**Original Article**

**The FOUR Score and Glasgow Coma Scale to Evaluate the Patients with Intubation at Emergency Room**

Niti Matheesiriwat¹² and Somsak Kuptniratsaikul³

¹ Emergency Physician, Department of Trauma and Emergency, Phramongkutklao Hospital;
² Master of Science Program in Health Development Student, Faculty of Medicine, Chulalongkorn University;
³ Chulalongkorn Hospital, Faculty of Medicine, Chulalongkorn University

**Objective**: To evaluate the reliability between the FOUR score and the GCS score for evaluating patients with intubation in emergency room and to predict in-hospital mortality. **Methods**: Eighty patients with intubation in the ER were enrolled. Each patient was evaluated by random pairs of raters using both the FOUR and the GCS scores. The statistical analysis included Willcoxon’s sign rank test to determine the differences between the groups, Spearman’s and Kendall’s correlation to determine the correlation between FOUR and GCS, weighted kappa and intra-class correlation coefficient (ICC) to determine the degree of agreement and the logistic regression for in-hospital deaths. **Results**: Inter-rater agreements were good to excellent for the FOUR and GCS scores. The correlation between score system was good. For every 1-point increase in the total scores, there was a reduction in the odds of an in-hospital mortality. **Conclusion**: The FOUR score is reliable for evaluating patients with intubation in the ER. The FOUR score can be used to predict in-hospital mortality.

**Key Words**: • FOUR score • GCS score • Emergency room • Intubation


---

**Introduction**

The Glasgow Coma Scale (GCS) is a worldwide tool used for evaluating the alterations of consciousness in patients. It is widely popular particularly in pre-hospital care, with emergency physicians and neuroscience physicians. There are some limitations in some cases such as patients with intubation or lock- in syndrome.¹ Because of these limitations the overall decline can’t be fully evaluated in these patients.

Many scores have been developed to assess the patients such as RAPS (The Rapid Acute Physiology Score), REMS (Rapid Emergency Medicine Score)², APACHE (Acute Physiology and Chronic Health Evaluation II)³ and IHSS (In House Score System)⁴ but no score has been used as the gold standard and some are difficult to use. In many scores, the GCS is also part of the assessment.

Since 2005, the Full Outline of UnResponsiveness (FOUR) score has been developed to reduce some of the limitations of the GCS in the intensive care setting compared to the GCS for evaluating patients. The FOUR score can correlate well with the GCS to be assessed by evaluation by several specific groups, such as nurses and neuroscience physicians. There has been widespread increased use of the FOUR score in evaluating patients in the ICU MED⁵, Neuro ICU⁶, Pediatric population⁷ and ER setting⁸ in comparison to the GCS assessment in the alteration of consciousness.
divided into groups such as alertness, drowsiness, stuporous and coma. All studied results were effective.

The FOUR score consists of four components e.g. eyes, motor, brainstem and respiration pattern and the evaluator assigns a score of 0 to 4 to each of the four functional categories. The maximum total score is 16 and the minimum score is 0. (Figure 1)

This study compared the FOUR score and the GCS score used for patients with intubation in the emergency room. Our objective was to study the intra-observer reliability, validity and functional outcome at hospital discharge.

Materials and Methods

An observational study was conducted after approval by ethic committee. The data was collected from patients who had alterations of consciousness with intubation in the emergency room, Phramonkutklao Hospital between May 2011 and November 2011.

Inclusion criteria
1. Patients with intubation
2. Age over 18 years old
3. Thai language understandable
4. Visited to emergency room at Phramongkutklao Hospital

Exclusion criteria
1. Patients received sedative or neuromuscular blocking agents.
2. Patients with effects from anesthesia within the period of 24 hours.

A total of 80 patients were recruited for the study.

The FOUR and GCS scores would be assessed by three types of raters, each with two personnel i.e. emergency medicine residents (D), nurses (N) and externs (E). To protect patients from over-assessment, only 2 raters would independently examine and assign both the FOUR and the GCS scores to each patient in the ER within 30 minutes. (Figure 2)

The raters all participated in training, provided by the investigator, related to the use of the GCS and the FOUR score assessment tool. The raters were given a copy of the GCS and the FOUR score instruction card for reference during the assessment of the patients.

To reduce bias, raters were blinded to the other’s score and were not aware of the diagnosis of the patients. A randomization sheet was used to select the rater pair (D/D, D/E, D/N, E/E, E/N or N/N) that would assess the
The FOUR Score and Glasgow Coma Scale to Evaluate the Patients with Intubation at Emergency Room

All alteration of consciousness patients with intubation at emergency room

Inclusion/Exclusion criteria

Two raters both score evaluation within 30 min for each patient
(Nurse/Nurse, Nurse/extern, Nurse/resident EM, Extern/Extern, Extern/resident, Resident/resident)

FOUR score

Glasgow Coma Scale (GCS) score

Initial treatment

Data collection

age, sex, FOUR and GCS total score
Level of consciousness
Diagnosis after discharge from hospital
Dead/alive from discharge

Data analysis

Result & Report

Figure 2 Flow Diagram

patients. During the patients evaluation each rater would follow the instructions and complete the scoring sheet. Informed consent releases were required to be signed by every patient or the legal relative and the study was approved by the Ethics Committee for Clinical Research of Phramongkutklao Hospital.

Analyses

The statistical analysis of the study used descriptive statistics for the baseline data which will be presented as mean, SD, or median in continuous data and the categorical data will be presented as frequency (%). We used Willcoxon’s sign rank test to determine the differences between groups, Spearman’s and Kendall’s correlation to determine correlation between the FOUR and the GCS, weighted kappa and Intra-class Correlation Coefficient (ICC) to determine the degree of agreement and the logistic regression for in-hospital death. Their analyses were conducted using the SPSS software (version 15) and STATA 12 trial version.

Results

There were 80 patients and 160 sets of data available for final analysis. The age of the patients in the study was 62.23 ± 17.92 years with a range of 19-92 years. Forty-five patients (56%) were male. All patients were intubated with a mechanical ventilator and categorized into 3 stages of consciousness as follows; 64% drowsy, 10% stupor and 26% coma. The top three underlying
diseases of the patients were hypertension, diabetic mellitus and ischemic heart disease. Following the results of the treatment at the hospital after discharge, there were 21 patients who died. (Table 1)

The overall rater agreement was good to excellent for both scores. Inter-rater reliability using weight kappa ($K_w$) and intra-class correlation coefficient (ICC) for the FOUR score ($K_w = 0.8$, ICC = 0.96, 95% CI 0.93-0.97) and the GCS score ($K_w = 0.83$, ICC = 0.92, 95% CI 0.96-0.98). Weight kappa agreement value according to the following: values of 0.4 or less is considered poor, values between 0.4 and 0.6 are considered fair to moderate, those between 0.6 and 0.8 suggest good interobserver agreement, and values greater than 0.8 suggest excellent agreement. (Table 2)

Table 3 presents the FOUR score and the GCS within subject differences by rater types (ED Resident, ED Nurse and Extern on duty at ER). Each pair of raters used both scores to evaluate each subject and present the mean differences. The mean differences of the FOUR score range from 0.00 (Extern/Extern) to 1.15 (Resident/Resident) and the GCS range from 0.07 (Resident/Nurse) to 0.42 (Nurse/Extern). There were no statistical significant differences between rater types. (Table 3)

Table 4 presents that good to excellent internal consistency was found by measuring Cronbach’s alpha for the FOUR score and the GCS between pairs of raters (Cronbach’s alpha ≥ 0.8). The Cronbach’s alpha values ≥ 0.9 as excellent internal consistency, 0.9 > α ≥ 0.8

### Table 1 Baseline characteristics of patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N=80 (%)</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>19</td>
<td>92</td>
<td>62.23 ± 17.92</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>45 (56.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>35 (43.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of consciousness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drowsiness</td>
<td>51 (63.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stuporous</td>
<td>8 (10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coma</td>
<td>21 (26.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underlying diseases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>none</td>
<td>7 (8.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>33 (41.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>30 (37.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 3</td>
<td>10 (12.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome (hospital discharge)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>death</td>
<td>21 (26.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alive</td>
<td>59 (73.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Min. = minimum; Max. = maximum; SD= standard deviation

### Table 2 Inter-rater reliability using weighted kappa and intra-class correlation (ICC)

<table>
<thead>
<tr>
<th></th>
<th>FOUR score</th>
<th>GCS score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eye Resp. Motor Total</td>
<td>Eye Verbal Motor Total</td>
</tr>
<tr>
<td>Weighted kappa</td>
<td>0.88 0.57 0.81 0.8</td>
<td>0.84 0.81 0.87 0.83</td>
</tr>
<tr>
<td>ICC [95% CI]</td>
<td>[0.93-0.97] [0.93-0.97] [0.96-0.97]</td>
<td>[0.93-0.97] [0.93-0.97] [0.96-0.98]</td>
</tr>
<tr>
<td></td>
<td>[0.65-0.95] [0.92-0.97] [0.92-0.97]</td>
<td>[0.92-0.97] [0.96-0.98] [0.96-0.98]</td>
</tr>
</tbody>
</table>

FOUR = Full Outline of Unresponsiveness; GCS = Glasgow Coma Scale; ICC = Intra-class Correlation Coefficient; CI = Confident Interval
The FOUR Score and Glasgow Coma Scale to Evaluate the Patients with Intubation at Emergency Room

149

as good, $0.8 > \alpha \geq 0.7$ as acceptable, $0.7 > \alpha \geq 0.6$ as questionable, $0.6 > \alpha \geq 0.5$ as poor and as unacceptable if $\alpha < 0.5$. (Table 4)

Table 5 presents the correlation between the FOUR score and GCS which was measured by the rater.

Table 6 shows the relationship between the total FOUR score and the GCS with an outcome of in-hospital deaths. We found that each time the total FOUR score increased 1-point it resulted in a $0.87$ (95% CI 0.68-1.10) times lower risk of experiencing in-hospital mortality under the unadjusted model. (Table 6)
Discussion

The overall rater agreement in this study was good to excellent for both the FOUR and GCS scores by using weighted kappa ($K_w$) and intra-class correlation coefficient (ICC) analyses. The respiration subscale of the FOUR score had the lowest score from $K_w = 0.57$ and ICC = 0.77. The raw data showed errors of the raters when evaluating the scores differently from the protocol in some critical situations in the ER. Previous studies\textsuperscript{5-8,10-11} have found results similar to our findings. All of the studies found that education and experience did not interfere with the agreement of the raters. The level of overall reliability is possibly caused by strong guidance, training and demonstration of the evaluation process for all raters.

The internal consistency for the FOUR score and the GCS score between pairs of raters calculated by Cronbach’ s alpha in this study was good to excellent as was the high scores’ s correlation between the FOUR score and the GCS score analyzed by Spearman’s rho and Kendall’s tau\textsubscript{b}. The results are similar to the prior study of Spearman’s rho\textsuperscript{3,6,9}, on the other hand, this study shows good correlation by Kendall’s tau\textsubscript{b} also.

The FOUR score and the GCS score were used to evaluate within the subject by rater type (Table 3) and showed no statistical significant difference between the raters. These results demonstrated that the FOUR score has no limitations and does not depend on the knowledge or expertise of the evaluators, as well as the GCS score.

The total FOUR score can be used to predict in-hospital mortality. We found that every 1-point increase in the total FOUR score will result in a 0.87 times lower risk of experiencing an in-hospital mortality in the unadjusted model. These results were similar to previous studies.\textsuperscript{5-7}

In our study, all of the patients were in an altered state of consciousness with intubation. This may have improved rater agreement values in the GCS with no interferences with the FOUR score.

Following the research objective, the author has proposed that the FOUR score is reliable regarding the differences of the raters and powerful for predicting in-hospital mortality. Further studies of FOUR score in varying situations should be done to verify the foregone conclusion of the new coma score instead of the GCS score in the near future.

Limitation

There were several limitations to this study. Observation could introduce a variety of bias including those related to ascertainment. Intra-raters reliability had not been evaluated because it was impossible for the rater to score the same patient in such a short time period without remembering the previous scores. This study was produced at one center and was time limited. Further studies may be produced at multicenter to verify the results.

Conclusion

The FOUR score is reliable for evaluating the alterations of the consciousness of patients with intubation in emergency rooms and can be used to predict in-hospital mortality. The Four score can an improvement in the emergency care system for triage, transportation and disaster or critical situations for evaluating the patients if the GCS is limited. The Four score can be used and applied in every emergency room for evaluating all patients.

<table>
<thead>
<tr>
<th>Table 6</th>
<th>FOUR score and GCS relation to In-hospital mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>N (N=80)</td>
</tr>
<tr>
<td>In-hospital death</td>
<td>21</td>
</tr>
</tbody>
</table>
Acknowledgements
Thai Clinical Epidemiology and Research Trial Consortium, Chulalongkorn University and Phramongkutklao Hospital supported the original study.

Disclosure
The author report no conflicts of interest in this work.

References
การเปรียบเทียบการใช้ฟอร์สกอร์กับกลาสโกว์โคมาสเกลในการประเมินผู้ป่วยที่ได้รับการใส่หัวช่วยหายใจที่ห้องฉุกเฉิน

นิติ  เมธีศิริวัฒน์ และ สังศักดิ์ คุปต์กิจศัยกุล

1 Emergency Physician, Department of Trauma and Emergency, Phramongkutklao Hospital;
2 Master of Science Program in Health Development Student, Faculty of Medicine, Chulalongkorn University;
3 Chulalongkorn Hospital, Faculty of Medicine, Chulalongkorn University

วัตถุประสงค์: เพื่อเปรียบเทียบการนำาโฟร์สกอร์กับกลาสโกว์โคมาสเกลมาใช้ในผู้ป่วยที่ได้รับช่วยหายใจในห้องฉุกเฉินจากผู้ประเมินหลายประเภท และการพยากรณ์อัตราการตายในโรงพยาบาลจากคะแนนรวม วิธีการศึกษา: ผู้ป่วย 80 รายได้รับการใส่หัวช่วยหายใจที่ห้องฉุกเฉินจะได้รับการประเมินระดับความรู้สึกทางระบบประสาทโดยใช้ทั้งฟอร์สกอร์และกลาสโกว์โคมาสเกลจากผู้ประเมินโดยการสุ่มเป็นคู่ แล้วนำข้อมูลวัดมาโดยใช้หลักสถิติเพื่อหาความแตกต่างระหว่างกลุ่ม ความสัมพันธ์ระหว่างการวัดระดับความรู้สึกทางระบบประสาท คะแนนรวม หาค่าความน่าเชื่อถือและความสัมพันธ์ในกลุ่มแต่ละกลุ่มและหาอัตราการตายในโรงพยาบาล ผลการศึกษา: คะแนนรวมของผู้ประเมินมีทิศทางไปแนวเดียวกันถึงดีมากทั้ง 2 แบบประเมิน และไม่พบความแตกต่างระหว่างผู้ประเมินทุกๆ 1 คะแนนรวมของผู้ประเมินผู้ป่วยทั้ง 2 แบบที่เพิ่มขึ้นจะมีผลต่อการลดลงของอัตราการตายในโรงพยาบาล สรุปการศึกษา: แบบประเมินฟอร์สกอร์มีความน่าเชื่อถือในการนำาไปใช้ประเมินผู้ป่วยที่ได้รับการใส่ช่วยหายใจที่ห้องฉุกเฉินและยังสามารถนำาไปใช้เพื่อทำนายโอกาสการเสียชีวิตของผู้ป่วยเมื่อเข้ารักษาในโรงพยาบาล

Key Words: ฟอร์สกอร์ กลาสโกว์โคมาสเกล ห้องฉุกเฉิน การใส่หัวช่วยหายใจ เวชสารแพทย์ทหาร版 2555;65:145-52.