SHORT COMMUNICATION

Inter-rater reliability of the Johns Hopkins Highest Level of Mobility Scale (JH-HLM) in the intensive care unit

Stephanie Hiser [a,b,1], Chi Ryang Chung [b,c,d,1], Amy Toonstra [e],
Lisa Aronson Friedman [b,d], Elizabeth Colantuoni [b,f], Erik Hoyer [a,b,g],
Dale M. Needham [a,b,d,*]

[a] Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[b] Outcomes After Critical Illness and Surgery (OACIS) Group, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[c] Departments of Medicine and Critical Care Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Gangnam-gu, Seoul, Republic of Korea
[d] Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[e] Concordia University, St. Paul, MN, USA
[f] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
[g] Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract
Background: The Johns Hopkins Highest Level of Mobility (JH-HLM) scale is used to document the observed mobility of hospitalized patients, including those patients in the intensive care unit (ICU) setting.
Objective: To evaluate the inter-rater reliability of the JH-HLM, completed by physical therapists, across medical, surgical, and neurological adult ICUs at a single large academic hospital.
Methods: The JH-HLM is an ordinal scale for documenting a patient's highest observed level of activity, ranging from lying in bed (score = 1) to ambulating $\geq$250 feet (score = 8). Eighty-one rehabilitation sessions were conducted by eight physical therapists, with 1 of 2 reference physical therapist rater simultaneously observing the session and independently scoring the JH-HLM. The intraclass correlation coefficient was used to determine the inter-rater reliability.
Results: A total of 77 (95%) of 81 assessments had perfect agreement. The overall intraclass correlation coefficient for inter-rater reliability was 0.98 (95% confidence interval: 0.96, 0.99), with similar scores in the medical, surgical, and neurological ICUs. A Bland–Altman plot revealed a mean difference in JH-HLM scoring of 0 (limits of agreement: −0.54 to 0.61).

* Corresponding author at: 1830 East Monument Street, 5th Floor, Baltimore, MD 21287, USA.
 E-mail: dale.needham@jhmi.edu (D.M. Needham).
[1] Theses authors contributed equally to this project.

Conclusion: The JH-HLM has excellent inter-rater reliability as part of routine physical therapy practice, across different types of adult ICUs.

Introduction

Advances in critical care medicine have resulted in more patients surviving critical illness; however, such survivors frequently experience sustained impairments in their physical functioning.[1] Evaluating physical function early and longitudinally throughout the patients' entire hospital stay is critical in assessing effectiveness of rehabilitation interventions and considering functional prognosis.[2]

A few scales were designed specifically for measuring the highest level of mobility achieved by a patient in the intensive care unit (ICU).[2] However, these scales (e.g. ICU mobility scale, Manchester Mobility Score, Surgical Optimal Mobility Scale) have not been evaluated both inside and outside of the ICU setting.[3–5] Evaluating and using one scale to measure mobility throughout the entire hospital stay would be beneficial. Notably, as there are often different physical therapists treating the same patient throughout the patient's ICU stay, the scale must be reliable among different physical therapists; thus, it is essential to evaluate instrument's inter-rater reliability.

The Johns Hopkins Highest Level of Mobility (JH-HLM) scale is used for both documenting observed patient mobility and establishing mobility goals for patients in the hospital.[6] The JH-HLM has been used by physical therapists, in ICUs, step-down units, and hospital wards, for at least five years.[6] The JH-HLM has demonstrated excellent inter-rater reliability (intraclass correlation coefficient [ICC] 0.99; 95% confidence interval [CI] = 0.98, 0.99) among non-critically ill neurological patients.[7] The validity of the JH-HLM was also evaluated demonstrating strong correlations (>0.60) with the Activity Measure for Post-Acute Care (AM-PAC), Inpatient Mobility Short Form (IMSF), and Katz Activities of Daily Living (Katz ADL) Scale. The JH-HLM's standard error of measurement (SEM) and minimal detectable change (MDC) have been reported to be 0.2 and 0.6 points, respectively.[7] However, there is little evaluation of the JH-HLM in the ICU setting or across a variety of diagnoses where patients' mobility may be more limited than in prior psychometric evaluations. Hence, our objective was to evaluate the inter-rater reliability of the JH-HLM, completed by physical therapists as part of usual practice, across medical, surgical, and neurological ICUs at a single, large academic hospital.

Methods

Design and participants

As done in our prior publication,[8] inter-rater reliability was prospectively assessed across 4 ICUs with a total of 66 ICU beds (for patients with medical, surgical, and neurological diagnoses) at a single large academic hospital in the United States. Clinical physical therapists, who had been previously trained on the JH-HLM, scored this scale during routine patient physical therapy sessions while one of two reference rater physical therapists independently scored the JH-HLM based on observation. The reference raters [SH and AT] were two of the investigators in this evaluation and had extensive knowledge and experience using the JH-HLM. All clinical physical therapists are trained on the JH-HLM upon hire and are required to document patient status using this tool during every physical therapy session. Prior to starting this evaluation, the two reference raters provided a one-time, one-hour, in-person training session that reviewed the instrument and its frequently asked questions document, based on publicly available JH-HLM information.[9]

All eight clinical physical therapists worked across all ICUs. Both the physical therapists and reference raters were blinded to each other's score.[8] These inter-rater reliability assessments were completed whenever both a physical therapist and reference rater were available for the assessment during the 6-month time period from November 2016 to May 2017. Because this instrument is used for all patients, there were no restrictive eligibility criteria for patient selection. Consistent with the objectives of this project, any patient receiving evaluation by a physical therapist, as part of usual clinical care, was eligible to be included. The patients were selected by the clinical physical therapists as part of their routine patient care and availability of reference rater. However, the reference raters were careful as to not assess the same patient with the same clinical physical therapist on consecutive days to avoid recall of patients' previous performance. The evaluations were part of usual care and quality assurance and did not require patient consent. Collection of patient data from the electronic medical record for purposes of publication was approved by the Institutional Review Board at Johns Hopkins University, Baltimore, Maryland, USA (NA00048180). Additionally, the guidelines for reporting reliability and agreement studies (GRRAS) was followed.[10]

Instrument

The JH-HLM is a 1-item scale with eight ordinal responses scored based on patient's observed highest level of activity, as follows: ''1 = only lying, 2 = bed activities, 3 = sitting at edge of bed, 4 = transferring to chair, 5 = standing for greater than or equal to 1 min, 6 = walking 10 or more steps, 7 = walking approximately 7.5 m or more (25 ft or more), and 8 = walking approximately 75 m or more (250 ft or more)''.[6] The physical therapist scores the patient based on highest level achieved during the physical therapy session.

Statistical analysis

Statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC). Patient and physical therapist characteristics were sumarized using descriptive statistics. The inter-rater reliability was calculated using linear regression with a random intercept for each physical therapy session. The ICC (1, 1) was calculated as the ratio of the estimated variance for the random physical therapy sessions to the estimated total variance (total of the variances of physical therapy sessions plus residual error).[11]

To assess agreement between physical therapists and reference raters, a Bland–Altman plot was constructed, including 95% limits of agreement for the mean difference between the physical therapist and reference rater scores.[12]

As in our prior publication,[8] a simulation was performed to calculate the margin of error for the ICC's 95% CI using the following assumptions: (1) JH-HLM administered during 80 physical therapy sessions, (2) JH-HLM variation was consistent with prior data,[7] and (3) true ICC = 0.80.[8] The calculated margin of error was 0.05, indicating that a sample size of 80 physical therapy sessions was appropriate for using ICC to estimate inter-rater reliability.

Results

The eight physical therapists had a median [interquartile range (IQR)] of 2 [1, 3] years of ICU clinical experience. Among 76 patients there were a total of 81 physical therapy sessions observed by a reference rater physical therapist, with five patients evaluated twice due to readmission or extended ICU course. The patient cohort had a median [IQR] age of 64 [56, 71] years with 57% male, and majority having respiratory failure or a neurological disorder (26% and 21%, respectively) as their primary ICU admission diagnostic category. The median [IQR] length of ICU and hospital stay prior to the assessment was 6 [3, 14] and 8 [4, 24] days, respectively.

The mean ± standard deviation (SD) JH-HLM score for each type of ICU and across all assessements was 5 ± 2 points, with 10 assessments (12%) achieving the highest score of 8 point, and none of the assessment (0%) achieving the lowest possible score of 1 point.

A total of 77 of 81 (95%) assessments had perfect agreement. Across all evaluations, the ICC was 0.98 (95% CI: 0.96, 0.99), with similar values for each type of ICU (Table 1). The mean difference in JH-HLM scoring was 0 (95% limits of agreement: −0.54 to 0.61) and proportional bias of 1 (0.96–1.03) [see Bland–Altman plot (Fig. 1)].

Discussion

We evaluated the inter-rater reliability of the JH-HLM performed by physical therapists during routine clinical practice across different types of adult ICUs. The overall inter-rater reliability was excellent (0.98; 95% CI: 0.96, 0.99) among 8 clinical physical therapists with a median of 2 years of ICU clincal experience.

The JH-HLM has comparable measurement properties to other instruments evaluating mobility of ICU patients. The SICU Optimal Mobility Score is a single-item, 5-point ordi-
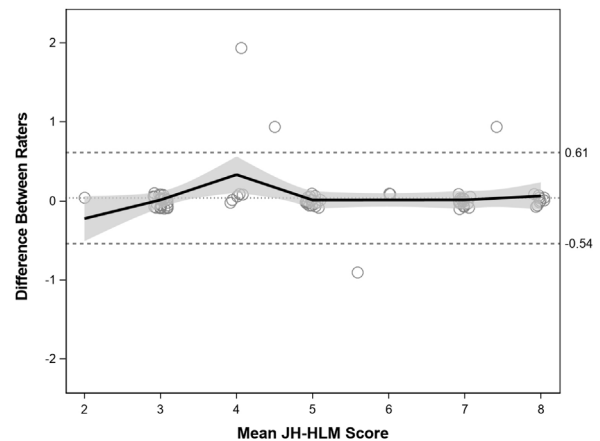

Figure 1 Bland–Altman plot of agreement between JH-HLM scores between 8 physical therapists and a reference rater physical therapist.
Legend: The mean difference between the physical therapists and reference rater is represented by the dotted line. The upper and lower limits of agreement around the mean difference are illustrated by 2 dashed lines. The bold black line represents a fitted locally weighted regression Loess curve, with the shaded area representing 95% confidence bounds. Abbreviation: JH-HLM, Johns Hopkins Highest Level of Mobility Scale.

nal scale ranging from 0 (no activity) to 4 (ambulation).[13] A previous study in 113 patients reported kappa statistics of 0.80 and 0.85, between two nurses and between nurses and an expert team, respectively.[4] The Manchester Mobility Score is a 7-point scale ranging from 1 (in bed interventions) to 7 (mobilizing >30 m) has a kappa statistic of 1.0 among 111 patients in various ICUs.[3] The ICU Mobility Scale is an 11-point scale ranging from 0 (lying in bed) to 10 (walking independently without a gait aid) with inter-rater reliability between junior and senior physical therapists having a weighted kappa (95% CI) of 0.83 (0.76, 0.90).[5] These studies limited their assessment to patients in the ICU without published evidence, to-date, of measurement properties across other non-ICU inpatient settings, as previously demonstrated for the JH-HLM.[7]

The JH-HLM may have utility across the hospital stay (e.g. both in the ICU and outside of the ICU) to facilitate documenting observed patient mobility, setting mobility goals,[14] and improving inter-professional communication of mobility status. The ICC in the current evaluation was consistent with a prior study reporting an ICC of 0.99 (95% CI: 0.98, 0.99) among non-critically ill neurological inpatients.[7] JH-HLM can be assessed quickly and easily given its simplicity. The JH-HLM is similar to the Manchester Mobility Score and ICU Mobility Scale with hierarchical ordinal scales. However, unlike the JH-HLM, these two scales consider the level of assistance required to perform the tasks. The SICU Optimal Mobility Score is different from the other scales because it was designed to evaluate the patient's capacity to perform tasks, rather than as a direct measure of observed mobility.

This study was conducted at a single academic hospital with 8 physical therapists, who had relatively little ICU clinical experience and had received training in use of the JH-HLM. Hence, generalizability of these findings may be

Table 1 Inter-rater reliability of the Johns Hopkins Highest Level of Mobility Scale, by ICU Type.[a]

Type of ICU	Assessments completed, n (%)	Intraclass correlation coefficient[b] (95% confidence interval[c])	Total variance estimate
Surgical	27 (33%)	1.00 (1.00, 1.00)	3.13
Medical	34 (42%)	0.97 (0.93, 0.99)	3.67
Neurological	20 (25%)	1.00 (1.00, 1.00)	2.84
Overall	81 (100%)	0.98 (0.96, 0.99)	3.39

ICU, intensive care unit.

[a] Inter-rater reliability evaluated for 8 clinical physical therapists compared with 1 of 2 reference rater physical therapists.

[b] Calculated via a random effects linear regression model.

[c] Calculated using a non-parametric bootstrap method.

limited. Further evaluation with other physical therapists and in other institutions is encouraged.

Conclusion

The JH-HLM has excellent inter-rater reliability across different adult ICU settings. This scale may allow physical therapists to reliably measure patient mobility in the ICU and across the hospital as part of routine clinical practice.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

References

1. Fan E, Dowdy DW, Colantuoni E, et al. Physical complications in acute lung injury survivors. *Crit Care Med*. 2014;42(4):849–859, http://dx.doi.org/10.1097/CCM.0000000000000040.

2. Parry SM, Huang M, Needham DM. Evaluating physical functioning in critical care: considerations for clinical practice and research. *Crit Care*. 2017;21(1):249, http://dx.doi.org/10.1186/s13054-017-1827-6.

3. Mcwilliams D, Weblin J, Atkins G, et al. Enhancing rehabilitation of mechanically ventilated patients in the intensive care unit: a quality improvement project. *J Crit Care*. 2015;30(1):13–18, http://dx.doi.org/10.1016/j.jcrc.2014.09.018.

4. Kasotakis G, Schmidt U, Perry D, et al. The surgical intensive care unit optimal mobility score predicts mortality and length of stay. *Crit Care Med*. 2012;40(4):1122–1128, http://dx.doi.org/10.1097/CCM.0b013e3182376e6d.

5. Hodgson C, Cardio P, Physio B, et al. Feasibility and inter-rater reliability of the ICU mobility scale. *Hear Lung J Acute Crit Care*. 2014;43(1):19–24, http://dx.doi.org/10.1016/j.hrtlng.2013.11.003.

6. Hoyer EH, Friedman M, Lavezza A, et al. Promoting mobility and reducing length of stay in hospitalized general medicine patients: a quality-improvement project. *J Hosp Med*. 2016;11(5):341–347, http://dx.doi.org/10.1002/jhm.2546.

7. Hoyer EH, Young DL, Klein LM, et al. Toward a common language for measuring patient mobility in the hospital: reliability and construct validity of interprofessional mobility measures. *Phys Ther*. 2018;98(133).

8. Hiser S, Toonstra A, Friedman LA, Colantuoni E, Connolly B, Needham DM. Interrater reliability of the functional status score for the intensive care unit. *J Acute Care Phys Ther*. 2018;9(4):186–192, http://dx.doi.org/10.1097/JAT.0000000000000086.

9. Physical medicine and rehabilitation activity and mobility promotion (AMP). https://www.hopkinsmedicine.org/physical_medicine_rehabilitation/education_training/amp/toolkit.html.

10. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64(1):96–106, http://dx.doi.org/10.1016/j.jclinepi.2010.03.002.

11. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428 https://na-mic.org/w/images/4/4b/Shrout_and_fleiss_ICC.pdf.

12. Altman DG, Bland JM. Measurement in Medicine: The analysis of method comparison studies. *Source J R Stat Soc Ser D (The Stat)*. 1983;32(3):307–317 http://www.jstor.org/stable/2987937.

13. Schaller SJ, Anstey M, Blobner M, et al. Early, goal-directed mobilisation in the surgical intensive care unit: a randomised controlled trial. *Lancet*. 2016;388(10052):1377–1388, http://dx.doi.org/10.1016/S0140-6736(16)31637-3.

14. Pandullo SM, Spilman SK, Smith JA, et al. Time for critically ill patients to regain mobility after early mobilization in the intensive care unit and transition to a general inpatient floor. *J Crit Care*. 2015;30(6):1238–1242, http://dx.doi.org/10.1016/j.jcrc.2015.08.007.