out-of-Hospital: Sensitivity (SOE)</th>
<th>Number of Studies</th>
<th>Emergency Department: Sensitivity (SOE)</th>
<th>Number of Studies</th>
<th>Out-of-Hospital: AUROC (SOE)</th>
<th>Number of Studies</th>
<th>Emergency Department: AUROC (SOE)</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory Rate</td>
<td>RR &lt;10 or &gt;29</td>
<td>6 studies</td>
<td>RR &lt;10 or &gt;29</td>
<td>4 studies</td>
<td>0.70 (SOE: Low)</td>
<td>3 studies</td>
<td>0.61 (SOE: Moderate)</td>
<td>3 studies</td>
</tr>
<tr>
<td></td>
<td>Sen: 13% (SOE: Low)</td>
<td></td>
<td>Sen: 27% (SOE: Moderate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sp: 96% (SOE: Low)</td>
<td></td>
<td>Sp: 95% (SOE: Moderate)</td>
<td></td>
<td>0.61 (SOE: Moderate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O2 Saturation</td>
<td>Sen: 13 to 99% (SOE: Low)</td>
<td>3 studies</td>
<td>Sen: 25 to 100% (SOE: Low)</td>
<td>2 studies</td>
<td>0.53 to 0.76 (SOE: Low)</td>
<td>3 studies</td>
<td>0.61 to 0.76 (SOE: Low)</td>
<td>2 studies</td>
</tr>
<tr>
<td></td>
<td>Sp: 85 to 99% (SOE: Low)</td>
<td></td>
<td>Sp: 39 to 94% (SOE: Low)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airway Support</td>
<td>Sen: 8 to 53% (SOE: Low)</td>
<td>4 studies (in 5 articles)</td>
<td>Sen: 32 to 57% (SOE: Low)</td>
<td>3 studies</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Sp: 61 to 100% (SOE: Low)</td>
<td></td>
<td>Sp: 85 to 96% (SOE: Low)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AUROC = area under the receiver operating characteristic curve; O2 = oxygen; RR = respiratory rate; Sen = sensitivity; SOE = strength of evidence; Sp = specificity

Note: Bold font = data from pooled estimates; italic font = range from unpooled studies

*Serious injury includes resource use (e.g., blood transfusion, intensive care unit [ICU] admission, and life-saving interventions) and injury severity measures (e.g., the Injury Severity Score [ISS], mortality, or combinations of any of these)
Our analysis of individual measures of circulatory and respiratory compromise (Key Question 1 and Key Question 2) included pooled analyses of SBP, shock index (SI), heart rate (HR), lactate, and respiratory rate (RR), and qualitative summaries of studies of heart rate variability/heart rate complexity, base deficit, and oxygen saturation. Other measures that were the subject of one or two studies were included but not synthesized.

Most of the strength of evidence assessments were “low” due to inconsistency in results across studies and imprecise estimates, though in some cases study limitations also contributed to the low rating. There were a few “moderate” ratings for measures where there were more studies and subjects and the results were consistent and the estimates more precise. There were no “high” strength of evidence ratings as we are not confident that the results will not change based on future studies of physiologic measures that are larger, better, and purposefully designed to study trauma triage.

Across all the measures, the pooled AUROC values we calculated generally fell into the ranges considered poor (0.60 to 0.69) or fair (0.70 to 0.79). Focusing on data collected out-of-hospital, the lowest pooled AUROCs were for SBP (0.67) and HR (0.67). The AUROCs were in the fair range for SI (0.72), lactate (0.77), and RR (0.70). We also pooled data to estimate sensitivity and specificity results for blood pressure and lactate at different thresholds (<90 and <100 mmHg for blood pressure and >2.0 or >4.0 mmol/L for lactate). Using the higher threshold of <100 mmHg for SBP did increase sensitivity compared the lower threshold of <90 mmHg (from 19% at the lower threshold to 35% at the higher threshold for out-of-hospital studies, and from 18% to 35% for ED studies) with a moderate decrease in specificity (from 95% at the lower threshold to 88% at the higher threshold for out-of-hospital, and from 97% to 89% in ED). For lactate, defining abnormal with a more extreme value of >4.0 mmol/L decreased sensitivity and increased specificity. The changes were more extreme in the out-of-hospital data (sensitivity was

### Table C. Key Question 3 results: overview of predictive utility of combination of circulatory, respiratory, and level of consciousness measures for serious injury* by setting

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Revised Trauma Score and Revised Trauma Score for Triage</td>
<td>RTS &lt;7.5, T-RTS &lt;12, Sen: 95 to 96% (SOE: Insufficient), Sp: 38 to 42% (SOE: Insufficient) 1 study (in 2 articles)&lt;sup&gt;98,99&lt;/sup&gt;</td>
<td>RTS &lt;5.68 or &lt;5.97, T-RTS &lt;8 or &lt;12, Sen: 19 to 84% (SOE: Low), Sp: 64 to 100% (SOE: Low) 6 studies&lt;sup&gt;28,34,48,61,101,102&lt;/sup&gt;</td>
<td>0.57 for Resource use (SOE: Low) 0.89 for Mortality (SOE: Low) 3 studies (in 4 articles)&lt;sup&gt;28,45,98,99&lt;/sup&gt;</td>
<td>0.88 for Mortality (SOE: Low) 7 studies&lt;sup&gt;48,69,70,83,101,103,104&lt;/sup&gt;</td>
</tr>
<tr>
<td>Glasgow Coma Scale, Age, and Arterial Pressure (GAP)</td>
<td>None</td>
<td>Sen: 75 to 98% (SOE: Low), Sp: 57 to 91% (SOE: Low) 2 studies&lt;sup&gt;101,105&lt;/sup&gt;</td>
<td>None</td>
<td>0.96 for both Mortality and Early Mortality (SOE: Moderate) 3 studies&lt;sup&gt;101,103,105&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

AUROC = area under the receiver operating characteristic curve; RTS = Revised Trauma Score; Sen = sensitivity; SOE = strength of evidence; Sp = specificity; T-RTS = Revised Trauma Score for Triage

Note: Bold font = data from pooled estimates; italic font = range from unpooled studies

*Serious injury includes resource use (e.g., blood transfusion, intensive care unit [ICU] admission, and life-saving interventions) and injury severity measures (e.g., the Injury Severity Score [ISS], mortality, or combinations of any of these)
74% for lactate >2.0 mmol/L and 23% for >4.0 mmol/L; specificity increased from 62% to 93%) than in the ED data (sensitivity was 74% for lactate >2.0 mmol/L and 50% at >4.0 mmol/L; specificity increased from 52% to 86%). However, the out-of-hospital estimates are from fewer studies and patients and the estimates are less stable and less precise.

We identified numerous combination measures (Key Question 3); however, most were analyzed in only one or two articles. The exception was the Revised Trauma Score (RTS) and variations on this score. Given that the formula for RTS cannot be calculated quickly without a calculator or app, some studies suggested and evaluated revisions that simplified the calculation. The produced minor decreases in AUROCs (from 0.90 for the RTS to 0.8899 for the simpler version, or from 0.75 to 0.7498). Another combination of potential interest is Glasgow Coma Scale, age, and arterial pressure (GAP), which combines the Glasgow Coma Scale (GCS) score, adds points if the patient is over 60 years of age (age is the A in GAP), and scores SBP as above or below 120 mmHg. Adding age means this is not purely a physiologic measure, but it is included as it is simple and there is small but growing evidence of its performance. While the data we reviewed is from a smaller number of studies than are available for other measures, and the measures were all collected in the ED, these initial indications are that the GAP performs well. Reported AUROCs were over 0.9 and sensitivities ranged from 75 to 98 percent and specificities from 57 to 91 percent across different indicators of serious injury.

We examined the utility of the measures or specific thresholds for pediatric and older trauma patients. The included studies that assessed measures in pediatric patients reported that the standard thresholds used for adults for SBP and base deficit resulted in low sensitivities in children. Lactate >2.0 resulted in higher sensitivities compared to the other measures, but the values were still low. Performance of this measure varied across indicators of serious injury and in age subgroups in the one study with subgroup comparisons; however, larger studies are needed to confirm these variations. The results of evaluations of respiratory rate are inconsistent, with reported sensitivities ranging from 2 to 76 percent. Combination measures performed better, with better results for a trauma score developed specifically for pediatrics.

In older adults, studies reported consistently low sensitivities and AUROCs for SBP, lactate, base deficit, respiratory rate, and assisted ventilation. Shock index also performed less well in older patients. Some variations of triage criteria modified for older adults by either changing thresholds or adding additional criteria (e.g., mechanism of injury) have demonstrated substantial increases in sensitivity (e.g., 76% to 92%), but this magnitude of improvement is not consistent across indicators of serious injury and came with similar substantial decreases in specificity (e.g., 78% to 42%).

**Discussion**

**Implications and Applicability**

For out-of-hospital clinical practice, our findings demonstrated that current circulatory and respiratory measures have low sensitivities but higher specificities. The evidence does not point to necessarily “better” cut-points for measures such as SBP, SI, and RR. In general, more liberal cut-points (e.g., SBP <110 mmHg vs. <90 mmHg) will raise sensitivity and lower specificity—an inevitable trade-off, but the magnitude of this trade-off may differ across tests.

However, based on the evidence we identified, no physiologic measures have high enough sensitivity that a negative result (e.g., normal physiologic value) could be confidently used to conclude that a patient is not at risk of being seriously injured, even with more liberal cut-points.

Our findings were based on a relatively large number of diverse studies. Having data from studies across a wide range of possible situations mirrored the reality of field triage and out-of-hospital assessment. While the diversity across the studies meant heterogeneity was high in the pooled estimates and the consistency across results was lower, the range was likely to reflect the variation that will be seen in trauma assessment and triage.

An approach to summarizing the data across studies and considering their impact is presented in Table D. This is a standard approach often used to present the implications of how well a screening test or triage tool performs. The pooled data are modeled to generate positive and negative likelihood ratios (LR+ and LR-). The positive likelihood ratio is Sensitivity/(1-Specificity) and the negative likelihood ratio is (1-Sensitivity)/Specificity. The likelihood ratios are then applied to different hypothetical pre-test probabilities and odds to produce post-test odds of the outcome (in this case serious injury) given a negative or positive test. The post-test probability if the test is negative (1-Negative Predictive Value) is also referred to as under-triage.
Table D. Post-test odds and probability of serious injury given pre-test assumptions

<table>
<thead>
<tr>
<th>Physiological Predictor (Test)</th>
<th>Serious Injury Indicator (Outcome)</th>
<th>Pre-Test Probability (Hypothetical)</th>
<th>Pre-Test Odds</th>
<th>LR+</th>
<th>LR-</th>
<th>Post-Test Odds (if a patient has positive test)</th>
<th>Post-Test Probability (PPV) (if a patient has positive test)</th>
<th>Post-Test Odds (if a patient has negative test)</th>
<th>Post-Test Probability (1-NPV) (if a patient has negative test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP &lt; 90</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>4.32</td>
<td>0.83</td>
<td>0.48</td>
<td>32%</td>
<td>0.09</td>
<td>8%</td>
</tr>
<tr>
<td>SBP &lt; 90</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>4.32</td>
<td>0.83</td>
<td>1.08</td>
<td>52%</td>
<td>0.21</td>
<td>17%</td>
</tr>
<tr>
<td>SBP &lt; 100</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>3.30</td>
<td>0.80</td>
<td>0.36</td>
<td>27%</td>
<td>0.09</td>
<td>8%</td>
</tr>
<tr>
<td>SBP &lt; 100</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>3.30</td>
<td>0.80</td>
<td>0.83</td>
<td>45%</td>
<td>0.20</td>
<td>17%</td>
</tr>
<tr>
<td>HR ≥ 110</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>1.37</td>
<td>0.91</td>
<td>0.15</td>
<td>13%</td>
<td>0.10</td>
<td>9%</td>
</tr>
<tr>
<td>HR ≥ 110</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>1.37</td>
<td>0.91</td>
<td>0.34</td>
<td>25%</td>
<td>0.23</td>
<td>19%</td>
</tr>
<tr>
<td>SI &gt; 1</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>3.13</td>
<td>0.71</td>
<td>0.34</td>
<td>26%</td>
<td>0.08</td>
<td>7%</td>
</tr>
<tr>
<td>SI &gt; 1</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>3.13</td>
<td>0.71</td>
<td>0.78</td>
<td>44%</td>
<td>0.18</td>
<td>15%</td>
</tr>
<tr>
<td>Lactate &gt; 2*</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>1.94</td>
<td>0.29</td>
<td>0.21</td>
<td>18%</td>
<td>0.03</td>
<td>3%</td>
</tr>
<tr>
<td>Lactate &gt; 2*</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>1.94</td>
<td>0.29</td>
<td>0.48</td>
<td>33%</td>
<td>0.07</td>
<td>7%</td>
</tr>
<tr>
<td>Lactate &gt; 4*</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>2.34</td>
<td>0.59</td>
<td>0.26</td>
<td>21%</td>
<td>0.07</td>
<td>6%</td>
</tr>
<tr>
<td>Lactate &gt; 4*</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>2.34</td>
<td>0.59</td>
<td>0.59</td>
<td>37%</td>
<td>0.15</td>
<td>13%</td>
</tr>
<tr>
<td>RR &lt; 10 or &gt; 29</td>
<td>Resource Use</td>
<td>10%</td>
<td>0.11</td>
<td>5.61</td>
<td>0.90</td>
<td>0.62</td>
<td>38%</td>
<td>0.10</td>
<td>9%</td>
</tr>
<tr>
<td>RR &lt; 10 or &gt; 29</td>
<td>Resource Use</td>
<td>20%</td>
<td>0.25</td>
<td>5.61</td>
<td>0.90</td>
<td>1.40</td>
<td>58%</td>
<td>0.23</td>
<td>18%</td>
</tr>
</tbody>
</table>

HR = heart rate; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; NPV = negative predictive value; PPV = positive predictive value; RR = respiratory rate; SBP = systolic blood pressure; SI = shock index

*Lactate >4 is based on emergency department data; lactate >2 is out-of-hospital*
Overall, our analysis demonstrated that physiologic measures have low sensitivity for identifying high-risk trauma patients (i.e., many patients will have normal physiology and prove to have serious injuries—there are higher numbers of false negatives), but have high specificity (i.e., patients with abnormal physiologic measures are likely to have resource needs, serious injuries, and are at higher mortality risk – there are few false positives). There was little evidence to suggest that one physiologic measure is significantly better than another (e.g., SBP versus SI versus lactate) because fewer studies compared these measures directly in head-to-head studies, the head-to-head studies were not amenable to pooling as they use different thresholds and outcomes, and the differences across our pooled estimate were small to moderate. However, combining different categories of physiologic measures (e.g., circulatory and level of consciousness) may increase predictive yield. Less extreme cut-points (e.g., lactate >2, SBP <110) raised sensitivity and lowered specificity, demonstrating that sensitivity and specificity have an inverse relationship when selecting dichotomous cut-points in continuous measures.

Limitations

The major limitations of the evidence base are the limited number of head-to-head comparisons and generally low strength of evidence available. As this review illustrates, there are a number of potential physiologic measures that could be used in triage and a range of indicators of serious injury used in this body of research. Our approach to this diversity was to focus on combining information for the same measure across studies and then looking across the measures. If we had limited our examination to comparable head-to-head comparisons we would have had small numbers of studies in each of a larger number of pairwise comparisons. However, there is a risk in comparing measure across studies rather than relying on comparisons within studies. Measures in different studies may produce similar results but for different populations. For example, if estimates of the AUROC for SBP and HR are similar, based on different studies with different populations, we could erroneously conclude that they will perform similarly across all patients when in truth SBP has this discriminant level for one subtype of patients while HR is similar but in a different subtype of patients. In order to assess this risk, we examined the results of the available head-to-head studies from the smaller number of studies that included direct comparisons and this did not change our conclusions. An overview of selected comparisons and all the results from these studies are included in the text and Appendix of the full report.

The literature available for analysis was dominated by studies that effectively limited their population to trauma patients who are transported by EMS. Most of the studies were based on data from trauma registries. While the specifics for inclusion vary across registries and also across studies that use administrative records in a similar way, standard practice seems to be inclusion of data collected on transported and/or admitted patients. The implication is that patients assessed by EMS but not transported are either not included at all or included inconsistently.

Another characteristic of the data in these registries is that it is usually collected prospectively but analyzed retrospectively, thus blurring the distinction between retrospective and prospective study types. In many cases the data sources are difficult to determine based on the published reports. Analysis is also complicated by the fact that the registry studies usually have large samples, while more clearly prospective studies we identified were often exploratory with small samples. The distinction matters because in other situations we might be able to make assumptions about the potential for differences in bias in prospective and retrospective studies, but in this literature the direction of the potential bias was not clear.

A substantial limitation in the evidence base was the lack of population-based samples where physiologic measures were collected in the out-of-hospital setting and patients were tracked across all hospitals (i.e., not limited to patients transported to major trauma centers), across phases of care (e.g., ED, hospital #1, transfer, hospital #2), and using population-based sampling to reduce selection bias. There was also limited detail about how the physiologic measure data were collected. Studies rarely reported details that could be important, such as what equipment was used, how and when the measurement was taken, and who was involved. Another important limitation of the research on this topic is the lack information on subpopulations, particularly children and older adults.

The evidence base also was inconsistent in how high-risk, seriously injured, trauma patients were defined, especially related to resource use. Studies tended to use a single indicator, such as need for a massive transfusion, rather than include multiple indicators, and even the definitions of given indicator varied across studies (e.g., what volume is considered massive and over what time period?). While the trauma research community has made efforts to come up with a comprehensive resource-based definition (i.e., the consensus-based criteria11 and lists of life-saving interventions), such a uniform definition is
not yet common in the trauma research. The result is that many studies may underestimate the utility of measures by requiring that they predict single or narrowly-defined indicators of severe injury.

There were also limitations to this review resulting from our decisions and processes. We included measurements in the ED as well as out-of-hospital measurements, but presented the ED and out-of-hospital results separately. We identified and included prognostic studies as they are similar but not identical to studies of predictive utility.

**Future Research Needs**

This review summarizes a sizable body of literature and it highlights several areas in which future research is needed.

One priority is for studies that compare, or at least document, differences in measurement (e.g., instrumentation, timing). This would allow the impact of these differences on the predictive utility of the measure to be considered.

Another priority is to encourage more research using the consensus-based criteria of the need for care in a major trauma center or a standardized list of life-saving interventions. If the indicators of high-risk patients were consistent, cleaner comparisons could be made both across studies and across measures. This would also permit an assessment of the utility of individual measures in a broader context.

Also, sampling patients in the out-of-hospital setting and tracking them through their hospital course and beyond, regardless of which hospital they were transported to, would help to reduce a large source of potential bias.

A key topic for additional research is the assessment of the utility of measures across age groups. While we did identify some studies that considered the use of physiologic measures for children and older adults, this is still a small subset of the literature and many questions remain. Age is often available or collected and if more researchers stratified analyses by age, even if age is not the focus of the study, a substantial amount of information would become available to inform decisions and improve care for children and older individuals.

**Conclusions**

While specifics vary across measures, settings, and populations, overall the predictive utilities of physiologic measures that are either currently used for trauma assessment and triage, or have been suggested, are moderate and not ideal. Measures of circulatory compromise (SBP, HR, SI, and lactate) and respiratory compromise (RR) have been evaluated in multiple studies, some with large numbers of patients. In general, these measures have low sensitivities, high specificities, and AUROCs in the fair-to-good range. Use of these measures should be guided by the understanding that when they are abnormal, that they are highly predictive of high-risk of serious injury in trauma patients, but that many patients with serious injuries will have normal physiologic measures. Combinations of these measures with assessments of consciousness seem to perform better, but how they would be implemented out-of-hospital needs to be determined, and then they need to be tested under field conditions to confirm their effectiveness and utility. Modification of triage measures for children or older adults is needed, given that these measures perform worse in these age groups than in adults; yet, the research has not yet identified better performing variations or replacements.

**References**


Full Report