*Comparative Effectiveness Review Disposition of Comments Report*

**Research Review Title:** *Glasgow Coma Scale for Field Triage of Trauma: A Systematic Review*

Draft review available for public comment from June 29, 2016, to July 26, 2016.

**Research Review Citation:** Chou R, Totten AM, Pappas M, Carney N, Dandy S, Grusing S, Fu R, Wasson N, Newgard C. Glasgow Coma Scale for Field Triage of Trauma: A Systematic Review. Comparative Effectiveness Review No.182. (Prepared by the Pacific Northwest Evidence-based Practice Center under Contract No. 290-2015-00009-I.) AHRQ Publication No. 16(17)-EHC041-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2017. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

## Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site or AHRQ Web site in draft form for public comment for a 3-4-week period. Comments can be submitted via the Web site, mail or E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2416*
*Published Online: January 12, 2017*

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #1** | General Comments | Your alleged differences in discrimination are almost certainly statistical noise rather than biologically correct, as there are two related reasons why your analysis is destined to underestimate the predictive value of shorter scales relative to the tGCS: | We agree; as noted in the Abstract/Conclusions and Discussion of the report, we state that the small differences in discrimination are unlikely to be clinically significant in several places, including the discussion. We reviewed the text to be sure this is clear. |
| **TEP Reviewer #1** | General Comments | 1) Your AUC analysis considers the discrimination of the full 13-point scale (i.e., 3 through 15) at all thresholds simultaneously; however the current National Field Triage Guidelines (your reference #4) recommend that it instead be applied in binary fashion: tGCS<=13 “yes” versus “no”. Thus, your analytic plan is not “real world”. | Very few studies reported diagnostic accuracy using binary cut points; however, we reported the results of the ones that did. We noted that few studies reported sensitivity/specificity at specified cut points, but findings were generally similar to those of the analysis based on the AUROC. |
| **TEP Reviewer #1** | General Comments | 2) AUC analyses inherently favor complexity. Any scale with more points will demonstrate more granularity and thus “smooth” each ROC curve—compared to a lesser scale with more abrupt linear segments, and thus lesser area under the curve. Further, there are substantial gaps between some points where an ROC curve extrapolates portions of the curve where decision-making really does not exist (e.g., tGCS cut points of 3.5 or 14.5), but yet nevertheless contributes substantial area under the curve volume attributable to such fictional points. | We revised the discussion and the “Limitations of the Review Process” section to note that the AUROC analyses may favor the tGCS and that few studies reported diagnostic accuracy using binary cut points. |
| **TEP Reviewer #1** | General Comments | Accordingly, due to the above numerical factors your alleged differences in discrimination are likely statistical noise rather than any biological difference. You could verify this by evaluating the tGCS the way it is actually used, by adding a sensitivity analysis of a simple contrast of the sensitivity and specificity of tGCS<=13 as a binary decision point versus GCSm<6 and versus SMS<3. (And/or the AUCs based upon these single thresholds.) | As noted in the Results (see also Table 5) and Discussion, few studies reported sensitivity and specificity based on standard cut points, but those that did found small differences, consistent with the analysis of the AUROC. The AUROC by definition is based on sensitivity and specificity calculated at different values of the measure, and is not based on a single threshold. |
| **TEP Reviewer #2** | General Comments | Excellent, very thorough and detailed review, and well done. Thanks for your efforts in this important effort | Thank you. |
| **TEP Reviewer #2** | General Comments | Is the report clinically meaningful? Yes through an in depth examination of the topic | Thank you. |
| **TEP Reviewer #2** | General Comments | Are the target population and audience explicitly defined? Yes | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2416*
*Published Online: January 12, 2017*

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #2** | General Comments | Are the key questions appropriate and explicitly stated? Yes | Thank you. |
| **TEP Reviewer #3** | General Comments | Authors have produced a very impressive summary of the evidence. Very comprehensive. My only concern is determining "who" assigned the quality assessment information and rated quality—was it a single reviewer or a dyad or a panel of experts. I was a bit lost based on mention of a single reviewer vs. two reviewers. Otherwise, this is a very nice piece of work. | Thank you. Regarding your concern, described in the Methods section, two reviewers independently rated the quality of each study, discussed any discrepancies, and came to a consensus if there was disagreement. |
| **TEP Reviewer #4** | General Comments | Field triage criteria need continual refinement as new evidence becomes important. Assessing the utility, reliability and ease of use of the tGCS and mGCS is necessary in order to come up with a parsimonious set of predictors for trauma resource need. Though not explicitly stated, the target audience can be inferred from the background section, i.e. those administrators responsible for establishing EMS field triage criteria. The key questions are explicitly stated, however the target populations for the systematic review are not. The populations under consideration for key questions 1-3 are trauma victims, however for key question 4 the target population is actually the emergency medical provider rendering care for the trauma victim. | The evidence reviewed in this report is relevant to clinicians, policymakers, and consumers, not just policymakers. Therefore, we did not specify a specific target population for the report, which differs from clinical practice guidelines, which are typically targeted at a specific audience. Key Question 4 assesses the reliability of the tGCS vs. the mGCS among trauma patients; therefore, the population of interest is trauma patients. Key Question 4a addresses subgroup effects related to the training and background of the person administering the instrument, but the overarching Key Question still addresses predictive utility among trauma patients. |
| **TEP Reviewer #5** | General Comments | I feel that the methods and research approach detailed in the document is exhaustive, objective and well prepared. My comments represent "less important" issues. Thus, judge them as you see fit: | Thank you, noted. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #5** | General Comments | Pg. 5 of 250, line 11. Use of the term "field" in the structured abstract suggests that the objective of this work was to assess GCS studies conducted in the out-of-hospital setting. We then bring in studies which were conducted in the ED…a much different treatment environment. Much of the premise (and conclusions) associated with this investigation hinges on the idea that the "field" setting is such that a full tGCS assessment may not be advisable, when compared to the ease of a mGCS assessment. Yet, while the ED can be a busy, rushed environment, it is vastly different than the milieu surrounding a "field" environment. Somewhere early in the manuscript, a rationale should be provided as to why research conducted in the ED was included in the review of the literature and in the assessment of findings. | The Introduction notes that evidence about field triage instruments frequently relies on extrapolation from studies conducted in EDs, but that performance might differ in these settings. As noted in the "Assessing Research Applicability" section, we evaluated factors important for understanding the applicability of studies, including the setting in which assessment occurred (field vs. ED), and presented stratified analyses based on where the assessment took place. As described in the Results section, findings were very similar for studies in which assessment was conducted in the field vs. in the ED. |
| **TEP Reviewer #5** | General Comments | Pg. 5 of 250, line 36. Here we appropriately indicate that the reviewed published literature using "patient mortality" as an endpoint is limited to only "in-hospital mortality". Yet, we lose this designation throughout the much of the remaining document…..instead using the term "mortality". For example, Table ES-1 (line 7) list "Mortality, Overall"….but this should indicate that this is only in-hospital mortality. It is widely understood that in-hospital mortality excludes pre-hospital deaths and post-discharge deaths, the frequency of the latter is highly dependent on hospital discharge practices. Deaths occurring pre- or post-hospitalization could significantly influence assessments of precision or reliability. Thus, I think we should carefully carry the designation of "in-hospital deaths" throughout the manuscript. | We reviewed the report to ensure that we refer to "in-hospital" mortality when appropriate for that specific outcome. |
| **TEP Reviewer #5** | General Comments | Pg. 5 of 250, line 38. When referring to "AUROC" by spelling out the acronym, we should be adding the word "curve". We spell out the acronym several times in the manuscript. Currently we indicate that "AUROC" stands for pooled mean difference in area "UNDER" the receiver operating characteristic. It is not the area under a "characteristic". It is the area under a "curve". | This has been corrected throughout the report. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #5** | General Comments | I have many other comments….but they are editorial in nature! | Thank you, noted. |
| **TEP Reviewer #5** | General Comments | Thank you for the chance to review this initial document! | Thank you for your review. |
| **TEP Reviewer #6** | General Comments | Yes. I think some discussion of the missing data rate or completeness of the GCS documentation within each study is needed. In the datasets I work with there is lots of missing GCS documentation. If from a predictive stand point tGCS and mGCS are similar but one is captured more often than the other that will increase utility. I think you somewhat capture this with ease of use but I think it is a little broader than that. Further the bias of missing data may also be a factor. We attempted to do a tree analysis to build a better guideline and data missing was frequently the better predictor. | The assessments for risk of bias include whether studies reported the amount of missing data, and how much data was missing (shown in Appendix K). In all of the studies of predictive utility, the mGCS was taken from the tGCS (they were not independently administered), and patients with missing GCS data were excluded from the analysis, so it is not possible to determine how missing data would have impacted predictive utility differentially for the different scales. We revised the Limitations of the Review Process section to note that studies that reported low missing data reported results similar to those with high or unclear missing data. |
| **TEP Reviewer #6** | General Comments | I also think that there needs to be some additional discussion of accuracy in determining GCS. There is the GCS score that is perceived by the provider and the "true" GCS score. The GCS score assigned by the provider may be in accurate but that inaccuracy may still work for predicting. I think you need to be a little clearer about who did the GCS measuring. I think they are mostly EMS provider determined during treatment but I am not sure that is the same as determined in the ED by a trained observer or some other method. | The studies on comparative predictive utility are based on the actual scores assigned by providers. Results were similar for studies based on GCS performed in the field by EMS personnel and studies based on GCS performed in the ED. As noted in the Applicability section, no study evaluated how predictive utility varied according to the level or training of personnel administering the GCS. |
| **Peer Reviewer #1** | General Comments | This is a very comprehensive review of the literature on an important topic. The manuscript is well organized and the methods are clear. Key questions are clearly and explicitly stated | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #2** | General Comments | This is a superb report. The paucity of data is a challenge, but your methods have addressed this and you managed to come up with fairly clear statements that will guide prehospital care. Thank you for your hard work. Please see my comments below, meant to guide you towards a strengthened report and improved readability and knowledge translation. | Thank you, noted. |
| **Peer Reviewer #3** | General Comments | While the report is generally well-done and clinically meaningful, it seems to be written in the context of the assumption that total GCS is a criterion standard when the recent literature has been suggesting that it is actually of fairly limited use (Ann Emerg Med 2011;58(5):427–430), especially in trauma. | As noted in the Background section, the GCS is currently used in field triage of trauma, the purpose of our report was to compare tGCS vs. mGCS or SMS. This is discussed in the article cited by the reviewer: “simpler scales perform just as well.” tGCS was not assumed to be a criterion standard; rather it was evaluated as a potential tool for use in field triage. |
| **Peer Reviewer #4** | General Comments | Regarding the phrasing of the key questions, I would consider modifying them as follows:<br><br>"...known or suspected trauma.... should probably be ...known or suspected head trauma....."<br><br>GCS applies to the evaluation of patients with head (brain) trauma. In my opinion, it would be useful to state that outright in the key questions by including the word head or brain with the phrase "known or suspected trauma". If the brain is not injured, GCS is not relevant. | The original Key Questions focused on brain trauma, but after discussions with the Technical Expert Panel, the scope was broadened to include all trauma, since the GCS is applied to all trauma in the field in clinical practice. We did perform stratified analyses based on the type of trauma and found no clear differences in findings. |
| **Peer Reviewer #5** | General Comments | Thank you for allowing me to serve as a reviewer on this important work. I only have a couple of specific edits that are included below. With regard to the overall work, I have a number of comments. Overall, a tremendous amount of work went into this document and the work was much needed. | Thank you, noted. |
| **Peer Reviewer #5** | General Comments | Unfortunately, work ended up primarily demonstrating that we have little high quality research on the subject at hand in the EMS community. Although, I would suggest that there is little work in any setting that looks at a specific patient assessment and evaluates its overall sensitivity/specificity. | Thank you for the comment. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #5** | General Comments | The conclusions, as presented, are consistent with the science presented but do amount to a collective "so what" when evaluated. The take home message is that there is a small but measurable difference between tGCS and mGCS and the main recommendation is for further research. | Thank you for the comment. We attempted to describe the implications of the findings in the Discussion and Conclusions. |
| **Peer Reviewer #5** | General Comments | I would more forcefully argue that the data would push things toward utilizing mGCS instead of tGCS. | The Conclusions state: "The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use." The purpose of our report was to synthesize the evidence; others may use the evidence to provide recommendations regarding which instrument to use. |
| **Peer Reviewer #5** | General Comments | The manuscript points out that simplicity may be a reason to consider the to mGCS from tGCS. Given the small difference between the two, then simplicity and reproducibility should be the major goal. | The Conclusions state: "The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use." |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #5** | General Comments | GCS is not likely to be a solely independent factor in a trauma triage decision, and therefore any attempt to tease out its impact is difficult at best. The fact that the best available data shows little difference implies that the mGCS portion of the tGCS is probably the biggest discriminator – or that if mGCS is off, then the other components will also be similarly depressed | The Conclusions state: "The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination is likely to be small, and could be offset by factors such as convenience and ease of use." |
| **Peer Reviewer #5** | General Comments | In practice, GCS is rarely fully calculated before the decision is made to transport directly to a trauma center. All charting is retrospective and the narrative will fit the ultimate decision. A study that tries to prospectively evaluate mGCS may be of benefit but I suspect would have similar findings and certainly would create a "groundbreaking" change. As a result, I believe that the herculean task that was performed is probably the best look that we will have at this subject and that the data shows that applying mGCS would simpler, almost equivalent and may even result in less over-triage. | Thank you for the comment. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Public Reviewer #3 Douglas Kupas Geisinger Health System** | General Comments | I believe it is a disservice to suggest that we need more clinical trials to make a change in field triage related to GCS total vs "does not follow commands", and I make this based upon the following points. - Studies of many hundreds of thousands of patients suggest there is a very small increased accuracy in identifying serious traumatic injuries using GCS =13 as opposed to GCS motor 6. - Studies have shown rather large discrepancies in EMS provider ability to accurately calculate total GCS. - Both of these are successful in identifying the vast majority of serious trauma patients that can be identified by using some component of mental status. - Identifying patients that don't follow commands (GCS motor 6) is a comparatively easy concept to teach, especially to lower level healthcare providers. I would ask the reviewers in this project to consider these issues and consider at least making a statement that the evidence would support either of these approaches for the purpose of trauma triage, as they both meet the criteria of having a 20% likelihood of identifying a patient with ISS 15, and there is reason to believe that any small improvement by using total GCS is likely negated by the larger degree of miscalculation of total GCS. Personally, I think that it could be taken even further - from similarity to support for the easier "does not follow commands" approach. Thank you for considering this point of view in your review and conclusions. | The purpose of this report is to synthesize the evidence on tGCS vs. mGCS or SMS, not to make clinical recommendations. We do not state that clinical trials are needed to change recommendations regarding which scale is used in field triage, though we do note that head-to-head studies (observational or randomized) would be helpful to understand effects on clinical decisionmaking and clinical outcomes. Our conclusions note that the differences between the tGCS and mGCS or SMS in discrimination are small, may not be clinically important, and may be offset by factors such as ease of use. |
| **TEP Reviewer #1** | Introduction | Excellent | Thank you. |
| **TEP Reviewer #2** | Introduction | Good. | Thank you. |
| **TEP Reviewer #3** | Introduction | Abstract -- Lines 15-32 – there is no mention of the framework used to evaluate the literature (evidence) / strength of evidence. Which framework was used? GRADE or some other methodology? On what instruction was the quality of literature rated and by whom? | As described in the Methods section, we assessed the strength of evidence using the approach described in the AHRQ Methods Guide, which is similar to the methods used by GRADE. The methods for assessing study quality are also described in the Methods section. (We used criteria adapted from the QUIPS tool for assessing prognostic studies.) |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #3** | Introduction | Abstract – Lines 33-34 – Consider not beginning a sentence with a number. | In order to stay concise, we kept this sentence as is. |
| **TEP Reviewer #3** | Introduction | My opinion is that the executive summary is a bit burdensome to read. I would raise the observation that most non-academics will have problems reviewing this. Consider maybe providing two different summaries – one directed to clinicians and administrators – a second to academics and investigators. Finally, maybe a one-pager for policy makers. | The Executive Summary summarizes the report, including background, methods, results, and conclusions. The Abstract provides a one-page summary of the report. AHRQ may commission a separate group, the Eisenberg Center, to develop translational products aimed at clinicians, policymakers, and consumers. |
| **TEP Reviewer #3** | Introduction | Executive Summary / Background – use of the phrase "EMS Provider" is associated with a term used by the Centers for Medicare and Medicaid Services (CMS). Consider using the term(s): EMS personnel or EMS clinicians. | We changed "EMS provider" to "EMS personnel." |
| **TEP Reviewer #3** | Introduction | Executive Summary / Scope of Review and Key Questions line 47-48. The first sentence seems awkward. Were the research questions selected mentioned above? Seems odd to begin a section/sentence with "The research questions……" without having previously listed those questions. If I missed it, ignore this comment. | We moved the Key Questions before the explanatory text in this section. |
| **TEP Reviewer #3** | Introduction | Same section / paragraph – there is very little information here about what was done (e.g., a systematic review) and how the literature was evaluated/rated. No need to be too elaborate. Simply stating use of a systematic review would be helpful – or take out any information that is even remotely related to methods in this section and put in methods section below. There is a bit of a mix of some methods information and some background information in this section. Consider moving some text elsewhere or re-arranging. | The Methods section describes the systematic review methods that were used, including the Literature Search Strategy, Risk of Bias Assessment, and Data Synthesis Methods, as well as methods for assessing the Strength of the Body of Evidence. Additional details are provided in the full report. |
| **TEP Reviewer #3** | Introduction | Executive summary – Risk of Bias Assessment of Individual Studies – why not mention the framework used to rate quality of studies? (e.g., In accordance with the XXXXXX framework, we rated XXXXXXXXX). | We revised to state: "Two investigators independently rated the quality of studies (good, fair, poor) using prespecified criteria developed for evaluation of studies on prognosis and diagnosis," with references to the QUIPS tool (prognosis) and the QUADAS tool (diagnosis). |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #3** | Introduction | Executive summary – Results section line 30-32 – I thought it was mentioned previously that a single reviewer abstracted details of papers retrieved? Now it says "After dual review of abstracts and titles……" Clarify? Any adjudication of disagreements between reviewers on include/exclusion of articles? | We omitted this detail from the ES, but added it back in: "All citations were independently reviewed by two investigators to determine eligibility for inclusion, and discrepancies resolved by discussion and consensus." The abstraction occurs after this dual review determines whether a study is included and is done by one person and checked by another. We reviewed the text to be sure the different tasks are clearly described. |
| **TEP Reviewer #4** | Introduction | The introduction is focused, well written and provides appropriate references. | Thank you. |
| **TEP Reviewer #6** | Introduction | I have concerns with the statement that clinical outcome of mortality prediction is the goal and your analytic framework picture showing that the main outcome is mortality prediction. The goal of field triage is to get the patient to a hospital with the resources to treat their condition which should reduce morbidity and mortality. GCS predicting mortality is not the goal of using it in the guideline. The goal is to predict resource needs of the patient. Predicting death is frequently used as a proxy for needing trauma center resources. Thus reducing under and over triage is the main goal of field triage which should in turn reduce morbidity and mortality. Defining the gold standard for that assessment is the trick. Mortality is one of the easier outcomes to obtain and is frequently used in the literature but whether it discriminates who does and does not need a trauma center is questionable. I would change you analytic framework to better reflect the intended use of this review to determine if the field triage guidelines should use tGCS or mGCS to reduce over and under triage which should then reduce morbidity and mortality. It is not that GCS predicts mortality it predicts who needs special stuff that is only available at select hospitals. | To clarify, we do not state that mortality prediction is the goal of field triage in the Analytic Framework. As indicated in the Analytic Framework, the ultimate goal of field assessment (outcome farthest to the right) is to impact final health outcomes, as addressed by Key Question 3. Key Question 2 addresses effects on resource utilization, an intermediate outcome measured by over- and under-triage. Key Question 1 addresses predictive utility, an even more intermediate outcome (farther to the left in the analytic framework). As described in the Results, no study evaluated effects of the tGCS vs. mGCS or SMS on clinical outcomes or rates of over- or under-triage; therefore, it is necessary to rely on studies that evaluate comparative predictive utility. |
| **Peer Reviewer #1** | Introduction | appropriately sets the stage for the review | Thank you. |
| **Peer Reviewer #2** | Introduction | The introduction is strong. I have no comments. | Thank you. |
| **Peer Reviewer #3** | Introduction | No concerns. | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #4** | Introduction | This work will be clinically meaningful when the findings are used to validate the role of GCS as a component of the revised National Trauma Field Triage Guidelines. | Thank you, noted. |
| **Peer Reviewer #5** | Introduction | Well presented. The rationale was well developed and the importance is properly emphasized. | Thank you. |
| **Peer Reviewer #5** | Introduction | A couple of minor points:<br>ES-1, line 25 and 1,28 – help identify potentially seriously | "Potentially" was added to the sentences in the ES. |
| **Peer Reviewer #5** | Introduction | ED-2, line 30 and 2,32– Advanced EMT/Paramedic are 2 distinct levels of EMS practitioner | We revised to reflect the current EMS personnel classifications: Emergency Medical Responder, EMT, Advanced EMT, and Paramedic. |
| **TEP Reviewer #1** | Methods | See general comments | Thank you. |
| **TEP Reviewer #2** | Methods | Are the inclusion and exclusion criteria justifiable? Yes | Thank you. |
| **TEP Reviewer #2** | Methods | Are the inclusion and exclusion criteria justifiable? Yes | Thank you. |
| **TEP Reviewer #2** | Methods | Are the definitions or diagnostic criteria for the outcome measures appropriate? Yes | Thank you. |
| **TEP Reviewer #2** | Methods | Are the statistical methods used appropriate? Yes | Thank you. |
| **TEP Reviewer #3** | Methods | No major comments. My only concern is determining "who" assigned the quality assessment information and rated quality—was it a single reviewer or a dyad or a panel of experts. I was a bit lost based on mention of a single reviewer vs. two reviewers (see prior comments regarding feedback on the executive summary). Overall, nicely done. | As described in the Methods section, two reviewers independently rated the quality of all studies, discussed any discrepancies, and came to a consensus if necessary. We also ensured that investigators did not review, assess, or screen papers that they were an author on and this has been added to the Methods. |
| **TEP Reviewer #4** | Methods | The inclusion and exclusion criteria are justifiable, however the authors need to specify that for key question 4, the population under study were emergency medical providers that assign GCS for trauma victims. Will need to address this in all pertinent sections of the methods. | Key Question 4 assesses the reliability of the tGCS vs. the mGCS among trauma patients; therefore, the population of interest is trauma patients. Key Question 4a addresses subgroup effects related to the training and background of the person administering the instrument, but the overarching Key Question still addresses predictive utility among trauma patients. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #4** | Methods | As a reviewer, I struggled with the use in the difference in the AUROC with CI >0 indicating significance, particularly when the confidence intervals for the AUROCs for most analyses were overlapping suggesting no difference in tGCS, mGCS, and SMS. At a minimum, this warrants a better justification in the methods section. | We reported the pooled differences in AUROC values with 95% CIs, which were statistically significant for comparisons involving tGCS vs. mGCS or SMS, as shown in Table 4. The differences in AUROC values in individual studies are shown in Table 3 and in the forest plots, including the point estimate and confidence intervals for the differences (e.g., Figure 3). It is common for some individual studies in a meta-analysis to report results that are not statistically significant and for the pooled result to be statistically significant; in fact, increasing statistical power is one of the reasons to perform meta-analysis. |
| **TEP Reviewer #4** | Methods | There is no mention of how key question 4 is evaluated. The authors need to introduce the measures of inter-rater reliability that were included, what a meaningful threshold for the measure is (e.g. a kappa with LCL >0.4 indicates moderate agreement between assessors) and what the a priori meaningful difference in inter-rater reliability is. | The "Outcomes" section of the Methods states that for Key Question 4, we included outcomes that assessed reliability (e.g., interrater and intrarater kappa). We revised the "Assessing Research Applicability" section to note that we considered a difference in kappa of ≥0.20 to be clinically meaningful. However, no study directly compared the reliability of the tGCS vs. the mGCS, so we were unable to apply this threshold to any findings. |
| **TEP Reviewer #6** | Methods | Methods are fine | Thank you. |
| **Peer Reviewer #1** | Methods | all methods appropriate, comprehensive assessment of clinically relevant outcomes | Thank you. |
| **Peer Reviewer #2** | Methods | The statistical methodology is strong. | Thank you. |
| **Peer Reviewer #2** | Methods | The systematic review methodology could be made a bit more clear. | We revised the text to address the specific methods issues brought up in separate comments. |
| **Peer Reviewer #2** | Methods | Was this systematic review registered with PROSPERO? | Yes, this is noted on page 6, lines 26-28. |
| **Peer Reviewer #2** | Methods | What happened if there was disagreement between the two abstract reviewers? | Discrepancies were resolved through discussion and consensus. This is stated on page 8, line 32. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #2** | Methods | Were Kappa scores calculated for the abstract/full text review process? | We did not calculate Kappa scores. The goal of the selection process is to identify the "correct" papers, not necessarily to have high agreement among reviewers. We are unaware of any literature showing that Kappa scores are associated with the likelihood of selecting the correct papers; the AHRQ methods guide recommends against calculation of Kappa scores. |
| **Peer Reviewer #2** | Methods | Were papers blinded/de-identified when screened? | Investigators were not blinded to papers, as this would require substantial additional resources and evidence indicates that this has little impact on the findings of reviews; the AHRQ methods guide does not recommend use of blinded/de-identified papers. |
| **Peer Reviewer #2** | Methods | Risk of Bias assessment is fairly described. | Thank you, noted. |
| **Peer Reviewer #2** | Methods | I am not a meta-analysis expert but the description seems strong to me. | Thank you, noted. |
| **Peer Reviewer #3** | Methods | Yes to all – the search appears to have been well done, and I am not aware of any relevant articles that have been missed. | Thank you, noted. |
| **Peer Reviewer #4** | Methods | Are the inclusion/exclusion criteria justifiable? yes | Thank you. |
| **Peer Reviewer #4** | Methods | Are search strategies explicitly stated and logical? yes | Thank you. |
| **Peer Reviewer #4** | Methods | Are the definitions or diagnostic criteria for the outcome measures appropriate? yes | Thank you. |
| **Peer Reviewer #4** | Methods | Are the statistical methods used appropriate? I think so. | Thank you. |
| **Peer Reviewer #5** | Methods | The methods were well developed. I am not an expert in search methodology but have participated in a number of GRADE methodology reviews and found the development of the questions to be reasonable and appropriate. Additionally, I agreed with the methodology, as described, as to when to exclude papers. | Thank you, noted. |
| **TEP Reviewer #1** | Results | See general comments | Thank you. |
| **TEP Reviewer #2** | Results | Is the amount of detail presented in the results section appropriate? Yes | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #2** | Results | Are the characteristics of the studies clearly described? Yes | Thank you. |
| **TEP Reviewer #2** | Results | Are the key messages explicit and applicable? Yes | Thank you. |
| **TEP Reviewer #2** | Results | Are figures, tables and appendices adequate and descriptive? Yes | Thank you. |
| **TEP Reviewer #2** | Results | Did the investigators overlook any studies that ought to have been included or conversely did they include studies that ought to have been excluded? Not that I have identified | Thank you, noted. |
| **TEP Reviewer #3** | Results | No major comments. The use of the appendices is impressive and helps provide transparency to the other sections. | Thank you. |
| **TEP Reviewer #4** | Results | Throughout results, sometimes the authors present both AUROC with confidence intervals and Difference in AUROC with confidence intervals and sometimes just the Difference in AUROC with confidence intervals. I would choose one standard way of reporting throughout the document. | We provide the pooled AUROC for tGCS, mGCS, and SMS, as well as the difference in the AUROC for the main results. We only provided the difference in AUROC for the sensitivity analysis and not the pooled AUROCs because they were similar to the main results and in order to focus on comparative predictive utility. |
| **TEP Reviewer #4** | Results | The results of the key question 1a are presented within key question 1 as well as within its own section on page 45. This is redundant and should be narrowed to one or the other. | We attempted to minimize redundancy between Key Question 1 and 1a to the extent possible. Some of the subgroup/sensitivity analyses related to age, type of trauma, and assessment setting are briefly presented in Key Question 1 in order to describe key stratified analyses and permit interpretation of overall findings, including robustness of findings, with more detail provided in Key Question 1a, which specifically addresses subgroup effects. |
| **TEP Reviewer #4** | Results | Despite the paucity of evidence for key question 2, a detailed assessment/overview of what is available is provided. This is lacking for key question 3 and should be added. | The Results section for Key Question 2 includes a discussion of two studies that did not meet inclusion criteria but might provide some contextual information regarding rates of over- and under-triage. No similar study on effects on clinical outcomes was available for Key Question 3. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #6** | Results | In one place you say the outcome used was ISS greater than 5 I suspect that is a typo and you meant 15. I would search on ISS and make sure they are correct throughout the document. (See page 69 line 20) | We corrected this typo and checked for other occurrences. |
| **Peer Reviewer #1** | Results | clearly presented, figures and tables are excellent | Thank you. |
| **Peer Reviewer #2** | Results | Can you help me understand ES09 line 55 bullet point, specifically if insufficient is appropriate when reliability appeared high? Perhaps SOE low is appropriate here? (The interrater reliability of tGCS and mGCS appears to be high, but evidence was insufficient to determine if there were differences between scales (SOE: Insufficient).) | Insufficient is appropriate for the strength of evidence because the question is about the comparative interrater reliability, for which there was insufficient evidence (no head-to-head studies). There was some evidence that interrater reliability of the tGCS and mGCS are high when each is looked at individually, but there is insufficient evidence to determine how they compare with one another. |
| **Peer Reviewer #2** | Results | The schematic on page 13 is excellent. | Thank you, noted. |
| **Peer Reviewer #2** | Results | Are the SOE's on Pages 39-41 consistent with the SOEs in the executive summary? | Yes, they are the same (note that the reviewer is referring to pages 39-41 of the PDF, which are pages 14-16 of the report). |
| **Peer Reviewer #2** | Results | The questions are answered fairly and concisely. | Thank you. |
| **Peer Reviewer #3** | Results | There is WAY too much detail provided – much (most?) of the “results” information should be summarized in a few easy-to-read tables. This is my biggest gripe with the paper – 250 pages, for what could be a regular-length journal article?? Also not sure why three sections are needed (abstract, executive summary, full report) that restate, often verbatim, the same material. | The Executive Summary and Abstract provide summarized results; we have added a Box on the title page summarizing the key messages from the report. We also plan to prepare a manuscript to submit for journal publication. Following the AHRQ format guidance, the full report (text 63 pages, including figures and in-text tables) provides full details including methods and more detailed results; the detailed evidence tables/appendices are available for those interested. |
| **Peer Reviewer #4** | Results | Level of detail is fine. | Thank you. |
| **Peer Reviewer #4** | Results | Material is clearly presented. | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #5** | Results | I guess the one thing that is not clear in the way the results are discussed is: what does a difference in the AUROC really mean? Again, not a methodologic expert, so I had a hard time determining what the real world implications really were. The number seemed small but would like a little more explanation. | The Applicability section states: "The differences between the tGCS versus the mGCS or SMS in mean AUROC ranged from 0.01 to 0.05. This indicates that the ability of the scales to distinguish patients who experience an outcome from those who do not based on a higher score is 1 percent to 5 percent higher with the tGCS than with the more abbreviated scales. These differences were statistically significant, in part due to the large sample sizes evaluated in the studies. Although we classified such differences as "small," based on a priori thresholds, such thresholds are by nature somewhat arbitrary. The importance of "small" differences in discrimination depends in part on the seriousness of the outcome evaluated, the degree to which triage and other treatment decisions are based on the field triage scale, and the degree to which such actions impact clinical outcomes." |
| **TEP Reviewer #1** | Discussion/ Conclusion | See general comments | Thank you. |
| **TEP Reviewer #2** | Discussion/ Conclusion | Are the implications of the major findings clearly stated? In the "Implications" section, I think clinicians would like for this to be more clear (and may not be able to at this point) about what this means-either do or don't use mGCS at this time based upon the evidence, along with your admonition for ongoing research, etc.-or that these cannot be stated based upon the evidence-just something more definitive. | The Conclusions note that differences in discrimination between the tGCS and mGCS or SMS are small and may be offset by greater ease of use. The purpose of this report is not to provide clinical recommendations, but to summarize the current evidence. |
| **TEP Reviewer #2** | Discussion/ Conclusion | Are the limitations of the review/studies described adequately? Yes | Thank you. |
| **TEP Reviewer #2** | Discussion/ Conclusion | In the discussion, did the investigators omit any important literature? Not that I have identified. | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #2** | Discussion/ Conclusion | Is the future research section clear and easily translated into new research? Yes, clearly identified the key need for a head-to-head study as the next step in this process. In addition to level of training, it might be helpful to assess targeted educational efforts and their impact. | We revised the Future Research Needs section to mention the impact of educational efforts on interrater reliability and ease of use. |
| **TEP Reviewer #3** | Discussion/ Conclusion | I think the initial paragraph summarizing findings is good, but could be more declarative. I did not feel like I left the initial paragraph with the main point of the review / findings. I had to read a few times. | The initial paragraph of the Discussion in the Executive Summary states: "Based on head-to-head studies, we found that the tGCS is associated with slightly better predictive utility than the mGCS, based on the AUROC, a measure of discrimination. The tGCS is better able than the mGCS to discriminate people with trauma who undergo neurosurgical intervention, have severe TBI, or undergo emergency intubation from people who do not experience these outcomes. However, the difference in the AUROC on each of these outcomes was small (<0.05)." We believe that this accurately and succinctly summarizes the findings on predictive utility. |
| **TEP Reviewer #4** | Discussion/ Conclusion | The first section of the discussion, Key Findings and Strength of Evidence is long and redundant. It should be shortened and hit the high points without recanting all of the findings. | We feel that the Discussion section warrants a more in-depth discussion of key findings and issues. The Conclusions section summarizes the main findings in a single paragraph. (Findings are also summarized in the Abstract.) |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #4** | Discussion/ Conclusion | While I agree with the conclusion, the discussion (particularly the first section) reads as if the tGCS is being advocated for. What I glean from the results is that there is near equipoise in the emergency setting between tGCS and mGCS (particularly given the overlapping confidence intervals for the AUROCs). Given the ease of use of the mGCS, I think that it is fair to advocate for the use of mGCS in field triage. However, it is also important to clearly state that in other environments, tGCS may be a better predictor of outcome or clinical course. | We do not feel that the Conclusion advocates for tGCS; it states: "The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use." The purpose of this report is to synthesize the evidence; others may utilize the evidence to provide recommendations about which assessment instrument to use. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #6** | Discussion/ Conclusion | In addition, I disagree with the recommendation that a head to head study be done with two groups of patients. That seems like a poor study design suggestion given that both scales can be measured in the same patient and compared with the outcome suggesting that some sort of RCT be done seems like an inefficient recommendation that would be costly and difficult to get sufficient sample to control for all of the potential confounders. | As noted in the Limitations section, the mGCS or SMS were retrospectively determined for each patient from the tGCS. Therefore, the scales were not performed independently, and it is uncertain how the mGCS or SMS would perform alone, since the other parts of the GCS could have impacted how the motor component was scored. It is also necessary to have studies in which different assessment scales are applied in different patients in order to understand effects on over- or under-triage, and on clinical outcomes. We revised the Future Research Needs section to be clearer that head-to-head studies could be observational or randomized. Observational studies that address confounders could be designed, and, given the equipoise between tGCS and mGCS based on predictive utility, we think that an RCT (in which confounders should be randomly distributed) would also be appropriate. The identification of Future Research Needs does not mean that clinicians and policymakers cannot make decisions about use of more simplified scales based on currently available evidence; however, the role of this report is to summarize the current evidence, not provide clinical or policy recommendations. |
| **Peer Reviewer #1** | Discussion/ Conclusion | Discussion and conclusion are appropriate, clearly more research is needed but this document defines the issues very well | Thank you. |
| **Peer Reviewer #2** | Discussion/ Conclusion | I thought the discussion was prudent and concise. The future research section is clear and poignant. | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #2** | Discussion/ Conclusion | The discussion does not discuss much about interrater reliability of a GCS, which may be a limitation to tGCS that favors mGCS despite its marginally worse AUROC. It would be worth noting something about the variability in EMS practitioner ability, as many would argue that tGCS is poorly and inconsistently applied by workers in different states/systems. While not directly part of the questions you started with, it is an important elephant-in-the-room that not all EMS systems are created equal. Thus, a tailored approach to selecting the use of tGCS vs mGCS/SMS based on "local provider capabilities" may be appropriate given the minimal clinical significance of the statistically significant differences. I realize this is discussed on page 55 but I think you can be less perfunctory here or repeat it in other parts of later discussion as it is KEY going forward. Perhaps expand on the sentence you include on page 63 in future directions. | To clarify, there was insufficient evidence to determine comparative reliability, although there was limited evidence that mGCS is easier to use. In addition to the Discussion on p 55, in the section on Implications for Clinical and Policy Decisionmaking, we note: "Limited evidence suggests that the mGCS may be easier to score correctly than the tGCS, which may offset disadvantages related to slightly lower predictive utility."<br><br>There was limited evidence to determine how accuracy of the tGCS and mGCS vary according to the EMS practitioner (p 55). In the Discussion we note that "Evidence on how factors related to patients, assessors, and settings impacts predictive utility is limited. However, even if such differences exist, there may be advantages to having a single scale that can be applied across trauma scenarios, instead of requiring field assessors to select from among different scales for particular situations, even if the predictive utility of the single scale is slightly lower in certain situations." The trade-offs between predictive utility and ease of use is also mentioned in the Conclusions.<br><br>We revised the Future Research Needs section: "Studies that evaluate how the predictive utility of the tGCS compared with the mGCS or SMS according to the level *or type of training* of assessing personnel in the field are also needed" ("type of training" added).<br><br>The purpose of this report was to summarize the current evidence; others may use the evidence to make recommendations regarding which scales to use in which settings. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #3** | Discussion/ Conclusion | Yes to all – except that some discussion is needed regarding the overall limitations of GCS in the context of assessing the trauma patient, to help the reader understand that overall it is falling out of favor. | The GCS is currently recommended in the CDC field triage guideline, as noted in the Background section introducing the report. The purpose of the report is to compare the tGCS vs. more abbreviated versions of the GCS; we did not address whether the GCS should be used in field triage at all. |
| **Peer Reviewer #4** | Discussion/ Conclusion | Findings are clearly stated. | Thank you. |
| **Peer Reviewer #4** | Discussion/ Conclusion | Limitations are well described. | Thank you. |
| **Peer Reviewer #4** | Discussion/ Conclusion | Literature search seems complete and comprehensive. | Thank you. |
| **Peer Reviewer #5** | Discussion/ Conclusion | See first entry. Would strongly recommend that the data shows little to no difference and that switching to mGCS would be a significant operational improvement. | The Conclusions state: "The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use." The purpose of this report is to synthesize the evidence, not provide recommendations regarding which instrument to use. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Public Reviewer #3 Douglas Kupas Geisinger Health System** | Discussion | As a participant in the expert panel that established the 2011CDC/ACS guidelines for field triage of the trauma patient, I was involved in discussions about the criteria to add an element to the guidelines. In general, to be used for field triage, and element needed to have a 20% chance or more of association with an ISS 15 to be included. By this criteria, both the GCS =13 or GCS motor 6 would have an appropriate association with serious trauma to be included in the guidelines. I believe that the issue is what recommendation should be made regarding two tests that give a very similar result. Statistically, it is quite easy to attain statistical significance when some of these studies have hundreds of thousands of patients, but the clinical differences are very small. | The Conclusions note that the differences in discrimination were small and probably clinically insignificant and may be offset by factors such as ease of use. The purpose of this report was to synthesize the current evidence, not make clinical recommendations. |
| **Public Reviewer #3 Douglas Kupas Geisinger Health System** | Discussion | The data in the included studies are from centers where individuals with the highest expertise are calculating the total GCS, and there is good evidence that the miscalculation of GCS by EMS providers (some of whom at the lower levels of training have had only a very basic education and experience with using GCS) likely far exceeds the small difference in clinical outcome. Your paper suggests that using the total GCS may identify 1-3 more patients that require neurosurgical intervention or die in hospital, but it also suggests that there are 6-27% more errors in calculating total GCS than using GCS - motor (This is further simplified that GCS motor 6 only requires a simple assessment of "does the patient follow commands?" I would suggest that the review that only compares the outcomes of total GCS vs GCS motor misses the importance of this error of calculating GCS, particularly by EMS providers. | Several studies utilized data from large regional or national databases that would appear to be fairly applicable. Estimates of discrimination from such studies incorporate any effects of errors in score calculation; the findings suggest that even though the tGCS may be harder to use in the field (and may be scored inaccurately more frequently than more simplified scales), that it is still associated with slightly higher discrimination. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Public Reviewer #3 Douglas Kupas Geisinger Health System** | Discussion | It seems that your analysis should also have reviewed the accuracy of total GCS calculation in the same way as the results of the primary comparison. If so, the conclusion that could be drawn is that total GCS is too complex and fraught with error to be trusted for field triage by EMS providers. Those of us who take radio reports from EMS providers regarding traumatized patients know how often the reported GCS is a guess rather than a calculated assessment. | Estimates of discrimination incorporate any effects of errors in score calculation. The findings suggest that even though the tGCS may be harder to use in the field (and may be scored inaccurately more frequently than more simplified scales), that it is still associated with slightly higher discrimination. |
| **TEP Reviewer #1** | Clarity and Usability | Superbly written | Thank you. |
| **TEP Reviewer #2** | Clarity and Usability | Clarity and Usability: Is the report well-structured and organized? Yes | Thank you. |
| **TEP Reviewer #2** | Clarity and Usability | Are the conclusions relevant to policy or practice decisions? Yes, they further the science and call for more research in field triage | Thank you. |
| **TEP Reviewer #2** | Clarity and Usability | Do they contribute new information or understanding? Yes through a thorough examination of the data around this issue; does not provide definitive clinical guidance. | Thank you, noted. |
| **TEP Reviewer #3** | Clarity and Usability | Consider having someone go through the entire document to reduce unnecessary use of adjectives or qualifiers (e.g., 'slightly' and 'however'). Removing some text would potentially lead to a cleaner read. | Our copy editor reviewed the text and made edits to improve the readability, as suggested. |
| **TEP Reviewer #4** | Clarity and Usability | Abstract is missing key information in the methods regarding the data synthesis and rating. | We revised the Abstract to note that strength of evidence was determined based on the risk of bias, consistency, directness, precision, and reporting bias. |
| **TEP Reviewer #4** | Clarity and Usability | In the executive summary, consider bullet pointing the results similar to the "Key Points" section in the main document; easier for the reader. | In the final report, the Results in the Executive Summary are bulleted. |
| **TEP Reviewer #4** | Clarity and Usability | In both the executive summary and the main document, information regarding considerations for the key questions are presented before the actual Key Questions are enumerated, which was difficult to follow. I would introduce the key questions, and then provide the clarifying information regarding the key questions (Move ES-2 lines 51-58 and ES-3 lines 3-24; Pg 3 lines 12-42 behind the key questions). | We moved the Key Questions to come before the explanatory text, as suggested. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **TEP Reviewer #4** | Clarity and Usability | Page 7, line 14 under Population and Conditions of Interest, I believe the authors meant to state "in which the proportion with trauma was less than 10 percent..." | This sentence was changed to clarify that we excluded studies in which the proportion without trauma was >10% |
| **TEP Reviewer #4** | Clarity and Usability | Page 8, line 50-52, "No study met inclusion criteria for key questions 2 or 3" are results, and should be moved to the results section. | This is noted here because otherwise we would need to describe methods for assessing risk of bias for Key Questions 2 and 3. |
| **TEP Reviewer #4** | Clarity and Usability | Figure 2 Literature Flow Diagram, perhaps consider stating the key questions in parenthesis in the flow diagram in abbreviated form so the figure can stand alone. | In order to save space and preserve clarity, we opted not to add the words of the Key Questions to the Figure. Instead, we present the Key Questions in the text for readers that wish to refer to them. |
| **TEP Reviewer #4** | Clarity and Usability | Table 2, page 21, percent sign missing from 6th row, 4th column results. | This was added, thank you. |
| **TEP Reviewer #4** | Clarity and Usability | In tables would consider reporting the country of origin for the study. | We added the country for each study to Tables 3, 5, and 6. |
| **TEP Reviewer #6** | Clarity and Usability | The report is clear. I think the contribution is unclear mostly since the studies were and still are limited. | The Conclusions note, “The tGCS is associated with slightly greater discrimination than the mGCS or SMS for in-hospital mortality, receipt of neurosurgical interventions, severe brain injury, overall injury severity, and emergency intubation, with differences in the AUROC ranging from 0.01 to 0.05. The clinical significance of small differences in discrimination are likely to be small, and could be offset by factors such as convenience and ease of use.” |
| **Peer Reviewer #1** | Clarity and Usability | very clear and easy to read | Thank you. |
| **Peer Reviewer #2** | Clarity and Usability | While it could probably be tightening up slightly, this is well written and clear. Some people may have trouble following the clinical significance of such small AUROC numbers but you have tried to make these relatable. | Thank you, noted. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #2** | Clarity and Usability | I suppose I would say in general a 250 page report on GCS scores is rather onerous to go through; you will have to be very careful to include your key findings in the ES as very few people will spot details that are tucked away deep in the full report. Further formatting or break-out boxes may aid in translating key findings to readers. This is very important; knowledge translation will be tricky without easy-to-read summaries. Consider releasing this report with an infographic that highlights the key findings, challenges and implementation findings. I don't know any EMS leaders who will read the entire document, but many agencies would reinforce practice or make changes based on a catchy, easy-to-read infographic released concurrently. I don't mean to discourage you as you have done a truly exceptional job with this document, but your focus must now turn to translation. See this link for ideas as to how you can translate very large, scientific, stats-heavy documents into readable, actionable graphics. http://canadiem.org/the-top-five-changes-project-2015-aha-guidelines-update-cpr-ecc-infographic-series/. | Thank you for this link. We are very interested in making the information in our reports more understandable and usable.<br><br>With this in mind, we have reviewed the Abstract and the Executive Summary, which summarize the key points more succinctly than the main report. We also plan to submit a manuscript based on this report for publication in a journal, and AHRQ may commission the Eisenberg Center to develop translational products aimed at clinicians, policymakers, and/or consumers. We have also added a Box with Key Message on the title page. |
| **Peer Reviewer #4** | Clarity and Usability | The report is well presented. | Thank you. |
| **Peer Reviewer #5** | Clarity and Usability | The report is well structured, well written and well, just voluminous. | Thank you, noted. |
| **Peer Reviewer #3** | General comments to authors | Substantial copy-editing is needed. There are numerous singular/plural mismatches, a number of missing words, and several ordinary nouns that do not need to be capitalized (such as Emergency Medical Services). There are also several instances where an acronym is defined, having already been defined earlier. For example, “TBI” is defined on page 26, line 18 but then defined again on page 32, line 33. “GCS” is defined on page 33, line 14 having been defined earlier. | Thank you. We have done additional copy editing and have made the suggested corrections. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #3** | General comments to authors | The word “curve” is missing when the term “area under the receiver operator characteristic [AUROC]” is used. It is the “area under the ROC curve” that is of interest. | This was corrected. |
| **Peer Reviewer #3** | General comments to authors | Is “clinical decisionmaking” the same as “over- and under-triage”? The “Objectives” section of the Executive Summary lists “clinical decisionmaking” as the fourth of the five study objectives, and then the term appears in KQ 2a, which seems to be referring to the over- and under-triage of KQ 2 – but then does not appear again in the paper (e.g. in the “Discussion” section). | For Key Question 2, we focused on destination decisions as a result of triage as measured by rates of over- and under-triage. The term “clinical decisionmaking” was inadvertently left in Key Question 2a; we had previously revised Key Question 2 to more specifically address the outcome “over- and under-triage.” We replaced “clinical decisionmaking” with “over- and under-triage” in Key Question 2a to be consistent with Key Question 2. |
| **Peer Reviewer #3** | General comments to authors | Key Question 1 lists “e.g. receipt of intracranial monitoring within 48 hours of admission, receipt of surgery within 12 hours of admission, or early intubation” as utilization indicators of severe injury – but only intubation is actually shown in the results section and table. There is “neurosurgical intervention, overall” but nothing for “receipt of surgery [presumably not just neurosurgery] within 12 hours of admission” or “receipt of intracranial monitoring within 48 hours of admission” (though one study, by Acker et al, is cited in the table which examined this). Were the items in the Key Question listed a priori, but then found not to be studied? | As described in the Methods, the outcomes were decided a priori. The data from the Acker study is provided in the Table but is not mentioned in the text because it was the only study that reported this outcome. Surgery was intended to focus on neurosurgery; we revised the text to clarify this. |
| **Peer Reviewer #3** | General comments to authors | It is noted that no study was found evaluating how using tGCS vs mGCS or SMS would affect clinical outcomes. I’m not surprised: I frankly can’t think of how such a study would be conducted, or even what question it would ask. | We think that a cohort study could evaluate clinical outcomes in patients who underwent field triage with the tGCS vs. a more abbreviated scale, controlling for potential confounders. Given the likely clinical equipoise, we think an RCT would also be justified. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #3** | General comments to authors | The "Detailed Synthesis" and "Detailed Assessment" (variously labelled – should be consistent) sections are unnecessarily long. Most of the information between pages 39 and 81 could be shown in tables, which would be much simpler to read and digest. | We changed "Detailed Assessment" to "Detailed Synthesis" throughout the paper for consistency. There are a number of summary tables in this section already; the purpose of the "Detailed Synthesis" text is to provide more extensive textual detail. The information is also summarized in bullet points at the beginning of each section. We have added a Box on the title page summarizing the key points of the report. |
| **Peer Reviewer #3** | Specific comments to authors | Page 11, lines 8-9: Why should decisions be based on how field triage instruments compare to tGCS? Is tGCS really a criterion standard? I don't think so - | We revised the text so that this sentence focuses only on the comparison of tGCS with simplified versions of the tGCS, rather than tGCS vs. field triage instruments in general: "Decisions regarding the use of the tGCS versus more simplified versions of the tGCS should be based on their relative performance." |
| **Peer Reviewer #3** | Specific comments to authors | Page 11, lines 28-29 and page 85, lines 25-26: The current (since 2005 – see the National EMS Scope of Practice Model published by NHTSA) "levels" of EMS personnel are Emergency Medical Responder, Emergency Medical Technician, Advanced Emergency Medical Technician, and Paramedic. Not sure why citation #24 (about lactate) is used here to support the definition of EMT. | We removed the reference to the Neville article which was an error; we cite the National Emergency Medical Services Workforce Data Definition document from NHTSA and revised the description of EMS personnel to be consistent with the 2005 Scope of Practice Model. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #3** | Specific comments to authors | Page 12, line 15 (and page 32, line 51): Proportions transferred to higher and lower levels care are very poor estimates over under- and over-triage, respectively. Particularly in the case of over-triage, no higher-level trauma center routinely transfers minimally injured patients to lower-level centers; instead, these patients are simply treated and released. Citation #39 (Lerner) does not support this concept – not sure why it is cited here. I would consider this a fatal flaw in methodology, except that no studies were found examining this in the context of tGCS, mGCS, or SMS. | We revised the Scope and Methods sections to be clearer that we assessed rates of over- or under-triage based on the proportion of patients that were transported to a higher or lower than appropriate level of care based on a standardized definition for trauma center need (the Lerner reference discusses a criterion-based standard for trauma center need). We also included rates of transfer of care as a marker of over- or under-triage since we anticipated that few studies would evaluate over- or under-triage using a standardized definition for trauma center need. We also revised to note that interpretation of transfer rates is challenging because over-triaged patients may be discharged directly home. As the reviewer notes, no studies compared rates of over- or under-triage using either of these definitions. We also revised the Future Research Needs section to clarify that studies should utilize standardized, validated measures to determine the appropriateness of transport and triage decisions. |
| **Peer Reviewer #3** | Specific comments to authors | Page 14, line 21: Should this be “suspected OR undetected”, not “suspected OF undetected”? | This typo has been corrected. |
| **Peer Reviewer #3** | Specific comments to authors | Page 32, lines 12-15: I do not understand the last of the exclusion criteria: “studies of patients with and without trauma in which the proportion without trauma was less than 10 percent and results were not reported separately for patients with trauma.” What about a study with 50 or 90 percent nontrauma –why exclude one with <10% nontrauma but not one with significantly more nontrauma? Should it maybe be “proportion WITH trauma was less than 10 percent”? | We corrected to say that we excluded studies in which >10% of patients did not have trauma. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| **Peer Reviewer #3** | Specific comments to authors | Page 73, line 34: The "National Trauma Triage Protocol" does not actually exist – the "field triage decision scheme" (2006 ACS term) evolved to the "guidelines for field triage of injured patients (2011 CDC term) – the authors of citation #54 appear to have made up the term "NTTP" on their own, despite citing the 2011 CDC MMWR article with the correct new term. Probably best to use the current term. | We replaced the term "National Trauma Triage Protocol" with "2011 CDC guidelines for field triage." |
| **Peer Reviewer #3** | Specific comments to authors | Page 85, line 13: GCS was not originally developed for assessment of TBI, but instead for assessment of altered mentation of any source. Teasdale & Jennett, in citation #11, specifically discuss how "a wide range of conditions may be associated with coma or impaired consciousness". | As stated by Teasdale & Jennett in the 1974 Lancet article that is cited (reference #11), the GCS was developed for assessment of patients with acute brain injury. It has been widely applied to other situations associated with impaired consciousness, but that was not what it was originally developed for. We added a citation by Teasdale et al. from 2014 that reviews the history of the GCS in additional detail. |
| **Public Reviewer #3 Douglas Kupas Geisinger Health System** | References | I have uploaded an article that my colleagues and I recently published related to this specific question. Kupas DF, Melnychuk EM, Young AJ. Glasgow coma scale motor component ("Patient does not follow command") performs similarly to total Glasgow coma scale in predicting severe injury in trauma patients. Ann Emerg Med, 2016 published online. http://dx.doi.org/10.1016/j.annemergmed.2016.06.017 Since this article has almost 400,000 patients to add to this review, we are hopeful that its results can be included. | Thank you. We have added this article to our review and updated our Results accordingly. |