Reliability of Robotic Telemedicine for Assessing Critically Ill Patients with the Full Outline of UnResponsiveness Score and Glasgow Coma Scale

Amelia K. Adcock, MD,¹ Heidi Kosiorek, MS,² Prachi Parikh, MD,³ Alyssa Chauncey, BSc,⁴ Qing Wu, MD, ScD,⁵ and Bart M. Demaerschalk, MD, MSc, FAHA, FRCP(C)^{1,6}

¹Divisions of Cerebrovascular Diseases and Neurocritical Care, Department of Neurology, Mayo Clinic College of Medicine, Phoenix, Arizona.

²Department of Research, Biostatistics, Mayo Clinic, Phoenix, Arizona.

³Department of Neurology, Cleveland Clinic, Cleveland, Ohio.

⁴Arizona State University, Tempe, Arizona.

⁵Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada, Las Vegas, Nevada.

⁶Center for Connected Care, Mayo Clinic, Rochester, Minnesota.

Abstract

Purpose: Telemedicine is increasingly utilized in the evaluation of critically ill patients, including those with decreased level of consciousness (LOC) or coma. Improving access to providers with neurologic expertise affords earlier triage and directed patient management. However, objective data regarding the reliability of using standardized coma scales, traditionally employed at the bedside for remote assessment, are largely lacking.

Hypothesis: Bedside and remote assessments of patients with decreased LOC, using either the Glasgow Coma Scale (GCS) or Full scale Of UnResponsiveness (FOUR), score are equivalent.

Methods: Prospective trial comparing the reliability of bedside and remote coma assessments using GCS or FOUR score clinical evaluation tools utilizing robotic telepresence technology. Total scores of the GCS and FOUR score were compared between bedside and remote physician assessors by paired t-test and Pearson correlation coefficient (PCC).

Results: One hundred subjects were enrolled. Mean age was 70.8 (± 15.9) years and the average examination time took 5.16 (± 2.04) minutes. Mean GCS total score at bedside was 7.5 (± 3.67) versus examination conducted remotely 7.23 (± 3.85); difference in scores was 0.25 (± 0.10); $p=0.01$. Mean FOUR total score at bedside was 9.63 (± 4.76) versus

remote 9.21 (± 4.74); difference in scores was 0.40 (± 2.00); $p=0.05$. PCC for GCS was 0.966; $p<0.001$, and for FOUR score 0.912; $p<0.001$. Ninety-five percent of remote providers rated GCS and 89% rated FOUR score as good (4/5) regarding overall satisfaction and ease of use.

Conclusions: Differences between total bedside and remote GCS and FOUR scores were small. Furthermore, PCCs between remote and bedside assessments were excellent: 0.97 (GCS) and 0.91 (FOUR). These results suggest that LOC can be reliably assessed using existing robotic telemedicine technology. Telemedicine could be adopted to help evaluate critically ill patients in neurologically underserved areas.

Keywords: telemedicine, coma, critical care, Glasgow Coma Scale, FOUR Score, intensive care units, remote consultation, neurology

Introduction

Assessing patients with an altered level of consciousness (LOC) or coma in the intensive care unit (ICU) is an essential part of critical care. To date, no easily measurable biomarker or perturbation in vital signs exists to define the clinical syndrome of coma. Therefore, standardized scoring systems are employed to track a patient's progress and make predictions regarding prognosis. These data are based on serial clinical observations at the bedside.

The Glasgow Coma Scale (GCS) is the most commonly used scoring system; however, due to several limitations, including inconsistent interobserver reliability, uncertainty surrounding the scoring of an intubated or sedated patient, and disregard of brainstem reflexes, the Mayo FOUR score has been proposed as an alternative scoring system.¹⁻⁶ The Full Outline of UnResponsiveness (FOUR) score is a coma scale consisting of four categories: (1) eye response, (2) motor response, (3) brainstem reflexes, and (4) respiration, with a maximum score of four points in each category; hence, an overall score ranges from 0 to 16. It has been validated in diverse ICU settings.⁷⁻⁹ Advantages compared

with the GCS include the absence of verbal response (difficult to evaluate in intubated patients) and improved prognostic value, especially in the lowest categories.¹⁰

Reliable assessment is one of the first steps toward improved care in critically ill patients. Prompt recognition of a change in clinical status could allow for timely interventions. In the dynamic setting of brain injury, every minute represents the opportunity to prevent or treat complications. In addition, reliable methods are necessary for predicting functional outcome, determining prognosis, and helping to define the goals of care for critically ill patients and their families. Recent advances in telemedicine networks have improved access to subspecialist (neurologist, neurosurgeon, or neurointensivist) emergency evaluation.¹¹ The need for critical care, together with population dispersion and relative shortage of ICU doctors and facilities with round-the-clock coverage, has created an opportunity for telemedicine.¹²

Tele-ICU services have been proven to shorten length of stay and decrease mortality.¹³ In addition, these connected care networks have been shown to decrease costs and be financially beneficial to both spoke and hub hospitals.¹⁴ Assessing whether clinical coma evaluation can be reliably measured by providers who are not physically at the bedside is a vital step toward improving acute care of patients with neurologic emergencies.

The objective of this study was to measure the reliability of a robotic telemedicine assessment of the Mayo FOUR score and GCS in comatose patients. Reliability between simultaneous independent assessments scored by a remote neurologist using a robotic telemedicine platform and assessments scored by a neurologist at the bedside were prospectively evaluated. In addition, feasibility of using robotic telemedicine technology for remote FOUR score and GCS was assessed.

Methods

The study protocol was approved by Mayo Clinic Institutional Review Board (ID: PR12-008262-03). The procedures were followed in accordance with institutional guidelines.

Subjects were identified among patients admitted to Mayo Clinic Hospital, Phoenix, Arizona, using the ICU, general hospital, neurology, and neurosurgery patient admission lists and prospectively enrolled after informed consent was obtained. Inclusion criteria were pragmatically broad; any adult patient with a reduced LOC (of any depth of coma, mild, moderate, or severe) beyond their emergency department course was eligible. All 16 evaluators were medical doctors (Mayo Clinic consultant staff, fellows, or residents) who participated in study training and passed a competency quiz on

both coma scoring systems. Once consent was obtained, two study physician evaluators were randomized using a random number generator (www.random.org) to either the bedside or remote assessment arms. Remote assessors used an InTouch Health RP7 robot endpoint to conduct their clinical coma evaluations. This is a real-time audio and visual robotic telepresence system that provides communication between patients, hospital staff, and a remote physician. It was approved by the FDA in 2012 for clinical use.¹⁵

Each assessor independently scored each subject using both GCS and Mayo FOUR scales with the assistance of a trained, bedside, telepresenter critical care nurse familiar with coma examination and scales. The bedside evaluator was face-to-face with the subject, while the remote evaluator was working from a different location in a hospital office desktop work station. The remote and bedside evaluators, together with the critical care nurse, were permitted to repeat any portion of their examination, but they were not permitted to otherwise communicate, render opinions, or to remark upon on their clinical observations or their scoring. As such, the two assessors were blinded to each other's scores. Duration of each clinical interaction was recorded. Bedside and remote scoring was statistically analyzed for agreement and correlation.

FOUR SCORE

The FOUR score with a possible score of 0–16 provides a standardized tool for tracking clinical course and lends predictions in the comatose patient.¹⁶

GLASGOW COMA SCALE

The GCS was originally designed to standardize clinical assessments of patients with head trauma in the 1970s, but is currently regarded as the most widely used scoring system for patients with a depressed LOC. It consists of independent observations of three categories: (1) eye opening, (2) best motor response, and (3) best verbal response.¹⁷ The range of possible scores is 3–15.

After each examination, the remote assessor rated their audio–video robotic telemedicine assessment experience using a satisfaction survey (1 = very poor, 2 = poor, 3 = neutral/no opinion, 4 = good, and/or 5 = very good) for the following categories: reception in the hospital, image quality, sound quality, ease of use, and ability to assess subjects using the Mayo FOUR score or GCS.

STATISTICAL METHODS

Bedside scoring was compared with remote GCS and Mayo FOUR scores to determine the intermethod agreement. Paired

t-test and Pearson correlation coefficient (PCC) were employed. Furthermore, weighted kappa (wk) for each GCS and Mayo FOUR category was calculated. wk scores were interpreted according to published standards ($wk > 0.75$ excellent agreement beyond chance, $wk > 0.4$ and < 0.75 moderate agreement beyond chance, and $wk < 0.4$ poor agreement beyond chance).¹⁸ Confidence intervals (95%) were computed. Bland–Altman limits of agreement for the total score were computed. Remote assessments were considered equivalent to bedside assessment if the 95% limits of agreement were within ± 1 point. A sample size of 100 consecutive subjects ensured that the margin of error for correlation was no larger than ± 0.05 if the correlation in the target population is at least 0.90. This sample size also ensured a margin of error no wider than $\pm 4\%$ if the incidence of disagreement in the target population was no more than $\pm 5\%$. SAS version 9.4 (Cary, NC) was used for analysis.

Results

Between November 1, 2012, and February 14, 2014, 100 consecutive subjects admitted to Mayo Clinic Hospital, Phoenix, Arizona, and determined to have a depressed LOC during their hospital course, were enrolled (mean age 70.8 years). Average examination time to conduct both coma scales was 5.16 (± 2.04) minutes (Table 1).

Table 1. Patient Characteristics

Age, mean (SD)	70.8 (± 15.9)
Gender, <i>n</i> (%)	
Male	52 (52)
Female	48 (48)
Cause of coma, <i>n</i> (%)	
Ischemic Stroke	23 (23)
Seizure	16 (16)
Intracerebral hemorrhage	15 (15)
Metabolic-endocrine	13 (13)
Encephalitis	9 (9)
Sepsis	7 (7)
Drugs/poisoning	6 (6)
Anoxic-Ischemic insult	6 (6)
Subarachnoid hemorrhage	3 (3)
Unknown	1 (1)
Length of examination (minutes), mean (SD)	5.2 (2.0)

SD, standard deviation.

Mean GCS total score at bedside was 7.5 (± 3.67) versus examination conducted remotely 7.23 (± 3.85); difference in scores was 0.25 (± 0.10); $p = 0.01$. Mean FOUR total score at bedside was 9.63 (± 4.76) versus remote 9.21 (± 4.74); difference in scores was 0.40 (standard deviation = 2.00); $p = 0.05$. PCC for GCS was 0.97 (95% CI: 0.95, 0.98); $p < 0.001$, and for FOUR score was 0.91 (95% CI: 0.87, 0.94); $p < 0.001$ (Fig. 1). Bland–Altman limits of agreement were -0.52 to 0.51 for GCS and -0.83 to 0.82 for Mayo FOUR score, within the ± 1 point range. Eighty-nine percent (95% CI: 81.2, 94.4) of patients had a < 1 point difference between bedside and remote total score for GCS, whereas 66% (95% CI: 56.5, 75.8) for Mayo FOUR; 10.1% of patients had a ≥ 2 point difference in total score for GCS and 34% for Mayo FOUR, reflecting the overall lower correlation between bedside and remote assessments for Mayo FOUR compared with that of GCS.

Weighted kappa of each category, for GCS, was eyes 0.89 (95% CI: 0.84, 0.94), verbal 0.89 (95% CI: 0.83, 0.95), and motor 0.83 (95% CI: 0.75, 0.90) and, for FOUR score, was eyes 0.83 (95% CI: 0.76, 0.91), motor 0.80 (95% CI: 0.71, 0.89), brainstem 0.64 (95% CI: 0.51, 0.78), and breathing 0.87 (95% CI: 0.78, 0.95). Except for Mayo FOUR brainstem, all subscales had excellent agreement beyond chance.

94.6% of remote providers rated GCS and 89.2% rated FOUR score as good (4/5) for overall satisfaction and ease of use. Mean scores ranged from 4.2 (± 0.72) for ability to assess GCS remotely to 4.8 (± 0.55) for sound quality and ability to assess Mayo FOUR remotely (Fig. 2).

Discussion

The results demonstrate that patients with depressed LOC can be reliably assessed utilizing robotic telemedicine. Utilization of telemedicine might be an appropriate way to expand access to acute neurologic care. Although telemedicine is routinely employed in the evaluation of acute stroke, the telemedicine assessment of patients with depressed levels of consciousness has not been formally validated before. Differences between total bedside and remote GCS and FOUR scores were small. Furthermore, PCCs between remote and bedside assessments were excellent; 0.97 (GCS) and 0.91 (FOUR).

Although both scoring systems were reliable, it is interesting to note that the GCS outperformed the FOUR score, which is contrary to prior in-person assessments asserting the FOUR score's superior inter-rater reliability.^{8,14,19–21} We postulate our results are most readily explained by the FOUR score's relatively weaker weighted kappa of brainstem reflexes; 0.64 (brainstem FOUR). In spite of the high-definition

Fig. 1. Correlation between bedside vs remote evaluation using GCS and Mayo FOUR Score. GCS, Glasgow Coma Scale; FOUR, Full scale of UnResponsiveness.

camera, zoom, and pan-tilt technology of the InTouch Health robotic telemedicine and adjustment of the environmental lighting conditions, brainstem and particularly pupillary responses proved challenging to judge remotely. Overall, technical satisfaction was high and evaluation was completed in a timely manner. GCS was rated slightly more favorably than the FOUR score (95% vs. 89%), which is likely a further reflection of the challenges associated with the remote brainstem assessment. These results suggest that LOC can be reliably assessed using existing InTouch Health robotic telemedicine. Telemedicine evaluation could be employed to improve early evaluation of critically ill patients. Validation represents an important first step in banishing uncertainty and helping to establish telemedicine coma assessment with GCS and FOUR score as a valid clinical use case much like the National Institutes of Health Stroke Scale (NIHSS) for stroke severity in years past.^{22–24}

In addition, the results strengthen the use of telemedicine assessment of patients with a decreased LOC in clinical studies investigating telemedicine, tele-ICU, or acute treatment of neurologic emergencies. We would propose that

future studies incorporate commonly available peripheral technology such as a pupilometer to potentially improve the brainstem assessment, the FOUR score's lowest performing category. Other important future work would be to promote research designed to empirically determine what the minimal clinically significant change is in both coma scales. The mean difference between bedside GCS (7.50) and remote (7.23) was statistically significant; however, we fail to see how a GCS of 7.50, as opposed to 7.23, is clinically significant or relevant to the patient's care. For example, in GCS, the division of scores into categories of mild (13–15), moderate (9–12), and severe (3–8) suggests that unless mean score differences approached a range of 1–3, the clinical significance is relatively small. Uncertainty surrounding the definition of a clinically relevant change in either the FOUR score or GCS remains a major confounder for clinical implications in general.²⁵ Future analysis of our data and patient outcomes could help clarify this critical issue.

Advantages of this study include its real-world design with the majority of patients presenting with significantly decreased LOC (GCS <8) and unbiased approach. Examinations

Fig. 2. Remote satisfaction with GCS and Mayo FOUR. 1=very poor, 2=poor, 3=neutral/no opinion, 4=good, and 5=very good.

were conducted simultaneously by physicians who would routinely be taking care of these patients and using their assessments to dictate care decisions, as opposed to use of simulated patients. Limitations include the uncontrollable environment of the examination itself (family presence could have affected the assessment, order of examination components was not randomized). The lack of a standardized testing environment is a product of the study's design. A further limitation could be its generalizability due to the proclivity of neurologic diagnoses and neurologist assessors. Although our subject population represented a typical breakdown of conditions seen in a neurointensive care unit, it was not necessarily similar to populations in general, surgical, cardiac, or pediatric ICUs.

Conclusion

The results indicate that remote and bedside evaluation of patients with LOC using commonly employed clinical coma scales in a real-world environment demonstrates excellent agreement. Their reliability supports telemedicine expansion into the ICU and into other areas where acute neurologic care is required. The remote GCS modestly outperformed the FOUR score's in inter-rater reliability. We postulate that this is largely due to the difficulty of remote assessment of the FOUR score brainstem component. However, both scales were rated highly overall by our study's assessors. Evalua-

tions were completed in less than 5 min and were easy to administer. Studying the addition of a peripherally attached pupilometer's effect on reliability of robotic telemedicine assessments of patients in coma is recommended as future research.

Acknowledgments

The authors acknowledge Drs. T. Overall, P. Dhwan, H. Yancy, J. Lee-Iannotti, C. O'Carroll, D. Capampangan, M. Aguilar, C. Kramer, M. Rubin, S. Spritzer, T. Bravo, G. Schlossler, P. Barr, and M. Rady for their efforts in this research. This work was supported by a Mayo Clinic Center for Translational Science Activities grant (UL RR0244150; REDCap project).

Disclosure Statement

No competing financial interests exist.

REFERENCES

- Jennett B, Teasdale G. Aspects of coma after severe head injury. *Lancet (London, England)* 1977;1:878-881.
- Kelly CA, Upex A, Bateman DN. Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow Coma Scale. *Ann Emerg Med* 2004;44:108-113.
- Gill M, Martens K, Lynch EL, Salih A, Green SM. Interrater reliability of 3 simplified neurologic scales applied to adults presenting to the emergency department with altered levels of consciousness. *Ann Emerg Med* 2007;49:403-407, 407.e401.
- Batchelor J, McGuinness A. A meta-analysis of GCS 15 head injured patients with loss of consciousness or post-traumatic amnesia. *Emerg Med J* 2002;19:515-519.
- Green SM. Cheerio, laddie! Bidding farewell to the Glasgow Coma Scale. *Ann Emerg Med* 2011;58:427-430.
- Fleiss JL. *Statistical methods for rates and proportions*. New York, NY: John Wiley & Sons, Inc., 1981:212-236.
- Wijdicks EF, Bamlet WR, Maramattom BV, Manno EM, McClelland RL. Validation of a new coma scale: The FOUR score. *Ann Neurol* 2005;58:585-593.
- Wolf CA, Wijdicks EF, Bamlet WR, McClelland RL. Further validation of the FOUR score coma scale by intensive care nurses. *Mayo Clinic Proc* 2007;82:435-438.
- Stead LG, Wijdicks EF, Bhagra A, et al. Validation of a new coma scale, the FOUR score, in the emergency department. *Neurocrit Care* 2009;10:50-54.
- Wijdicks EF, Rabinstein AA, Bamlet WR, Mandrekar JN. FOUR score and Glasgow Coma Scale in predicting outcome of comatose patients: A pooled analysis. *Neurology* 2011;77:84-85.
- Wechsler LR, Tsao JW, Levine SR, et al. Teleneurology applications: Report of the Telemedicine Work Group of the American Academy of Neurology. *Neurology* 2013;80:670-676.

12. Fuhrman SA, Lilly CM. ICU Telemedicine Solutions. *Clin Chest Med* **2015**; 36:401–407.
13. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS. A multicenter study of ICU telemedicine reengineering of adult critical care. *Chest* **2014**;145:500–507.
14. Kumar S, Merchant S, Reynolds R. Tele-ICU: Efficacy and cost-effectiveness of remotely managing critical care. *Perspect Health Inf Manag* **2013**;1.
15. FDA approval of InTouch Telepresence 510 (k) Submission. Available at www.accessdata.fda.gov/.../K1232 (last accessed July 11, 2016).
16. Fischer M, Ruegg S, Czaplinski A, et al. Inter-rater reliability of the Full Outline of UnResponsiveness score and the Glasgow Coma Scale in critically ill patients: A prospective observational study. *Crit Care* **2010**;14:R64.
17. Juarez VJ, Lyons M. Interrater reliability of the Glasgow Coma Scale. *J Neurosci Nurs* **1995**;27:283–286.
18. Fleiss JL. *Statistical methods for rates and proportions*. New York, NY: John Wiley & Sons, Inc., **1981**:212–236.
19. Cohen J. Interrater reliability and predictive validity of the FOUR score coma scale in a pediatric population. *J Neurosci Nurs* **2009**;41:261–267.
20. Marcati E, Ricci S, Casalena A, Toni D, Carolei A, Sacco S. Validation of the Italian version of a new coma scale: The FOUR score. *Intern Emerg Med* **2012**;7:145–152.
21. Sadaka F, Patel D, Lakshmanan R. The FOUR score predicts outcome in patients after traumatic brain injury. *Neurocrit Care* **2012**;16:95–101.
22. Schwamm LH, Holloway RG, Amarenco P, et al. A review of the evidence for the use of telemedicine within stroke systems of care: A scientific statement from the American Heart Association/American Stroke Association. *Stroke* **2009**;40:2616–2634.
23. Wang S, Lee SB, Pardue C, et al. Remote evaluation of acute ischemic stroke: Reliability of National Institutes of Health Stroke Scale via telestroke. *Stroke* **2003**;34:18.
24. Shafqat S, Kvedar JC, Guanci MM, Chang Y, Schwamm LH. Role for telemedicine in acute stroke. Feasibility and reliability of remote administration of the NIH stroke scale. *Stroke* **1999**;30:2141–2145.
25. Prasad K. The Glasgow Coma Scale: A critical appraisal of its clinimetric properties. *J Clin Epidemiol* **1996**;49:755–763.

Address correspondence to:

Amelia K. Adcock, MD
Mayo Clinic College of Medicine
West Virginia University
1 Medical Center Drive
Suite 7500, PO Box 9180
Morgantown, WV 26506

E-mail: akadcock@hsc.wvu.edu

Bart M. Demaerschalk, MD, MSc, FAHA, FRCP(C)
Mayo Clinic Hospital
5777 East Mayo Boulevard
Phoenix, AZ 85054

E-mail: demaerschalk.bart@mayo.edu

Received: October 14, 2016

Revised: November 4, 2016

Accepted: November 7, 2016