Emergency Medicine Journal Club
Wednesday, October 26, 2016

This journal club is sponsored by RCRMC, and will be held in the upstairs room at Mu Restaurant (309 W. State Street, Redlands) on October 26th from 6 to 9 pm. The topic is:

Glasgow Coma Scale: Origin, Application, and Limitations

The GCS has been a fundamental part of the culture of prehospital care and emergency medicine for decades. Where did it come from? Why is it so popular? How is it useful? What are its limitations? What can we learn from the GCS that applies to other scores and scales? The following individuals have been assigned to present articles (maximum of 5-10 minutes each). Please read the articles (particularly those to be presented by residents) and to be prepared to critically discuss them! Bring an electronic or paper copy with you.

<table>
<thead>
<tr>
<th>Page</th>
<th>Presenter: Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Steve Green: Sasser. CDC Guidelines for Field Triage of Injured Patients: Recommendations of the National Expert Panel on Field Triage, 2011</td>
</tr>
</tbody>
</table>
Described uses for the GCS

- originally described for monitoring depth of coma over time in a neurosurgical unit
- categorises severity of traumatic brain injury into mild (13-15), moderate (9-12) and severe (8 or less)
- used in Brain Trauma Foundation guidelines as part of the indications for ICP monitoring (e.g. GCS 8 or less and abnormal CT head)
- used for determining the need for CT head in TBI by tools such as the Canadian CT Head Rule (GCS<15 at 2 hours post-injury)
- Traditional ATLS mantra is “GCS 8, intubate”
- used in APACHE II

CDC Guidelines for Field Triage of Injured Patients: Recommendations of the National Expert Panel on Field Triage, 2011

The Panel recommended transport to a facility that provides the highest level of care within the defined trauma system if any of the following are identified:

**Step One: Physiologic Criteria**

- **Glasgow Coma Scale ≤13,** or
- SBP of <90 mmHg, or
- respiratory rate of <10 or >29 breaths per minute (<20 in infant aged <1 year), or need for ventilatory support.

**Step Two: Anatomic Criteria**

- all penetrating injuries to head, neck, torso, and extremities proximal to elbow or knee;
- chest wall instability or deformity (e.g. flail chest);
- two or more proximal long-bone fractures;
- rushed, degloved, mangled, or pulseless extremity;
- amputation proximal to wrist or ankle; pelvic fractures;
- open or depressed skull fractures; or paralysis.
significant correlations between surface area or weight and B.T.T.

Discussion

Japanese migrants to Hawaii—especially their families, the second-generation Nisei—develop disease patterns similar to Caucasians. The Hawaii-Japanese lose the high gastric-cancer risks of Japan and acquire “Western” malignancies such as those of prostate, breast, and colon. The Hawaii-Japanese have more colonic polyps and diverticuloses than Japanese living in Japan.

It would be expected, then, that the more traditional Issei would have a faster B.T.T. than the Nisei and that the Hawaii-Japanese would have transit-times similar to those of Caucasians. We were surprised to find no differences between the Issei and Nisei, and the Hawaii-Japanese had rapid transit-times comparable with those in rural Africans. The B.T.T. differences between the Japanese and Caucasians could not be explained by education, occupation, body-weight, surface area, or the number of bowel movements per day.

With respect to the Hawaii-Japanese experience, B.T.T.s do not seem to be related to the pathogenesis of colonic disease.

We thank Mr Harry Ito of Kuakini Hospital for engineering the stool-collection equipment; and Dr Edgar Childs, Dr Donald Ikeda, and Dr David Sakuda of the Department of Radiology, Kuakini Hospital, for their technical assistance. This work was supported in part by National Institutes of Health contracts E-71-2170 and PH-43-65-1003-C.

Requests for reprints should be addressed to the Japan-Hawaii Cancer Study, Kuakini Hospital, 347 North Kuakini Street, Honolulu, Hawaii 96817, U.S.A.

REFERENCES


ASSESSMENT OF COMA AND IMPAIRED CONSCIOUSNESS

A Practical Scale

GRAHAM TEASDALE BRYAN JENNETT
University Department of Neurosurgery, Institute of Neurological Sciences, Glasgow G51 4TF

Summary

A clinical scale has been evolved for assessing the depth and duration of impaired consciousness and coma. Three aspects of behaviour are independently measured—motor responsiveness, verbal performance, and eye opening. These can be evaluated consistently by doctors and nurses and recorded on a simple chart which has proved practical both in a neurosurgical unit and in a general hospital. The scale facilitates consultations between general and special units in cases of recent brain damage, and is useful also in defining the duration of prolonged coma.

Introduction

A wide range of conditions may be associated with coma or impaired consciousness. Apart from acute brain damage due to traumatic, vascular, or infective lesions, there are metabolic disorders such as hepatic or renal failure, hypoglycaemia or diabetic ketosis, and drug overdose. In gauging deterioration or improvement in the acute stage of such conditions, as well as in predicting the ultimate outcome, the degree and duration of altered consciousness usually overshadow all other clinical features in importance. It is therefore vital to be able to assess and to record changing states of altered consciousness reliably.

Need for a Clinical Scale

Impaired consciousness is an expression of dysfunction in the brain as a whole. This may be due to agents acting diffusely, such as drugs or metabolic imbalance; or to the combination of remote and local effects produced by brain damage which was initially focal. Such focal brain damage may affect some of the responses which are used to assess the level of consciousness, and any scale devised for general use must allow for this possibility. A simpler scale might suffice for metabolic or drug coma, when the likelihood of structural brain damage is small, but in an emergency there may be insufficient information to assign patients confidently to a particular diagnostic category. Moreover, coma of mixed origin is not uncommon, as when head injury is suspected of being associated with ingestion of drugs or alcohol, or with a vascular accident. These seem good reasons for devising a generally applicable scheme of assessment.

Existing Systems

The development of equipment for monitoring various functions in critically ill patients has not altered the need for doctors and nurses to assess the level of consciousness. There is an abundance of alternative terms by which levels of coma or impairment of consciousness are described and recorded. Systems for describing patients with impaired consciousness are not consistent. Indeed, many clinicians retreat from any formal scheme in favour of a general description of the patient's state, without clear guidelines as to what to describe and how to describe it.

In practice, such unstructured observations commonly result in ambiguities and misunderstandings when information about patients is exchanged and when groups of patients treated by alternative methods are compared, or reported from different centres. There is no general agreement about what terms to use, nor are those in common use interpreted similarly by different workers. Most report of patients in coma offers yet another classification. Most wide the spectrum of altered consciousness into a series of steps, which in the reports we reviewed ranged from 3 to 17 and were often described in terms which defied clear definition. Many assume the existence of constellations of clinical features which are unique to each "level", whilst...
others distinguish between coma and consciousness on the basis of only one aspect of behaviour.

The importance of careful and complete neurological examination in determining the nature and site of the lesion causing coma has been described at length by Fisher and Plum and Posner, who emphasised that tests of brainstem function, not usually included in routine examination, can be useful in the diagnosis of stupor or coma. Neither, however, was primarily concerned with repeated bedside assessment of the degree of conscious impairment, which is the subject of our paper.

**Glasgow Coma Scale**

To be generally accepted, a system must be practical to use in a wide range of hospitals and by staff without special training. But the search for simplicity must not be the excuse for seeking absolute distinctions where none exist: for that reason no attempt is made to define either consciousness or coma in absolute terms. Indeed, it is conceptually unsound to expect a clear watershed in the continuum between these states. What is required instead is an effective method of describing the various states of impaired consciousness encountered in clinical practice. Moreover, this should not depend on only one type of response because this may, for various reasons, be untestable. The three different aspects of behavioural response which we chose to examine were motor response, verbal response, and eye opening, each being evaluated independently of the other. These feature in many previous reports on coma but not in the formal system we propose. This depends on identifying responses which can be clearly defined, and each of which can be accurately graded according to a rank order that indicates the degree of dysfunction.

**Motor Responses**

The ease with which motor responses can be elicited in the limbs, together with the wide range of different patterns which can occur, makes motor activity a suitable guide to the functioning state of the central nervous system. Indeed, every one of the reported scales which we reviewed included some aspect of motor responsiveness as a criterion.

*Obeying commands* is the best response possible, but the observer must take care not to interpret a grasp reflex or postural adjustment as a response to command. The terms "purposeful" and "voluntary" are avoided because we believe that they cannot be judged objectively.

If there is no response to command, a painful stimulus is applied. The significance of the response to pain is not always easy to interpret unless stimulation is applied in a standard way and is maintained until a maximum response is obtained. Initially pressure is applied to the fingernail bed with a pencil; this may result in either flexion or extension at the elbow. If flexion is observed stimulation is then applied to the head and neck and to the trunk to test for localisation. In brain death, a spinal reflex may still cause the legs to flex briskly in response to pain applied locally.

For this reason, and because the arms show a wider range of responses, it is wise always to test them, unless local trauma makes this completely impossible.

*A localising response* indicates that a stimulus at more than one site causes a limb to move so as to attempt to remove it.

*A flexor response* may vary from rapid withdrawal, associated with abduction of the shoulder, to a slower, stereotyped assumption of the hemiplegic or decorticat posture with adduction of the shoulder. Experienced observers may readily distinguish between normal and abnormal flexion, but for general use in the first few days after brain damage has been sustained it is sufficient to record only that the response is flexor.

*Extensor posturing* is obviously abnormal and is usually associated with adduction, internal rotation of the shoulder, and pronation of the forearm. The term "decerebrate rigidity" is avoided because it implies a specific physiopathological correlation.

*No response* is usually associated with hypotonia and it is important to exclude spinal transection as an explanation for lack of response; and also to be satisfied that an adequate stimulus has been applied.

When recording motor response as an indication of the functional state of the brain as a whole, the best or highest response from any limb is recorded. During a single examination some patients give variable responses, these usually becoming better as the patient becomes more aroused; responses from the right and left limbs may also differ. Any difference between the responsiveness of one limb and another may indicate focal brain damage and for this purpose the worst (most abnormal) response should be noted. But for the purpose of assessing the degree of altered consciousness it is the best response from the best limb that is recorded.

**Verbal Responses**

Probably the commonest definition of the end of coma, or the recovery of consciousness, is the patient's first understandable utterance; speech figured in nearly all the reported scales which we reviewed. Certainly the return of speech indicates the restoration of a high degree of integration within the nervous system, but continued speechlessness may be due to causes other than depressed consciousness (e.g., tracheostomy or dysphasia).

*Orientation* implies awareness of the self and the environment. The patient should know who he is, where he is, and why he is there; know the year, the season, and the month. The words "rational" and "sensible" are avoided because they cannot be clearly defined.

*Confused conversation* is recorded if attention can be held and the patient responds to questions in a conversational manner but the responses indicate varying degrees of disorientation and confusion. It is here that verbatim reporting of the individual patient's responses can be useful.

*Inappropriate speech* describes intelligible articulation but implies that speech is used only in an exclamatory or random way, usually by shouting and swearing; no sustained conversational exchange is possible.

*Incomprehensible speech* refers to moaning and groaning but without any recognisable words.
Eye Opening

Spontaneous eye opening, with sleep/wake rhythms, is most highly scored on this part of the scale and it indicates that the arousal mechanisms in the brainstem are active. But arousal does not imply awareness, and we believe it is wise to try to decide whether a patient is attentive on the basis of eye movements. Patients in the persistent vegetative state,17 who are subsequently shown to be structurally decorticate, have often been believed by relatives, nurses, and even by doctors to be reacting visually to people around them; probably primitive ocular-following reflexes may be executed at subcortical level.

Eye opening in response to speech is a response to any verbal approach, whether spoken or shouted, not necessarily the command to open the eyes.

Eye opening in response to pain should be tested by a stimulus in the limbs, because the grimacing associated with supraorbital or jaw-angle pressure may cause eye closure.

Practical Applications of the Scale

Different observers were able to elicit the responses in this scale with a high degree of consistency, and the likelihood of ambiguous reporting appears to be small. This was demonstrated by having several doctors and nurses examine the same group of patients. Disagreements were rare.18 This was in pronounced contrast to what happened when the observers were asked instead to judge only whether patients were conscious or unconscious; one in five observers then disagreed with the majority opinion. This 20% disagreement-rate compared with rates of 20–35% which have been reported in various different clinical situations,19 whilst in one study extensor plantar responses showed only 50% consistency when observations were repeated.20

One or other components of this scale may be untestable, and this fact can be recorded. Limbs may be immobilised by splints for fractures, tracheostomy may preclude speech, and eyelid swelling or bilateral third-nerve lesions make eye opening impossible.

In the rare "locked-in syndrome," a patient with totally inactive limbs may obey commands to move the eyes and may even be able to signal his needs.21

The nurses in our intensive-care unit have willingly adopted this method of formalising observations which they previously used to record as a descriptive comment. They now plot them on a chart (see accompanying figure) somewhat similar in format, but not content, to one proposed by Bouzarth,2 and which also provides for conventional recording of temperature, pulse and respiration, of the pupil size in mm, and of focal motor signs. This method has already been adopted successfully for making observations on head injuries in a neighbouring general hospital. In such hospitals patients with head injuries form a considerable proportion of acute surgical admissions, and observations there depend on medical and nursing staff who have no special experience of neurology and neurosurgery.

Discussion

Apart from its practical use in the management of recently brain-damaged patients, this scale allows the duration of coma to be defined more precisely, in terms of how long different levels of responsiveness have persisted. There is evidence that this is a crucial criterion when it comes to predicting the ultimate outcome of coma, particularly after head injury.22 It would make it possible also to examine critically claims for good recovery after weeks or months "in coma," by enabling the alleged coma to be more accurately assessed. In such cases as we have scrutinised, it has been clear, even retrospectively, that there had been evidence of much earlier recovery, on at least one component of the coma scale, than had been recognised. By resolving the problem of defining "prolonged coma" the scale also makes it possible to distinguish between the various states which this term embraces, such as akinetic mutism and the persistent vegetative state.27

Some may have reservations about a system which seems to undervalue the niceties of a full neurological examination. It is no part of our case to deny the value of a detailed appraisal of the patient as a whole, and of neurological function in particular, in reaching a diagnosis about the cause of coma, or in determining the probable site of brain damage. However, repeated observations of conscious level are usually made by relatively inexperienced junior doctors or nurses; these staff are not only few in number but they change frequently even during the course of a day. There are therefore good reasons for restricting routine observations to the minimum, and for choosing those which can be reliably recorded and understood by a range of different staff.

We are grateful to many colleagues for their assistance in developing this scale; particularly Dr Fred Plum of New York, Dr Reinder Breukman of Rotterdam, and Dr David Shaw of Newcastle upon Tyne, in whose units its practical value has also been confirmed. We thank the consultants of the division of neurosurgery, Glasgow, for their cooperation. This scale was devised as part of a study of severe head injuries supported by the National Fund for Research into Crippling Diseases.

Requests for reprints should be addressed to B. J.

REFERENCES
Interrater Reliability of Glasgow Coma Scale Scores in the Emergency Department

Study objective: Emergency physicians often use the Glasgow Coma Scale (GCS) to help guide decisions in patient care, yet the reliability of the GCS has never been tested in a typical broad sample of emergency department (ED) patients. We determined the interrater reliability of the GCS between emergency physicians when adult patients with altered levels of consciousness are assessed.

Methods: In this prospective observational study at a university Level I trauma center, we enrolled a convenience sample of ED patients older than 17 years who presented with an altered level of consciousness. Two residency-trained attending emergency physicians independently assessed and recorded the GCS score and its components (eye, verbal, and motor) in blinded fashion within a 5-minute period. Data were analyzed for interrater reliability by using standard ordinal calculations. We also created scatter plots and Bland-Altman plots for each GCS component and for the GCS score.

Results: One hundred thirty-one patients were screened and enrolled in the study, with 15 excluded because of protocol violations. Of the 116 remaining patients, the agreement percentage for exact total GCS was 32% (τ-b=0.739; Spearman ρ=0.864; Spearman ρ^2=75%). Agreement percentage for GCS components were eye 74% (τ-b=0.715; Spearman ρ=0.757; Spearman ρ^2=57%), verbal 55% (τ-b=0.587; Spearman ρ=0.665; Spearman ρ^2=44%), and motor 72% (τ-b=0.742; Spearman ρ=0.808; Spearman ρ^2=65%). Our Spearman’s analyses found that only approximately half (44% to 65%) of the observed variance could be explained by the relationship between the paired component measures. For GCS components, only 55% to 74% of paired measures were identical, and 6% to 17% of them were 2 or more points apart.

Conclusion: We found only moderate degrees of interrater agreement for the GCS and its components.

INTRODUCTION

Background

Since its creation in 1974 by Teasdale and Jennett\(^1\) for the purpose of assessing the depth and duration of impaired consciousness and coma, the Glasgow Coma Scale (GCS) has become a universal tool for the quantification of altered level of consciousness. The GCS is increasingly being tested for its ability to predict outcomes from traumatic brain injury,\(^2\) nontraumatic coma,\(^25-27\) cerebrovascular events,\(^16\) cardiac arrest,\(^28\) and toxic ingestions.\(^29\) Classification schemes and prediction trees for head injury have been developed according to GCS scores,\(^12,23,30\) with radiographic studies and ultimately patient treatment based on the patient's GCS score.\(^12,21,23,31-33\)

Importance

The Brain Trauma Foundation recently stated that the GCS can be "fairly reliably measured by trained medical personnel"\(^26\); however, we found no evidence that it has ever been evaluated in a typical broad sample of emergency department (ED) patients with an altered level of consciousness. A single ED-based study included only patients with a GCS score of 13 or greater.\(^32\) Other authors have studied the reliability of the GCS in the ICU\(^34-39\) or through the use of videotaped assessments,\(^36,40\) but these studies are unlikely to be applicable to the unique environment of the ED. Because the GCS is universally applied and incorporated into trauma and life support education programs,\(^41,42\) it is important for this scale to be valid and reliable.

Goals of This Investigation

We wished to quantify the interrater reliability of the GCS and its components in adult patients presenting to the ED with altered levels of consciousness.

MATERIALS AND METHODS

Study Design and Setting

This prospective observational study examined the interrater reliability of the GCS at a Level I trauma center with an annual ED census of 53,000 patients. Patients were entered into the study between March 2002 and March 2003. This study was approved by the hospital's institutional review board.

Selection of Participants

Patients were included if they were older than 17 years and were judged by an enrolling physician to have an altered level of consciousness because of either medical or traumatic illness. Exclusion criteria were the administration of paralytic agents within 1 hour of assessment or the failure of both physician-observers to complete their evaluations of the patient within a 5-minute period.

Patients were enrolled by convenience sampling, with every attempt to enroll them at the earliest possible time after arrival at the ED. Two residency-trained attending emergency physicians (as available on a convenience basis) independently assessed the GCS within 5 minutes of each other while blinded to each other's scores. Physicians assessed and recorded GCS scores in accordance with their standard practice, did not receive notice before the study, and were intentionally not provided with any additional training or instruction in this assessment. They recorded separate GCS components on data forms listing GCS score categories (eye, verbal, and motor) together with their standard currently accepted definitions (Figure 1).\(^43\)

Methods of Measurement

We assessed interrater reliability separately for the eye, verbal, motor, and total GCS scores. First, we calculated agreement percentage and the unweighted and weighted k statistic. The unweighted k treats agreement in a binary (perfect agreement versus no agreement) manner and therefore weights all misses equally, where-
as the weighted κ adjusts the calculation (being off by 1 GCS point is not as bad as being off by ≥2 points). We then calculated Kendall’s τ-b, created scatter plots, and calculated Spearman’s ρ. We also report the confidence intervals (CIs) of Spearman’s ρ and the square of this ρ (i.e., the percentage of the observed variance explained by the relationship between the paired measures). We then created histograms and Bland-Altman plots44 of the differences between paired scores. Given the possibility that intubation might adversely affect the reliability of verbal scores,45-48 we performed a similar subset analysis of this measure in nonintubated patients.

Data Analysis and Processing

All analyses were performed with Stata 8 software (Stata Corporation, College Station, TX), except for CIs around the Spearman ρ estimates. These were calculated with JACKBOOT, a bootstrapping macro in SAS software (version 8.2, SAS Institute, Inc., Cary, NC) that uses a bias-corrected accelerated method.

RESULTS

A total of 131 patients were screened and enrolled in the study. We then excluded 15 protocol violations (1 underage patient and 14 with paired ratings >5 minutes apart). The remaining 116 patients (71 men, 45 women) are described in Table 1. Nineteen attending emergency physicians participated in the study.

Eighteen (16%) of the 116 patients were intubated at their assessments; in all circumstances, both raters assigned a verbal score of 1 for these patients.

Measures of interrater reliability demonstrated moderate levels of agreement (Table 2), with approximately half (44% to 65%) the observed variance explained by the relationship between the paired component measures. Scatter plots are shown in Figure 2, and differences between paired scores are displayed in Figures 3 and 4.

Table 1. Patient characteristics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
<th>GCS Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of patients</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>Median age, y (range)</td>
<td>54 (18–90)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>71 (61)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>45 (39)</td>
<td></td>
</tr>
<tr>
<td>Intubated at assessment</td>
<td>18 (16)</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>32 (28)</td>
<td>3–15</td>
</tr>
<tr>
<td>Infection/sepsis</td>
<td>19 (16)</td>
<td>3–15</td>
</tr>
<tr>
<td>Metabolic disturbance</td>
<td>16 (14)</td>
<td>3–15</td>
</tr>
<tr>
<td>Drug or alcohol intoxication</td>
<td>11 (9)</td>
<td>3–14</td>
</tr>
<tr>
<td>Cerebrovascular accident</td>
<td>10 (9)</td>
<td>3–15</td>
</tr>
<tr>
<td>Nontraumatic intracranial hemorrhage</td>
<td>7 (6)</td>
<td>4–15</td>
</tr>
<tr>
<td>Seizure disorder</td>
<td>6 (5)</td>
<td>3–14</td>
</tr>
<tr>
<td>Dementia</td>
<td>4 (3)</td>
<td>7–15</td>
</tr>
<tr>
<td>Other*</td>
<td>11 (9)</td>
<td>3–15</td>
</tr>
<tr>
<td>Head CT scan results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head CT scan not done</td>
<td>39 (34)</td>
<td>3–15</td>
</tr>
<tr>
<td>Intracranial hemorrhage</td>
<td>20 (17)</td>
<td>3–15</td>
</tr>
<tr>
<td>Normal</td>
<td>18 (16)</td>
<td>3–15</td>
</tr>
<tr>
<td>Atrophy</td>
<td>17 (15)</td>
<td>3–15</td>
</tr>
<tr>
<td>Infarction</td>
<td>11 (9)</td>
<td>3–15</td>
</tr>
<tr>
<td>Other†</td>
<td>11 (9)</td>
<td>3–15</td>
</tr>
</tbody>
</table>

*CT, Computed tomography.
†Included hypotension, shunt malfunction, hypoxia, dehydration, malignancy, and unknown etiology.
‡Included cerebral contusion, abnormal with no interval changes, and hydrocephalus.

Table 2. Measures of interrater reliability between paired ratings.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Eye</th>
<th>Verbal*</th>
<th>Motor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement, %</td>
<td>74</td>
<td>55</td>
<td>72</td>
<td>32†</td>
</tr>
<tr>
<td>Kendall’s τ-b</td>
<td>0.715</td>
<td>0.587</td>
<td>0.742</td>
<td>0.739</td>
</tr>
<tr>
<td>Spearman’s ρ</td>
<td>0.757</td>
<td>0.665</td>
<td>0.808</td>
<td>0.864</td>
</tr>
<tr>
<td>(95% CIs)</td>
<td>(0.612–0.851)</td>
<td>(0.519–0.765)</td>
<td>(0.700–0.877)</td>
<td>(0.804–0.904)</td>
</tr>
<tr>
<td>Spearman’s ρ², %</td>
<td>57</td>
<td>44</td>
<td>65</td>
<td>75</td>
</tr>
<tr>
<td>k, unweighted</td>
<td>0.59</td>
<td>0.37</td>
<td>0.58</td>
<td>0.00⁵</td>
</tr>
<tr>
<td>k, weighted⁶</td>
<td>0.72</td>
<td>0.48</td>
<td>0.63</td>
<td>0.40</td>
</tr>
</tbody>
</table>

*A subset analysis of this measure in the subset of nonintubated patients (n=98) yielded slightly less reliable results: percentage agreement 47%, k = 0.24 (“fair”), Kendall’s τ-b 0.452, Spearman’s ρ 0.515, ρ² 27%.
†Thirty-two percent had perfect agreement of all 3 score elements; an additional 6% had identical total GCS scores with differing components.
‡Pc .001 for all measures.
§The square of Spearman’s ρ represents the percentage of the observed variance explained by the relationship between the paired measures.
⁵Calculated with the requirement that all 3 components have agreement. If the overall agreement of the total is considered even when components disagree, the k is 0.29 or “fair.”
⁶Exact matches weighted 1.0, misses of 1 point weighted 0.5, and misses of ≥2 points weighted as 0. The calculation for total GCS is based on the overall agreement of the total even when components disagree.

Figure 1. The Glasgow Coma Scale.⁴³
assigned verbal scores of 1 in intubated patients, effectively enhancing (if only slightly) the reliability of this assessment for the total sample when compared with just the nonintubated patients (Table 2).

Thus, our data suggest that the GCS verbal score is not unreliable for intubated patients but instead is simply not applicable or helpful. We suggest that, for simplicity, this portion of the GCS be omitted in intubated patients.

A second potential limitation is that we intentionally chose not to provide our physicians with specific training on techniques of GCS score assignment because we wanted to duplicate the way the GCS score is currently assigned in actual practice. It is possible that a struct-

**LIMITATIONS**

One limitation of our study is that we chose to include intubated patients despite questions about the reliability of verbal score assessments in this patient subset. We included them with the intent of realistically mimicking the full spectrum of ED clinical scenarios. We noted that our physicians uniformly

---

**Figure 2.**
Plots of paired ratings (n=116). The smoothing line was generated by LOWESS (locally weighted least squares regression). The slight offset of data points (ie, jitter) is created solely to display data points that might otherwise overlap. **A,** Eye component; **B,** verbal component; **C,** motor component; **D,** total.
tured intensive training program might result in improved interrater reliability; however, given the simplicity and straightforwardness of this test, we believe that this result is unlikely.

Because our paired measures could occur up to 5 minutes apart, it is possible that some of our observed differences represent actual patient changes in function over time rather than interrater variation. However, we believe that appreciable changes within such a short span are unlikely.

A final consideration is that we performed our study in a clinically diverse population, including a mix of ages and diagnoses representative of our ED population. It is possible that the GCS demonstrates greater reliability when used in specific patient subsets or conditions.

---

**DISCUSSION**

We measured the interrater reliability of the GCS in a broad sample of ED adults with altered level of consciousness and were disappointed to find just moderate degrees of test reliability. Our Spearman's analyses found that only approximately half of the observed variance could be explained by the relationship between the paired component measures. For GCS components, only 55% to 74% of paired measures were identical, and 6% to 17% of them were 2 or more points apart.

Although our paired measures were naturally statistically associated, the magnitude of this association was substantially less than we would consider desirable.

Although we have also reported the \( \kappa \) statistics for the GCS and its components (Table 2), these values

---

**Figure 3.**

*Histograms of the difference between the paired ratings.* **A,** Eye component; **B,** verbal component; **C,** motor component; **D,** total.
must be interpreted with caution. The unweighted \( \kappa \) technique classifies differences of just a single point as complete disagreement, and accordingly its summary values are applicable only for the purposes of judging identical rather than partial agreement. When we viewed results in this fashion, we found that the total GCS exhibited a \( \kappa \) value of 0.00, which is consistent with random guessing. However, complete agreement may be too stringent a requirement for the GCS because 1- or 2-point differences in these scales may be considered clinically acceptable. Furthermore, the unweighted \( \kappa \) values we discovered contrast with the visual appearance of overall correlation seen in our scatter plots (Figure 2), which is likely because an unweighted \( \kappa \) counts all errors as equally bad (ie, being off by 1 is just as bad as being off by 10). Thus, the ordinal measures of reliability (ie, Kendall’s \( \tau-b \), Spearman) and weighted \( \kappa \) reported in Table 2 should be considered more accurate a depiction of GCS test performance than the unweighted \( \kappa \). A disadvantage of the weighted \( \kappa \), however, is that the results will differ, depending on the arbitrary weighting scheme selected.

An alternative manner of putting the current results in some sort of context is by comparing them with the reported reliability of other clinical features. The GCS components are approximately as reliable as pulmonary auscultation for rales (\( \kappa \)=0.24 to 0.65),\(^4^9\) the examination of tympanic membranes (\( \kappa \)=0.30),\(^5^0\) and murmur detection (\( \kappa \)=0.30 to 0.48).\(^5^1\) GCS components appear, in general, less reliable than the assessment of National

---

**Figure 4.**

Bland-Altman plots of absolute differences versus means. The slight offset of data points (ie, jitter) is created solely to display points that might otherwise overlap. **A**, Eye component; **B**, verbal component; **C**, motor component; **D**, total.
Emergency X-Radiography Utilization Study (NEXUS) cervical spine injury risk criteria (percentage agreement 83% to 96%; $k=0.58$ to 0.86) and the ascertain-ment of witnessed loss of consciousness from patient history ($k=0.83$). If this alternative and more forgiving standard of comparison is used, the GCS score perhaps does not seem as unreliable. However, some degree of imprecision must be tolerated in these other clinical features in that there is no better alternative for their assessment. The GCS, in contrast, is an arbitrary scale that is widely believed to enhance clinical prac-tice, and we believe that it should be held to a more rigorous standard.

Three sets of researchers have evaluated interrater reliability of the GCS in nonacute care settings by using a “disagreement rate” ranging from 0 to 0.5. Introduced in the original GCS study by Teasdale and Jennett, and no longer generally accepted as a measure of inter-rater reliability, disagreement rates are defined by 1 of the authors as “low” when they fall between 0 and 0.299 and high when between 0.3 and 0.5. Teasdale et al found that 7 neurosurgeons displayed low disagreement when assessing 12 ICU patients with the GCS according to this calculation (disagreement rates: eye=0.143, verbal=0.0054, and motor=0.109). Rowley and Fielding noted similar results in an unspecified number of neurosurgical ward patients (eye=0.005 to 0.163, verbal=0.000 to 0.035, motor=0.004 to 0.064) when evaluated by trained or untrained observers. For 7 videotaped ICU patients, Juarez and Lyons found that disagreement rates for individual GCS components varied, depending on the patient being evaluated (eye=0.00 to 0.29, verbal=0.00 to 0.06, and motor=0.00 to 0.22). Juarez and Lyons also reported in summary fashion that $k$ scores ranged from 0.39 to 0.79 for the total GCS but did not separately report values for GCS components.

Three other investigators have examined GCS scores in the ICU by using the more traditional $k$. Tessier et al report a $k$ of 0.35 between 2 physicians assessing the GCS in 46 ICU patients. Elliott reports a $k$ of 0.95 for GCS score agreement between 2 nurses for 22 intu-bated ICU patients. Lindsay et al report a $k$ of 0.69 between neurosurgery attending physicians and resi-dents for rating 15 patients with subarachnoid hemor-rhage into the following 3 categories: GCS score of 6 to 12, 13 to 14, and 15.

Our study is novel because, to our knowledge, it is the first to include a broad spectrum of patients with an altered level of consciousness, which is more likely to represent the typical ED population than are results from previous ICU studies. Accordingly, we believe that our results may be more applicable to typical ED prac-tice.

The widespread acceptance and implementation of the GCS is likely due to its face validity, because it merely quantifies those assessments of responsiveness that clinicians already routinely perform. However, the clinical utility of a test requires more than face validity, specifically, criterion-related validity, construct valid-ity, and content validity.

Criterion-related validity refers to the correspond-ence between the test in question and a reference crite-ron standard. GCS fails this best-available measure of true validity in that there is no criterion standard for the quantification of altered level of consciousness, and therefore it has been impossible to calculate the accu-racy of this test. Indeed, through default the GCS has effectively become the essential criterion standard for this purpose.

Construct validity refers to the consistency of one test with respect to another test that measures similar char-acteristics. In the case of the GCS, one would consider that it has construct validity in regard to demonstrated statistical associations with outcomes such as the Glasgow Outcome Score and the need for neurosurgical interventions. Despite this evidence, it is not clear that the GCS can accurately predict these specific outcomes in any given patient. In this manner, it may be similar to the WBC count, which, despite being statistically associated with appendicitis, has such a weak association that the test cannot be used to accurately guide clinical decisionmaking for specific patients.
Content validity refers to the ability of the test to directly assess the true qualities that it intends to measure. In the case of the GCS, this would be the ability to quantify the true functional level of the central nervous system. However, the GCS does not measure complete neurologic capacity but rather includes 3 practical surrogate markers that cannot be expected to appraise the full functional domain of the central nervous system. In fact, Teasdale and Jennett\(^4\) have stated that “we have never recommended using the GCS alone, either as a means of monitoring coma, or to assess the severity of brain damage or predict outcome.” Nonetheless, clinicians widely use the GCS for all of these things.

Despite its popularity and face validity, the GCS lacks the degree of criterion-related validity, construct validity, and content validity that would be necessary to consider it a truly valid clinical tool. Given the near-universal application of the GCS and its prominence in trauma and life support education programs,\(^{41,42}\) we must regard the GCS as a gross estimate of neurologic responsiveness rather than as a precise measurement. Our findings of only moderate interrater reliability may be regarded by some as insufficient to support the current widespread application of this technique for ED patient treatment and decisionmaking, as well as the prominence of this tool in trauma and life support education programs.

In retrospect, although our study found ample disagreement in GCS assessments between physicians, our methodology was unable to characterize whether such differences were caused by differential quality of physician assessment or just the inherent underlying subjectivity of these clinical measures. We could have explored these differences by having the 2 physicians review their scores with each other, reexamine the patient, and then report on the etiology of the discrepancy.

In conclusion, we found only moderate degrees of interrater agreement for the GCS and its components.

Received for publication April 24, 2003. Revisions received July 17, 2003; July 29, 2003; and August 5, 2003. Accepted for publication August 7, 2003.

Presented at the Society for Academic Emergency Medicine western regional meeting, Scottsdale, AZ, April 2003; and the Society for Academic Emergency Medicine national meeting, Boston, MA, May 2003.

The authors report this study did not receive any outside funding or support.

Address for correspondence: Michelle R. Gill, MD, Department of Emergency Medicine, Loma Linda University Medical Center, 11234 Anderson Street, PO Box 2000, Rm. A-108, Loma Linda, CA 92354; 909-558-4344, fax 909-558-0121; E-mail michelle_gill@yahoo.com.

REFERENCES


Author contributions: MRG and SMG conceived of and designed the study. MRG oversaw the data collection. SMG provided statistical advice on study design and analyzed the data. DGR assisted with data collection and manuscript development. All authors participated in drafting the manuscript and contributed to its revisions. MRG and SMG take responsibility for the paper as a whole.

We thank Jason S. Haukoos, MD, MS, for his assistance with statistical analysis and Renee Schroetlin for her assistance with data acquisition.
51. Etchells E, Bell C, Robb K. Does this patient have an abnormal systolic murmur? JAMA. 1997;277:584-57.
Randomized Controlled Trial of a Scoring Aid to Improve Glasgow Coma Scale Scoring by Emergency Medical Services Providers

Amanda Feldman, BS; Kimberly W. Hart, MA; Christopher J. Lindsell, PhD; Jason T. McMullan, MD*

*Corresponding Author. E-mail: jason.mcmullan@uc.edu.

Study objective: Emergency medical services (EMS) personnel frequently use the Glasgow Coma Scale (GCS) to assess injured and critically ill patients. This study assesses the accuracy of EMS providers’ GCS scoring, as well as the improvement in GCS score assessment with the use of a scoring aid.

Methods: This randomized, controlled study was conducted in the emergency department (ED) of an urban academic trauma center. Emergency medical technicians or paramedics who transported a patient to the ED were randomly assigned one of 9 written scenarios, either with or without a GCS scoring aid. Scenarios were created by consensus of expert attending emergency medicine, EMS, and neurocritical care physicians, with universal consensus agreement on GCS scores. \( \chi^2 \) and Student’s t tests were used to compare groups.

Results: Of 180 participants, 178 completed the study. Overall, 73 of 178 participants (41%) gave a GCS score that matched the expert consensus score. GCS score was correct in 22 of 88 (25%) cases without the scoring aid. GCS was correct in 51 of 90 (57%) cases with the scoring aid. Differences in accuracy were most pronounced in scenarios with a correct GCS score of 12 or below. Subcomponent accuracy was eye 62%, verbal 70%, and motor 51%.

Conclusion: In this study, 60% of EMS participants provided inaccurate GCS score estimates. Use of a GCS scoring aid improved accuracy of EMS GCS score assessments. [Ann Emerg Med. 2015;65:325-329.]

Please see page 326 for the Editor’s Capsule Summary of this article.

A podcast for this article is available at www.annemergmed.com.

INTRODUCTION

Background and Importance

First introduced in 1974, the Glasgow Coma Scale (GCS) is commonly used to describe the level of consciousness in and to predict outcomes of a wide variety of patients, including those in the out-of-hospital setting. \(^1\) Out-of-hospital providers use baseline and changes in GCS score assessments to indicate the severity of injuries or illness, and to aid in patient triage. \(^2\) In addition to its clinical utility, the GCS is commonly used in research for ascertainment of participant eligibility and as an outcomes assessment or adjustment for baseline severity. \(^5\)

Because the GCS can play a role in the initial and ongoing treatment of the patient, quick and accurate evaluation is necessary. The ability of out-of-hospital providers to accurately score the GCS has not been well reported, yet anecdotally they are often criticized for inaccurate GCS score assessment. There are only limited data characterizing the degree of emergency medical services (EMS) GCS inaccuracy. \(^6\) Furthermore, interrater reliability of GCS scoring is known to be low, including in the out-of-hospital setting. \(^7\) An aid to facilitate quick recall of the GCS in real time could improve scoring accuracy. \(^9\)

Goals of This Investigation

This study assessed the accuracy of EMS providers’ GCS scoring of written scenarios and estimated the potential for a GCS scoring aid to improve accuracy. We hypothesized that providers who were assisted by a GCS scoring table would assess GCS more accurately than those who were not.

MATERIALS AND METHODS

Study Design and Setting

This randomized controlled study of the utility of a GCS scoring table aid was conducted in the emergency department (ED) of an urban, academic Level I trauma center. The University of Cincinnati Institutional Review Board approved the study.

Selection of Participants

Participants were emergency medical technicians or paramedics who had transported a patient to the ED. We enrolled subjects during times of study personnel availability, which included weekdays, nights, and weekends. Providers were permitted to participate only once and had to be older than 18
Editor’s Capsule Summary

What is already known on this topic
Emergency medical services (EMS) personnel assessments of Glasgow Coma Scale (GCS) score are often inaccurate.

What question this study addressed
Does the use of a scoring aid improve the accuracy of EMS GCS score assessments?

What this study adds to our knowledge
Among mock written scenarios evaluated by 178 EMS personnel, GCS score accuracy was higher with (57%) than without (25%) a scoring aid.

How this is relevant to clinical practice
This scenario-based study found that GCS scoring accuracy for EMS personnel was low even when assisted by a scoring aid, further undermining the value of this already poor tool for neurologic assessment.

Methods of Measurement

Nine standardized brief patient scenarios (Table E1, available online at http://www.annemergmed.com) were modified from 3 widely used EMS textbooks.10-12 Attending physicians specializing in emergency medicine, EMS, and neurocritical care reviewed the scenarios and provided revisions until there was universal agreement on GCS scores. The scenarios depicted patients with GCS scores corresponding to mild (GCS score 13 to 15), moderate (GCS score 9 to 12), and severe (GCS score 3 to 8) traumatic brain injuries. The test scenarios and expert consensus GCS scores were verified by an independent team of paramedic instructors.

Scenarios with or without the scoring table were placed into sequentially numbered, sealed envelopes for distribution to participants. Each participant was randomly assigned to determine GCS scores on one of the 9 scenarios, with or without the scoring table. No blinding methods were used after randomization. Participants were asked to provide the total GCS score of the patient in the scenario, as well as the eye, verbal, and motor subcomponent scores. The participants’ demographic information was collected, including experience, level of training, and EMS practice habits.

Outcome Measures

The primary outcome was the absolute agreement between the participants’ assigned GCS scores and the correct GCS score determined by the attending physician review. Secondary outcomes included the frequency of scores falling within 1 point of the correct score, accuracy of subcomponent scores, and accuracy for the different levels of severity.

A sample size of 90 in each group would have 80% power to detect an absolute difference of 15% of the proportion of subjects able to correctly determine the GCS score with or without the GCS scoring aid when $\alpha=.05$ and conservatively assuming a wide SD.

Data Collection and Processing

The 1:1 randomization sequence was generated with nQuery Adviser (version 7.0; Statistical Solutions, Boston, MA) and designed to ensure equal distribution of the 9 scenarios among those receiving the GCS aid and those not receiving it. Participants’ responses were entered into an electronic database (REDCap; Vanderbilt University, Nashville, TN). Out-of-range GCS scores were queried and confirmed. Missing data were minimal and left missing. We compared participant GCS score with expert consensus GCS ratings, using the $\chi^2$ test or Fisher’s exact test, as appropriate, to test for differences in proportions and calculated 95% confidence intervals (CIs) for the effect size. We adjusted for multiple comparisons with Sidak’s method.

All statistical analyses were conducted with SPSS (version 22.0; IBM Corporation, Armonk, NY). Graphics were created with R (gplots). Differences in means and proportions and 95% CIs were calculated.

RESULTS

Characteristics of Study Subjects

Between April 2013 and June 2013, 261 subjects were screened; 16 declined participation and 65 did not meet inclusion criteria. Of 180 subjects enrolled, 2 participants were excluded because of incomplete GCS scores, leaving 178 cases in the analysis (Figure E1, available online at http://www.annemergmed.com).

Participant characteristics are described in Table 1. Approximately half (52%) were paramedics. Participants were drawn from 41 EMS departments or agencies, which were diverse and included rural, suburban, and urban settings; paid and volunteer staffing models; and annual call volumes ranging from less than 500 to greater than 55,000. The mean length of experience was 12 years (SD 8). Most participants (70%) stated they had been refreshed on GCS material through a course, recertification, or training within the past year, and 56% reported they had been refreshed on GCS material through a course, recertification, or training within the past year, and 56% stated they consistently use some sort of aid in the field to help determine the GCS score. The 2 study arms were well matched in experience and certification levels, and no protocol deviations occurred.

Overall, 73 of 178 participants (41%) gave a GCS score that matched the correct GCS score (Table 2; Figure). Among participants who did not receive the standard GCS scoring table as an aid, the GCS score was correct in 22 of 88 cases (25%) compared with 51 of 90 (57%) for those who did receive the table aid (difference in proportions 32%; 95% CI 18% to 46%).
Overall, 123 of 178 scores (69%) fell within 1 point of the correct GCS score. There was equal likelihood to overestimate (29.2%) and underestimate (29.8%) the total GCS score. More scores were correct within 1 point in the group who received the table aid than the group who did not (82.2% versus 55.7%; difference 26.5%; 95% CI 13.5% to 39.6%). The mean difference between actual and participant-assigned GCS scores for the group without the table was 2.6, and the mean participant-assigned GCS scores for the group with the table was 2.1 (difference of means 0.5; 95% CI –0.3 to 1.3).

The difference in accuracy between the 2 groups was most pronounced in the moderate (GCS score 9 to 12) and severe (GCS score 3 to 8) scenarios (Table 2; Figure). Twelve participants (7%) gave subcomponent scores that are not possible on the scale. Eye component accuracy improved from 43% without the table aid to 80% with the table aid, the verbal from 55% to 86%, and motor from 31% to 70% (Table 2; Figure).

### LIMITATIONS

Scoring a written scenario in a controlled environment and assigning a GCS score during the immediate evaluation and treatment phase of an acutely ill or injured patient are inherently

### Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th></th>
<th>No Table Aid (n = 88)</th>
<th>Table Aid (N = 90)</th>
<th>Total (n=178)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD), y</td>
<td>37 (10)</td>
<td>36 (9)</td>
<td>36 (9)</td>
</tr>
<tr>
<td>Race, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>72 (81.8)</td>
<td>76 (84.4)</td>
<td>148 (83.1)</td>
</tr>
<tr>
<td>Black</td>
<td>15 (17.0)</td>
<td>11 (12.2)</td>
<td>26 (14.5)</td>
</tr>
<tr>
<td>American Indian/Alaskan Native</td>
<td>0 (0)</td>
<td>2 (2.2)</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1.1)</td>
<td>0 (0)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>0 (0)</td>
<td>1 (1.1)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Male, No. (%)</td>
<td>80 (90.9)</td>
<td>77 (85.6)</td>
<td>157 (88.2)</td>
</tr>
<tr>
<td>Level of EMS certification, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMT-basic</td>
<td>44 (50.0)</td>
<td>39 (43.3)</td>
<td>83 (46.9)</td>
</tr>
<tr>
<td>EMT-intermediate</td>
<td>2 (2.3)</td>
<td>0 (0)</td>
<td>2 (1.1)</td>
</tr>
<tr>
<td>Paramedic</td>
<td>42 (47.7)</td>
<td>51 (56.7)</td>
<td>93 (52.0)</td>
</tr>
<tr>
<td>Years of experience, mean (SD)</td>
<td>12 (8)</td>
<td>11 (7)</td>
<td>12 (8)</td>
</tr>
<tr>
<td>Refreshed on GCS material within the past year, No. (%)</td>
<td>58 (65.9)</td>
<td>67 (74.4)</td>
<td>125 (70.2)</td>
</tr>
<tr>
<td>EMS instructor, No. (%)</td>
<td>6 (6.8)</td>
<td>7 (7.8)</td>
<td>13 (7.8)</td>
</tr>
<tr>
<td>Use aid to determine the GCS in the field, No. (%)</td>
<td>54 (61.4)</td>
<td>45 (50.0)</td>
<td>99 (50.0)</td>
</tr>
</tbody>
</table>

### Table 2. Scoring of patient scenarios by EMS providers.*

<table>
<thead>
<tr>
<th></th>
<th>Total (n=178)</th>
<th>No Table Aid (n = 88)</th>
<th>Table Aid (n = 90)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All GCS scenarios</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>73 (41.0)</td>
<td>22 (25.0)</td>
<td>51 (56.7)</td>
<td>31.9</td>
</tr>
<tr>
<td>Eye</td>
<td>110 (61.8)</td>
<td>38 (43.2)</td>
<td>72 (80.0)</td>
<td>37.3</td>
</tr>
<tr>
<td>Verbal</td>
<td>125 (70.2)</td>
<td>48 (54.5)</td>
<td>77 (85.6)</td>
<td>31.6</td>
</tr>
<tr>
<td>Motor</td>
<td>90 (50.6)</td>
<td>27 (30.7)</td>
<td>63 (70.0)</td>
<td>39.7</td>
</tr>
<tr>
<td>Mild TBI scenarios (GCS score 13–15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>32 (54.2)</td>
<td>13 (44.8)</td>
<td>19 (63.3)</td>
<td>14.3</td>
</tr>
<tr>
<td>Eye</td>
<td>41 (69.5)</td>
<td>16 (55.2)</td>
<td>25 (83.3)</td>
<td>18.5</td>
</tr>
<tr>
<td>Verbal</td>
<td>47 (79.7)</td>
<td>21 (72.4)</td>
<td>26 (86.7)</td>
<td>28.2</td>
</tr>
<tr>
<td>Motor</td>
<td>44 (74.6)</td>
<td>17 (58.6)</td>
<td>27 (90.0)</td>
<td>29.3</td>
</tr>
<tr>
<td>Moderate TBI scenarios (GCS score 9–12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17 (28.8)</td>
<td>3 (10.3)</td>
<td>14 (46.7)</td>
<td>31.4</td>
</tr>
<tr>
<td>Eye</td>
<td>37 (62.7)</td>
<td>12 (41.4)</td>
<td>25 (83.3)</td>
<td>34.9</td>
</tr>
<tr>
<td>Verbal</td>
<td>41 (69.5)</td>
<td>15 (51.7)</td>
<td>26 (86.7)</td>
<td>36.3</td>
</tr>
<tr>
<td>Motor</td>
<td>21 (35.6)</td>
<td>6 (20.7)</td>
<td>15 (50.0)</td>
<td>40.0</td>
</tr>
<tr>
<td>Severe TBI scenarios (GCS score 3–8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24 (40.0)</td>
<td>6 (20.0)</td>
<td>18 (60.0)</td>
<td>40.0</td>
</tr>
<tr>
<td>Eye</td>
<td>32 (53.3)</td>
<td>10 (33.3)</td>
<td>22 (73.3)</td>
<td>42.0</td>
</tr>
<tr>
<td>Verbal</td>
<td>37 (61.7)</td>
<td>12 (40.0)</td>
<td>25 (83.3)</td>
<td>43.3</td>
</tr>
<tr>
<td>Motor</td>
<td>25 (41.7)</td>
<td>4 (13.3)</td>
<td>21 (70.0)</td>
<td>56.7</td>
</tr>
</tbody>
</table>

TBI, Traumatic brain injury.

*Results are presented as the proportion of absolutely correctly assigned composite and component GCS scores and further stratified by mild, moderate, and severe TBI scenarios.
different. Although we have mirrored previously used methodologies, our approach may overestimate accuracy because the EMS providers are not subject to the task saturation of clinical care. Conversely, it is possible that the information gained from examining a live patient is more useful than that presented in a written scenario, which could improve the accuracy of GCS estimates.

We did not record whether subjects sought help in scoring the scenarios either from another EMS provider or from their personal GCS scoring aids, although no such activity was witnessed. Use of a scoring aid in the group not given one as part of the study would bias toward improved accuracy, and failure to use the aid provided would cause the opposite. In either case, the actual effect sizes would be greater than we observed.

Although we identified a deficiency in GCS scoring by EMS providers, we are unable to speculate about the reasons such a deficiency exists. Use of a scoring aid does improve accuracy, but discovery of other potential causes—and solutions—would be useful.

DISCUSSION

These results suggest that GCS score assessment with a scoring table improves the accuracy of EMS providers’ GCS scoring of patients, using written scenarios. However, even with use of a GCS table, accuracy of GCS scoring by EMS providers was low.

The GCS has been criticized as somewhat complex, and the fundamental utility and appropriateness of GCS scores in emergency medicine, and out-of-hospital care by proxy, have been challenged. We agree that GCS is imperfect and we support calls for a better tool. However, despite these criticisms, GCS is still the tool that has been universally adopted in clinical care. Until the GCS can be replaced, accurate scoring using the GCS should be emphasized. The inability of EMS providers to accurately assess an injured patient and communicate the findings is a problem regardless of the tool used.

Our data empirically quantify the inaccuracy, offering providers information that should be useful in interpreting an out-of-hospital GCS score. Clinically, a 1-point discrepancy in the GCS score may be acceptable, and when a scoring table aid was made available to providers, 82% of scores were within 1 point of the correct score. In other situations, even a 1-point error may prompt inappropriate field triage to a trauma center, exclusion from a clinical trial, or consideration of a procedure (ie, intubation). Providers were just as likely to overestimate and underestimate scores, and the magnitude of the difference was frequently enough to change the assigned category in the mild/moderate/severe classification scheme (Figure).

Figure. Dot plot of assigned composite and component GCS scores for each scenario. Each circle represents the score assigned by a single respondent.
Although our findings of inaccuracy when relying on memory alone are, unfortunately, not unique, we show that having a scoring table aid readily available more than doubles (25% to 57%) the number of accurate scores. To our knowledge, this is the first intervention shown to improve GCS scoring accuracy. In accordance with our observations, EMS providers should be given GCS scoring cards, with real-time use strongly encouraged.

Some healthcare providers have advocated abandonment of the full GCS and suggest simplifications or using only the motor component. In our sample, the motor score was the least reliable of the subcomponents. Proposed alternatives to the GCS that simplify assessment of consciousness include the FOUR score and the Emergency Coma Scale. However, these scoring methods may also suffer from accuracy limitations because the eye and motor components are similar to those of the GCS. Additionally, many of the articles that compare GCS with a newer tool of mental status assessment rely on retrospectively recorded out-of-hospital GCS values as the criterion standard. Our results do not support abandoning the full GCS in favor of these alternatives.

Our findings provide the key insights about the inaccuracy of GCS scoring by EMS and support the need for improved tools for evaluating out-of-hospital patients with neurologic emergencies. Until a new method of evaluating altered mental status in the setting of trauma is developed, validated, and adopted, use of a GCS scoring aid may help to improve the accuracy of the EMS GCS score assessments.

Supervising editor: Henry E. Wang, MD, MS

Author affiliations: From the Northeast Ohio Medical University, Akron, OH (Feldman); and the University of Cincinnati Department of Emergency Medicine, Cincinnati, OH (Hart, Lindsell, McMullan). Ms. Feldman was formerly affiliated with the University of Cincinnati Department of Emergency Medicine.

Author contributions: ALF, CJL, and JTM conceived and designed the study. ALF implemented the study, conducted all data collection, and drafted the article. KWH conducted the statistical analysis. All authors contributed to the interpretation of results, provided critical review and revisions for intellectual content for the article, and approved the final version. ALF takes responsibility for the paper as a whole.

Funding and support: By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist and provided the following details: This project was supported in part by an Institutional Clinical and Translational Science Award, NIH/NCRR grant 5UL1RR026314-03.

Publication dates: Received for publication February 18, 2014. Revisions received June 23, 2014; and July 9, 2014. Accepted for publication July 29, 2014. Available online September 6, 2014.

Presented as a poster at the National Association of EMS Physicians annual meeting, January 2014, Tucson, AZ.

REFERENCES

**Figure E1.** CONSORT flow diagram of subjects.
### Table E1. Patient scenarios randomly provided to on-duty EMS providers for GCS scoring, with the correct component and composite scores for each.

<table>
<thead>
<tr>
<th>Mild TBI</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>You respond to the scene of a 50-y-old man who was injured in a bicycle accident. A car pulled in front of him, forcing him off the road, and he fell in a grassy median. He is sitting up, inspecting his helmet, and is only complaining of road rash to his arms and legs. He is giving the police a description of the car involved and limps to the ambulance to have his wounds dressed.</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>You respond to an 18-y-old man involved in a single-car MVC, in which he struck a tree. There is moderate damage to the car, and he is sitting on the curb. When you ask him the date, he has slurred speech and states, “December 12, 2002” (it is actually February 14, 2013). When you ask him to show you 2 fingers, he giggles and flips you off with both hands. You note a strong odor of alcohol on his breath and possible track marks on his arms.</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>You respond to a 28-y-old woman struck in the head by a canoe oar. She was pulled from the river by bystanders, who state she was unconscious. She was wearing a life jacket. She awakens when you ask whether she is okay and asks what happened over and over again. She shows you 2 fingers on each hand when you ask her to do so.</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moderate TBI</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>You respond to a 16-y-old woman who was a pedestrian struck by a car and find her lying in the street. She has a boggy hematoma to the right side of her scalp and abrasions to her arms and legs. Her pupils are equal, round, and reactive to light. She opens her eyes when you call her name but she is confused. She knows her name and the year, but not the date or month. When you ask her to show you 2 fingers, she looks confused and does nothing. When you attempt to insert an IV line in her right arm, she quickly pulls away, swats at you with her left hand, and says, “Stop it!”</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>You respond to a 24-y-old male assault victim who was struck in the side of his head with a baseball bat during a bar fight. You find him breathing but unconscious. When you perform a sternal rub, he opens his eyes and tries to pull away, and stops when you stop stimulating him. When you constantly sternally rub him, he will talk to you and thinks the year is 1963 when asked (it is 2013).</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>You respond to a call for a “man down” by the railroad tracks and find a disheveled 45-y-old man facedown and parallel to the tracks. As you approach, he is moaning but you cannot understand what he is trying to say. He does not open his eyes to a deep sternal rub, he awakens when you apply a sternal rub, he reaches for your hand and briefly opens his eyes; his pupils are equal.</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severe TBI</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>You respond to a 22-y-old woman who was pushed down the stairs during a fight with her boyfriend. She fell down 12 wooden steps and landed on the cement basement floor. She is bleeding from the nose and mouth and has an obvious deformity to her left wrist. She will briefly open her eyes to a sternal rub, and her pupils are normal-sized and sluggishly reactive to light. She tries to pull away when you pinch her shoulder and mumbles something you cannot understand, but settles when you stop applying stimulation.</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>You respond to a motorcycle accident in which an unhelmeted rider hit a car that unexpectedly pulled out of a parking lot. He is found lying supine in the road 20 feet from the site of impact. Initially, you notice that he has irregular, snoring respirations and has obvious trauma to his head, face, and right leg. He is unresponsive and does not open his eyes to a deep sternal rub. You pull his eyelids open and discover that his left pupil is 2 mm larger than his right. Other than noisy respirations, he makes no sounds at any time. When you apply a sternal rub, his arms pull into his chest and his legs straighten out.</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>You respond to an 18-y-old man who fell out of a tree and landed on his head. He is briskly bleeding from his scalp but has no other obvious injuries. His only response to a deep sternal rub is to moan and groan. On your secondary examination, you find that his pupils are equal, dilated, and sluggishly reactive to light. There are several empty beer bottles at the base of the tree, and there is a strong odor of alcohol on his breath.</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Physician Knowledge of the Glasgow Coma Scale

RONALD G. RIECHERS II,1,3 ANTHONY RAMAGE,2 WILLIAM BROWN,2 AUDREY KALEHUA,3 PETER RHEE,4 JAMES M. ECKLUND,5 and GEOFFREY S.F. LING1,3,5

ABSTRACT
Appropriate triage is critical to optimizing outcome from battle related injuries. The Glasgow Coma Scale (GCS) is the primary means by which combat casualties, who have suffered head injury, are triaged. For the GCS to be reliable in this critical role, it must be applied accurately. To determine the level of knowledge of the GCS among military physicians with exposure and/or training in the scale we administered a prospective, voluntary, and anonymous survey to physicians of all levels of training at military medical centers with significant patient referral base. The main outcome measures were correct identification of title and categories of the GCS along with appropriate scoring of each category. Overall performance on the survey was marginal. Many were able to identify what “GCS” stands for, but far fewer were able to identify the titles of the specific categories, let alone identify the specific scoring of each category. When evaluated based on medical specialties, those in surgical specialties outperformed those in the medical specialties. When comparing the different levels of training, residents and fellows performed better than attending staff or interns. Finally, those with Advanced Trauma Life Support (ATLS) certification performed significantly better than those without the training. Physician knowledge of the GCS, as demonstrated in this study, is poor, even in a population of individuals with specific training in the use of the scale. It is concluded that, to optimize outcome from combat related head injury, methods for improving accurate quantitation of neurologic state need to be explored.

Key words: brain trauma; Glasgow Coma Scale; human; outcome; triage

INTRODUCTION
TRAUMATIC BRAIN INJURY (TBI) accounts for up to 20% of injuries incurred in battle by the U.S. military in Operation Iraqi Freedom (OIF) and Operation Enduring Freedom (OEF). TBI is also a leading cause of death and disability among civilians (Kraus and McArthur, 1996). When treating TBI casualties in the far forward environment, triage is paramount. For triage to be optimal, military medical first providers require an accurate and reliable clinical assessment of TBI. This is being done using the Glasgow Coma Scale (GCS) (Teasdale and Jennett, 1974). An important but unsubstantiated assumption is that the GCS is being properly administered.
Proceedings of the 133rd meeting of the Society of British Neurological Surgeons, Brighton, 23–25 September 1998

Glasgow coma scale: a help or hinderance
S. Bassi, N. Buxton, J. A. Punt & G. O’Reilly (Queen’s Medical Centre, Nottingham, UK)

Introduction: Teasdale and Jennett described the Glasgow Coma Scale (GCS) in 1974, stating that it would facilitate referrals between general and specialist centres. They felt that the Glasgow Coma Scale was simple to use with little interobserver variability.

Objectives: We aim to assess the understanding of the GCS across different clinical specialties.

Design: A prospective semi-structured interview based study of 140 junior and senior doctors. The questionnaire posed 6 scenarios to each doctor, who then proceeded to convert the clinical information into a coma score, or a coma score into a clinical situation.

Results: Overall, doctors managed to score only 48% correct in their assessment of the GCS. The results were slightly better in doctors using the scale most frequently (neurosurgeons scored 56%). There was also evidence of a learning curve with consultants scoring around 60%. We also assessed the range of inaccuracies encountered and some doctors were scoring a patient up to 5 points lower or higher in the scale than others were.

Conclusions: The GCS is widely used but poorly understood, and this may stem from being poorly taught at the undergraduate and postgraduate level. Doctors have difficulty in converting a patient’s clinical condition into a GCS, and there is significant interobserver variability. An incorrect assessment of the patient’s GCS may lead to sub-optimal management of the patient, e.g. inadequate airway protection during transfer of a patient, or miscalculation of the grade of a patient with a subarachnoid haemorrhage. Better communication would occur with the use of a descriptive coma assessment, rather than relying upon a numerical value in verbal communication.
British hospitals and different versions of the Glasgow coma scale: telephone survey

Martin F Wiese

The Glasgow coma scale is a clinical scoring system for objectively assessing how conscious a patient is. Although limited in predicting functional outcome, the scale is useful when making decisions about management in the acute setting, particularly for patients with traumatic brain injuries. Patients who score < 15 need imaging or observation, and patients with scores < 9 need to be promptly considered for definitive airways management.

The original Glasgow coma scale, published in 1974, had 14 points. Two years later, its authors introduced a distinction between “normal” and “abnormal” flexion (withdrawal to pain and decorticate response) increasing the “best motor response” item by one point. This revised scale is central to important clinical guidelines and has been the accepted version for more than 25 years.

Nevertheless, anecdotal evidence suggests that neurological observation charts based on the original 14 point scale are still being used in British hospitals. This study establishes the prevalence of this practice because it carries the potential for errors in communication and conflict with guidelines.

Participants, methods, and results

I conducted a national observational study of the neurological observation charts used in hospitals which care for adult patients with traumatic brain injuries.

Using the 2001-2 directory of the British Association for Accident and Emergency Medicine, I identified all UK emergency departments which manage patients with traumatic brain injuries. We contacted a sister or charge nurse in each department. Using a structured telephone interview, we asked staff to determine which version of the Glasgow coma scale they used by checking the neurological observation chart visually. We also asked the nurse to name one ward providing observation for patients with traumatic brain injuries, where we repeated the interview. Finally, we telephoned one ward in every specialist neurosurgical unit in the United Kingdom (table).

Comment

The original, 14 point, Glasgow coma scale continues to be used in many British hospital units which manage patients with traumatic brain injuries. Many of the staff that we contacted were not aware that the version of the scale that they were using had been superseded.

The parallel use of two versions of the Glasgow coma scale in the United Kingdom has been virtually unnoticed, possibly because publication of the revised scale in 1976 was not accompanied by an explanation and did not result in a clarifying change of name.

No evidence has been published that the continued use of the 14 point scale may have caused harm to any patient. The practice does, however, lead to difficulties. Staff from one hospital told us about their recurrent problems communicating with the local neurosurgical unit because the two were using different scales. In another trust, the neurological observation charts were changed after a recent coroner’s inquest: relatives of a patient who had died of severe traumatic brain injury had raised questions about the logic of two different versions of the Glasgow coma scale being used within the same trust.

Users of the Glasgow coma scale need training to ensure consistency and reliability of scoring. The continued employment of two different scales can only add to the confusion. Although the 15 point Glasgow coma scale is not perfect, it should be used by everybody who manages patients with traumatic brain injuries until even better measures of consciousness are devised.

I thank D Wallis for help with the study design and E Gluckman and P Leman for their comments on the drafts. Many thanks also to all the hospital staff.

Contributors: MFW wrote the paper and is guarantor. Nicola Burger collected some data.

Funding: None.

Competing interests: None declared.


(Accepted 3 July 2003)
Glasgow Coma Scale: variation in mortality among permutations of specific total scores

Abstract  

Objective: The objective of this study was to determine whether different score permutations of the Glasgow Coma Scale (GCS) giving the same GCS total score were associated with significantly different mortality.  

Design: For each GCS total we compared the mortality associated with each of the different GCS permutations using a Fisher's exact test. The relationship between components of the GCS score and mortality was also examined using univariate logistic regression.  

Setting: Data were collected from the intensive care unit at Wellington Hospital, a multidisciplinary, tertiary referral unit.  

Patients: We analysed the GCS and mortality data from all admissions over a 4 year period (January 1994–January 1998). Patients with GCS scores of 3 or 15 were excluded, since these two total scores do not have multiple permutations, leaving 1390 patients with GCS scores of 4–14 for analysis.  

Results: The incidence of mortality was significantly different for the different permutations for total GCS scores of 7, 9, 11 and 14.  

Conclusions. It is possible for patients to have the same total GCS score, but significantly different risks of mortality due to differences in the GCS profile making up that score. This suggests that GCS scores may be more useful reported in terms of profiles rather than totals. This could also have implications for the use of other scoring systems such as Acute Physiology and Chronic Health Evaluation and Simplified Acute Physiology Score.  

Key words  

Glasgow Coma Scale · Outcome · Permutation

Introduction

The Glasgow Coma Scale (GCS) was developed by Teasdale and Jennett in 1974 [1] to assess the depth and severity of head trauma. The [GCS] assesses patient responsiveness through increasing degrees of dysfunction in three features: eye responses (E), motor responses (M) and verbal responses (V). Subsequently, scores were allocated to performance on the three separate components and the total Glasgow Coma Score (GCSc) is derived from the summation of these [2]. A correlation between GCSc and outcome has been demonstrated in patients with head injury [2]. In addition to being used in patients with head injuries, GCS has been studied in a number of other patient subgroups. Correlations between outcome and total GCSc (T) have been reported in patients with infectious disease [3], respiratory infections [4], poisoning [5] and in post cardiac arrest patients [6]. The GCS and GCSc are commonly used in the intensive care setting to assess a broad spectrum of patients with some form of brain failure, as a component in other scoring systems such as the Acute Physiology and Chronic Health Evaluation (APACHE), as a guide to severity of illness and as a useful and objective assessment of a patient’s progress over time.
A GCS score of 4 predicts a mortality rate of 48% if calculated 1+1+2 for eye, verbal, and motor, a mortality of 27% if calculated 1+2+1, but a mortality of only 19% if calculated 2+1+1.
Comments from the creators of the GCS

“We have never recommended using the GCS alone, either as a means of monitoring coma, or to assess the severity of brain damage or predict outcome.”


The total GCS “assumes an equal weighting for the three responses. More importantly, the information conveyed by the [total] coma score is less than that contained in the three responses separately … Indeed in Glasgow patients are always described by the three separate responses and never by the total.”

Glasgow Coma Scale Motor Component (“Patient Does Not Follow Commands”) Performs Similarly to Total Glasgow Coma Scale in Predicting Severe Injury in Trauma Patients

Douglas F. Kupas, MD*; Eric M. Melnychuk, DO; Amanda J. Young, MS

*Corresponding Author. E-mail: dkupas@geisinger.edu.

Study objective: Trauma victims are frequently triaged to a trauma center according to the patient’s calculated Glasgow Coma Scale (GCS) score despite its known inconsistencies. The substitution of a simpler binary assessment of GCS-motor (GCS-m) score less than 6 (ie, “patient does not follow commands”) would simplify field triage. We compare total GCS score to this binary assessment for predicting trauma outcomes.

Methods: This retrospective analysis of a statewide trauma registry includes records from 393,877 patients from 1999 to 2013. Patients with initial GCS score less than or equal to 13 were compared with those with GCS-m score less than 6 for outcomes of Injury Severity Score (ISS) greater than 15, ISS greater than 24, death, ICU admission, need for surgery, or need for craniotomy. We judged a priori that differences less than 5% lack clinical importance.

Results: The relative differences between GCS and GCS-m scores less than 6 were less than 5% and thus clinically unimportant for all outcomes tested, even when statistically significant. For the 6 outcomes, the differences in areas under receiver operating characteristic curves ranged from 0.014 to 0.048. Total GCS score less than or equal to 13 was slightly more sensitive (difference 3.3%; 95% confidence interval 3.2% to 3.4%) and slightly less specific (difference −1.5%; 95% confidence interval −1.6% to −1.5%) than GCS-m score less than 6 for predicting ISS greater than 15, with similar overall accuracy (74.1% versus 74.2%).

Conclusion: Replacement of the total GCS score with a simple binary decision point of GCS-m score less than 6, or a patient who “does not follow commands,” predicts serious injury, as well as the total GCS score, and would simplify out-of-hospital trauma triage. [Ann Emerg Med. 2016; :1-7.]

Please see page XX for the Editor’s Capsule Summary of this article.

INTRODUCTION

Background

Rapid and accurate assessment of a patient’s condition is essential for trauma triage in the field. Emergency medical services (EMS) providers of all levels must be able to quickly evaluate and classify patients for appropriate transport destination while providing medical care. Avoiding undertriage by transporting patients with potentially serious injuries to an appropriate trauma center reduces mortality, but overtriage causes a strain on resources and is inconvenient for patients.

The guidelines for field triage of injured patients were designed for use by EMS providers to identify patients with potentially serious injuries and determine the most appropriate level of care.1,2 The 2011 version consists of 4 steps to determine the appropriate destination for patients.

Step 1 includes physiologic criteria, including assessment of vital signs and the Glasgow Coma Scale (GCS) score, and recommends that a patient with a GCS score of less than or equal to 13 be transported to a trauma center, preferably to the highest level of care within the defined trauma system.

Importance

Recent research on step-specific field triage has shown the motor component of the GCS (GCS-m) to be a more specific and simpler tool for patient assessment.3-8 A calculated GCS score of less than or equal to 13 may be a statistically more sensitive and less specific indicator of serious injury than the GCS-m score, which may lead to overtriaging of patients and thus transporting patients to more distant resources that may not be needed for them. Small differences may not be clinically significant, and field
The Motor Component of the Glasgow Coma Scale

Editor’s Capsule Summary

What is already known on this topic
The Glasgow Coma Scale (GCS) is widely used as a criterion for field triage of injured patients to trauma centers.

What question this study addressed
Does a single GCS element (GCS motor component score <6 or “patient does not follow commands”) predict trauma outcomes, as well as the widely used threshold of total GCS score less than or equal to 13?

What this study adds to our knowledge
In this analysis of a 393,877-adult statewide trauma registry, the differences observed between the new decision point and GCS score less than or equal to 13 were all below the prespecified 5% threshold of clinical importance for 8 trauma outcomes.

How this is relevant to clinical practice
The full GCS is unnecessarily complicated for out-of-hospital field triage and can be effectively replaced by the single decision point “patient does not follow commands.”

use of GCS-m score may be more reliable than the total calculated GCS score. The GCS score is only 1 parameter of trauma triage; therefore, these relatively small differences in sensitivity have an even smaller influence on overall trauma triage sensitivity. The National Expert Panel on Field Triage considered emerging evidence for the use of GCS-m score during their literature review when developing the 2011 guidelines, but this group ultimately did not include use of the GCS-m score in the current guidelines because of “lack of confirmatory evidence, the long standing use of total GCS and its familiarity among current EMS practitioners, the inclusion of the motor score within the total GCS, and complications because of the difficulty of comparing scoring systems.” However, several studies have indicated a significant interobserver variability in tallying the total GCS score, with discrepancies as high as 3 points. Even the assessment of the GCS-m score suffers from lack of standardization, with variations based on type of painful stimuli applied to elicit responses and variations because of provider education. Gill et al studied the interrater differences among emergency physicians in determining the GCS score and found that the agreement percentage for exact total GCS score was 32%, whereas the agreement percentage for the motor component was 72%. It is generally accepted that the motor component of the GCS is the most influential one when a patient’s severity of injury is assessed.

Goals of This Investigation
We wished to compare the total GCS score less than or equal to 13 with the GCS-m score less than 6 (“patient does not follow commands”) in predicting trauma-related outcomes.

MATERIALS AND METHODS

Study Design and Setting
We retrospectively analyzed the prospectively maintained Pennsylvania Trauma System Foundation’s registry, which included trauma patients admitted to the state’s Level I, II, III, and IV trauma centers from 1999 to 2013. The Pennsylvania Trauma System Foundation registry captures all patients with a diagnosis of trauma who are admitted to a Foundation-accredited Level I, II, III, or IV trauma center and patients presenting to the trauma center dead on arrival. This includes all trauma transfer admissions and trauma deaths. Solitary hip fractures are excluded. Patients do not need a minimum Injury Severity Score (ISS) to be included into the registry. The majority of accredited trauma centers in Pennsylvania during this study were Level I and II. Level IV trauma centers were first recognized in Pennsylvania in November 2013, and during the study (the last 2 months), there was only 1 Level IV accredited center. Pennsylvania has an exclusive trauma system, and the Pennsylvania Trauma System Foundation data exclude patients who were treated only at facilities that are not accredited trauma centers, although statewide EMS triage criteria and hospital referral patterns generally direct seriously injured trauma patients to accredited trauma centers, either initially or by interfacility transfer. Additional description of the Pennsylvania Trauma System Foundation trauma registry can be found at http://www.ptsf.org/index.php/resources.

Quality assurance and improvement measures for the Pennsylvania Trauma Outcome Study include internal data validation of the data entry system at each trauma center. Each trauma center’s data are abstracted locally and collected in the Pennsylvania Trauma Outcome Study by trained trauma registrars. Additional reviews are performed at the central site, including a random sampling program, which generates case reviews. Data are also validated against objective coding software, and foundation staff randomly select several cases from an institution and review the medical records at sites for consistency, accuracy, and completeness.

2 Annals of Emergency Medicine
This study was approved by the Geisinger Health System institutional review board, with the specific determination that this study met exempt criteria for full institutional review board review. Data obtained from the Pennsylvania Trauma Systems Foundation State Registry were approved by Pennsylvania Trauma System Foundation.

**Selection of Participants**

The database contained 393,877 adults aged 18 years and older. The out-of-hospital total GCS score, out-of-hospital GCS-m score, and ISS were obtained from each patient record. The primary outcome by which we compared total GCS scores with GCS-m scores was ISS greater than 15.

Secondary outcomes were also collected from each patient, which included ISS greater than 24, death, ICU admission, need for craniotomy, any surgery (defined as intrathoracic, abdominal, vascular, or cranial surgery), intubation (at the scene or in the trauma care center), and a composite variable, trauma care need. Trauma care need was defined as ISS greater than 15, ICU admission 24 hours or greater, need for surgery, or death before discharge. Values of GCS and GCS-m scores, systolic blood pressure, and respiratory rate were captured at first report (at either the scene or trauma center). If these physiologic criteria were available in an out-of-hospital patient care report, then this was used. Otherwise, the first physiologic criteria obtained by the hospital providers on arrival to the trauma center were captured. In

### Table 1. Sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio of total GCS versus GCS-m only scores in predicting outcomes.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>GCS Score ≤13 (95% CI)</th>
<th>GCS-m Score &lt;6 (95% CI)</th>
<th>Relative Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ISS &gt;15, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>31.3 (31.0 to 31.6)</td>
<td>28.0 (27.7 to 28.3)</td>
<td>3.3 (3.2 to 3.4)</td>
</tr>
<tr>
<td>Specificity</td>
<td>91.3 (91.2 to 91.4)</td>
<td>92.8 (92.7 to 92.9)</td>
<td>-1.5 (-1.6 to -1.5)</td>
</tr>
<tr>
<td>LR+</td>
<td>2.54 (2.52 to 2.56)</td>
<td>2.56 (2.54 to 2.59)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.53 (0.53 to 0.54)</td>
<td>0.51 (0.51 to 0.52)</td>
<td></td>
</tr>
<tr>
<td><strong>ISS &gt;24, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>47.2 (46.7 to 47.6)</td>
<td>43.7 (43.3 to 44.2)</td>
<td>3.5 (3.3 to 3.6)</td>
</tr>
<tr>
<td>Specificity</td>
<td>89.3 (89.2 to 89.4)</td>
<td>91.1 (91.0 to 91.2)</td>
<td>-1.8 (-1.9 to -1.8)</td>
</tr>
<tr>
<td>LR+</td>
<td>4.97 (4.89 to 5.05)</td>
<td>5.11 (5.03 to 5.19)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.67 (0.66 to 0.67)</td>
<td>0.64 (0.64 to 0.65)</td>
<td></td>
</tr>
<tr>
<td><strong>Died, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>69.8 (69.2 to 70.4)</td>
<td>67.3 (66.7 to 67.9)</td>
<td>2.5 (2.2 to 2.7)</td>
</tr>
<tr>
<td>Specificity</td>
<td>88.1 (88.0 to 88.2)</td>
<td>90.1 (90.0 to 90.2)</td>
<td>-2.0 (-2.1 to -1.9)</td>
</tr>
<tr>
<td>LR+</td>
<td>12.88 (12.52 to 13.25)</td>
<td>13.551 (13.183 to 13.920)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.76 (0.75 to 0.76)</td>
<td>0.726 (0.722 to 0.730)</td>
<td></td>
</tr>
<tr>
<td><strong>ICU admission, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>27.3 (27.1 to 27.5)</td>
<td>23.9 (23.7 to 24.1)</td>
<td>3.39 (3.29 to 3.50)</td>
</tr>
<tr>
<td>Specificity</td>
<td>91.6 (91.4 to 91.8)</td>
<td>92.7 (92.6 to 92.9)</td>
<td>-1.15 (-1.24 to -1.06)</td>
</tr>
<tr>
<td>LR+</td>
<td>1.52 (1.52 to 1.53)</td>
<td>1.51 (1.50 to 1.51)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.37 (0.37 to 0.38)</td>
<td>0.37 (0.37 to 0.38)</td>
<td></td>
</tr>
<tr>
<td><strong>Intubation, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>83.7 (83.3 to 84.2)</td>
<td>81.3 (80.9 to 81.8)</td>
<td>2.4 (2.1 to 2.6)</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.0 (89.9 to 90.1)</td>
<td>92.0 (91.9 to 92.1)</td>
<td>-2.0 (-2.1 to -1.9)</td>
</tr>
<tr>
<td>LR+</td>
<td>28.62 (27.72 to 29.53)</td>
<td>28.70 (27.85 to 29.55)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.62 (0.62 to 0.62)</td>
<td>0.57 (0.57 to 0.58)</td>
<td></td>
</tr>
<tr>
<td><strong>Trauma care need, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>28.2 (27.9 to 28.4)</td>
<td>25.1 (24.9 to 25.3)</td>
<td>3.0 (2.9 to 3.1)</td>
</tr>
<tr>
<td>Specificity</td>
<td>93.7 (93.6 to 93.8)</td>
<td>95.0 (94.9 to 95.1)</td>
<td>-1.3 (-1.4 to -1.3)</td>
</tr>
<tr>
<td>LR+</td>
<td>2.19 (2.17 to 2.20)</td>
<td>2.21 (2.19 to 2.22)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.38 (0.37 to 0.38)</td>
<td>0.35 (0.34 to 0.35)</td>
<td></td>
</tr>
<tr>
<td><strong>Surgery, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>33.5 (33.0 to 34.0)</td>
<td>30.5 (30.0 to 31.0)</td>
<td>3.0 (2.8 to 3.2)</td>
</tr>
<tr>
<td>Specificity</td>
<td>86.5 (86.4 to 86.6)</td>
<td>88.4 (88.3 to 88.5)</td>
<td>-1.9 (-2.0 to -1.9)</td>
</tr>
<tr>
<td>LR+</td>
<td>2.81 (2.75 to 2.87)</td>
<td>2.89 (2.83 to 2.96)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.87 (0.87 to 0.88)</td>
<td>0.86 (0.86 to 0.87)</td>
<td></td>
</tr>
<tr>
<td><strong>Craniotomy, %</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>51.4 (50.2 to 52.5)</td>
<td>46.5 (45.4 to 47.7)</td>
<td>4.9 (4.3 to 5.4)</td>
</tr>
<tr>
<td>Specificity</td>
<td>85.9 (85.8 to 86.0)</td>
<td>87.8 (87.7 to 87.9)</td>
<td>-2.0 (-2.0 to -1.9)</td>
</tr>
<tr>
<td>LR+</td>
<td>6.03 (5.76 to 6.30)</td>
<td>5.88 (5.61 to 6.12)</td>
<td></td>
</tr>
<tr>
<td>LR-</td>
<td>0.94 (0.94 to 0.94)</td>
<td>0.94 (0.94 to 0.94)</td>
<td></td>
</tr>
</tbody>
</table>

CI, Confidence interval; LR+, positive likelihood ratio; LR−, negative likelihood ratio.
this study, the first reported physiologic data were available from out-of-hospital records 48% of the time, and the data from initial hospital assessment were used for the remainder. After exclusion of patients with missing GCS score, GCS-m score, respiratory rates, and systolic blood pressure; with trauma year out of range; who do not fit into the trauma registry inclusion and exclusion criteria according to the Pennsylvania Trauma Outcome Study documentation; and who were missing ISS, the analysis data set included 370,392.

Methods of Measurement

We used descriptive analyses for our data (Tables E1 and E2, available online at http://www.annemergmed.com) and contrasted the GCS-m score with the total GCS score with the threshold less than or equal to 13 and with receiver operating characteristic curves. Recognizing that our large sample would likely identify some small differences as statistically significant, we defined a priori differences of less than 5% (ie, <0.05 for the area under a receiver operating characteristic curve) as clinically unimportant regardless of statistical probability.

We performed sensitivity analyses excluding patients who meet criteria for transport directly to a trauma center because of other trauma triage criteria within the guidelines (including those with systolic blood pressure less than 90 mm Hg and respiratory rate less than 10 or greater than 29 breaths/min, as well as for anatomic reasons including flail chest, skull fracture, paralysis, amputation, pelvic fractures, bone fractures, and penetrating injuries), generating a sample of 315,034. In a second sensitivity analysis, we restricted the sampling to only patients with

![Figure. AUC for GCS-m and GCS scores.](image)
out-of-hospital reported total GCS score versus GCS-m score.

**Primary Data Analysis**

We used the SAS (version 9.4; SAS Institute, Inc., Cary, NC) for data analysis.

**RESULTS**

We found that the differences between total GCS score less than or equal to 13 and GCS-m score less than 6 were all below our prespecified 5% threshold for clinical importance, ranging from 2.5% to 4.9% for sensitivity and −1.2% to −2.0% for specificity (Table 1). All such differences had 95% confidence intervals that did not overlap zero. We found similar results in our 2 sensitivity analyses.

Differences in areas under receiver operating characteristic curves ranged from 0.014 to 0.048 (Figure, Tables 2 and E3 [available online at http://www.annemergmed.com]), all also below our prespecified 0.05 threshold for clinical importance.

**LIMITATIONS**

This study is from a single state and may not be representative elsewhere, although Pennsylvania includes large urban, suburban, and rural areas. A large proportion of the patients in the registry were victims of blunt trauma. In addition, for analysis the values of total GCS and GCS-m scores, systolic blood pressure and respiratory rate were captured at first report (either in the out-of-hospital setting or at the trauma center). A further limitation is that approximately half of the first reported GCS scores were determined by hospital providers rather than in the field by EMS providers, and the outcomes might have differed if EMS providers had routinely provided GCS scores.
However, our sensitivity analysis of just this subgroup showed similar results.

DISCUSSION

Total GCS score has historically been an important physiologic component of field triage used to predict trauma outcomes. This relatively complicated 13-point scale has shown inaccuracy among health care workers, however, putting its reliability in question.9

A simpler assessment of cerebral function is the binary clinical determination of whether a patient “follows commands” (GCS-m score=6) or does not. Although previous studies have shown greater agreement among emergency physicians assessing the GCS-m score compared with total GCS score, it is reasonable to posit that the straightforward assessment of “following commands” would be as accurate as or more accurate than the assessment of all parts of the GCS-m score. This simple binary assessment of whether a patient “follows commands” is practical and appealing when one considers ease of education and use by all levels of EMS providers when they make field triage determinations.

We found that the differences between total GCS score less than or equal to 13 and GCS-m score less than 6 were below our prespecified 5% threshold for clinical importance, despite statistically significant associations that predictably resulted from our extremely large sample size. Our data thus confirm the findings of previous studies that our simpler decision point is just as predictive of trauma outcomes as the full GCS.13,14 A simplified field triage score for battlefield casualties, which includes the GCS-m, has shown promising results for use as a practical instrument in the combat zone.15 Additionally, the use of motor response in children after they sustain a traumatic head injury has shown to predict long-term outcome, as well as the full GCS score, with better interobserver agreement.16 A 2012 study identified GCS-m score as part of a prognostic model for predicting mortality at 30 days and unfavorable outcome at 6 months after traumatic brain injury.17 A retrospective study found that if patients had a GCS-m score less than 6 and a systolic blood pressure less than 90 mm Hg, 95% of them needed a lifesaving intervention.18 A similar study showed that GCS-m and GCS verbal scale scores, along with pulse character, predicted need of out-of-hospital lifesaving intervention.19 The performance of GCS-m in previous studies and in this study, as well as ease of using GCS-m in the out-of-hospital setting, make a strong argument for simplifying the national Guidelines for Field Triage by changing to the use of GCS-m score less than 6 or “patient does not follow commands” when making out-of-hospital trauma triage decisions.

In conclusion, during trauma triage a simple binary decision point of GCS-m score less than 6, or a patient who “does not follow commands,” predicts serious injury similarly to the more complicated calculation of total GCS score. For all outcomes, the relative differences in specificity, sensitivity, and area under the receiver characteristic curve between GCS-m score and total GCS score were clinically unimportant; therefore, we recommend our simpler binary assessment as a replacement for the total GCS score for field trauma triage.

Table 2. Area under the receiver operating characteristic curve for each outcome.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>GCS Score ≤13</th>
<th>GCS-m Score &lt;6</th>
<th>Relative Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISS &gt;15</td>
<td>0.648 (0.646–0.650)</td>
<td>0.606 (0.605–0.608)</td>
<td>0.042 (0.041–0.043)</td>
</tr>
<tr>
<td>ISS &gt;24</td>
<td>0.719 (0.716–0.721)</td>
<td>0.680 (0.677–0.682)</td>
<td>0.039 (0.038–0.041)</td>
</tr>
<tr>
<td>Died</td>
<td>0.831 (0.828–0.834)</td>
<td>0.803 (0.800–0.806)</td>
<td>0.028 (0.026–0.030)</td>
</tr>
<tr>
<td>ICU admission</td>
<td>0.625 (0.623–0.626)</td>
<td>0.583 (0.581–0.584)</td>
<td>0.042 (0.041–0.043)</td>
</tr>
<tr>
<td>Intubation</td>
<td>0.904 (0.902–0.907)</td>
<td>0.884 (0.882–0.887)</td>
<td>0.020 (0.019–0.021)</td>
</tr>
<tr>
<td>Trauma care need</td>
<td>0.641 (0.639–0.642)</td>
<td>0.603 (0.602–0.604)</td>
<td>0.038 (0.037–0.039)</td>
</tr>
<tr>
<td>Surgery</td>
<td>0.612 (0.608–0.615)</td>
<td>0.597 (0.595–0.600)</td>
<td>0.014 (0.013–0.016)</td>
</tr>
<tr>
<td>Craniotomy</td>
<td>0.724 (0.718–0.730)</td>
<td>0.676 (0.670–0.682)</td>
<td>0.048 (0.044–0.052)</td>
</tr>
</tbody>
</table>

AUC, Area under the receiver operating characteristic curve.

Supervising editor: Steven M. Green, MD

Author affiliations: From the Department of Emergency Medicine (Kupas, Melnychuk) and Center for Health Research (Young), Geisinger Health System, Danville, PA.

Author contributions: DFK and EMM conceived and designed the study and obtained institutional review board approval and waiver. AJY obtained the data and provided statistical analysis. All authors analyzed and interpreted the data, drafted the article, and contributed substantially to its revision. DFK takes responsibility for the paper as a whole.

Funding and support: By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships...
in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist and provided the following details:

Dr. Kupas is a member of the board of directors of the Pennsylvania Trauma Systems Foundation.

**Publication dates:** Received for publication December 4, 2015. Revisions received June 1, 2016, and June 3, 2016. Accepted for publication June 13, 2016.

**REFERENCES**


Lessons From the Glasgow Coma Scale

To the Editor:

Dr. Green’s comments on the persistence of the Glasgow Coma Scale (GCS) were right on the mark but have much broader implications than he observed. Nothing persists in social systems unless it benefits someone, somewhere, somehow, so when a seemingly useless ritual such as the GCS lasts for years, especially in the face of sound arguments against it, we should look for who it benefits and how. As Dr. Green observed, the persistence of the GCS is due not to its clinical utility but rather its psychological comfort. It gives the appearance of reducing uncertainty, creating order, and rationalizing clinical practice. That it does none of these is immaterial to the benefits it provides in reducing cognitive and affective stress. There are undoubtedly social benefits as well, in the form of providing seemingly objective justification for actions taken or not taken.

I would argue that the same factors are playing out in many other areas of health care, eg, the proliferation of decision rules, the desire for guidelines, the quest for standardization and the aversion to variation or heterogeneity, the faith in “evidence-based medicine,” the yearning for quantitative measurement, the fascination with templates and checklists, and the magical thinking about information technology. Although there are sometimes beneficial aspects to these efforts, much of the impetus behind them stems from their ability to increase the appearance of rationality and order, to make the intractable seem tractable and the ineffable, effable.

This might not matter if there were not potential drawbacks hidden in these ideas. There are 3 issues here. First, they can lead to miscalibration, an unwarranted sense of confidence in one’s understanding of a situation, an illusion of control. Second, there is abundant evidence from multiple fields that people will accept guidance from external sources or decision aids that is inferior to that which they could have produced themselves, if unaided. Finally, these notions can produce a sort of socially agreed-on agnosia, a total dependence on rational understanding in which everything must be decomposed before it can be understood, and that which cannot be decomposed is presumed not important.

Sorting through the complex motivations lying behind these schemes to organize and rationalize care so we can identify the changes that might actually improve clinical work, rather than just make us feel better about it, should be an important task for those who wish to make emergency care better.

Robert L. Wears, MD, MS
Department of Emergency Medicine
University of Florida
Jacksonville, FL
Clinical Safety Research Unit
Imperial College
London, UK

Funding and support: By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The author has stated that no such relationships exist.