Journal of Critical Care 29 (2014) 475.e1–475.e5

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.jccjournal.org

Initial interrater reliability for a novel measure of patient mobility in a cardiovascular intensive care unit ☆

Ricardo Kenji Nawa, PT, MSc [a], Colleen Lettvin, MSPT, CCS [b], Chris Winkelman, RN, PhD, CCRN, FCCM [c], Paulo Roberto Barbosa Evora, MD, PhD [a], Christiane Perme, PT, CCS [b,*]

[a] Ribeirão Preto School of Medicine, University of São Paulo, Ribeirão Preto, SP, Brazil
[b] Houston Methodist Hospital, Department of Rehabilitation Services, Houston, TX, USA
[c] Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, USA

ARTICLE INFO

Keywords:
Early mobility
Intensive care unit
Mobility status
Physical therapy
Early ambulation

ABSTRACT

Purpose: The purpose of this study was to determine the interrater reliability of the Perme Intensive Care Unit Mobility Score.
Materials and methods: This was a prospective observational study. Data were collected from 20 patients admitted in a cardiovascular intensive care unit. The interrater reliability was tested using the intraclass correlation coefficient with 95% confidence interval and the Cohen κ coefficient of 0.9, using a 2-tailed α of .05 to provide a 90% power.
Results: The 15 items of the Perme ICU Mobility Score were individually analyzed. Interrater reliability (Cohen κ coefficient) for items 2, 3, 5, 6, 7, 8, 13, and 15 was 1.0000; for item 1, 0.8276; item 4, 0.8000; item 9, 0.6000; item 10, 0.7297; item 11, 0.7260; item 12, 0.7872; and item 14, 0.9048; the intraclass correlation coefficient (95% confidence interval) was 0.9880 (0.97743-0.99859).
Conclusions: The Perme ICU Mobility Score is a reliable tool to assess mobility status of patients admitted to the cardiovascular intensive care unit in a specific moment in time, which can be an important tool for research and clinical practice.

1. Introduction

Early mobility activities in intensive care unit (ICU) are advocated as a strategy to improve functional recovery during and after prolonged critical illness, reduce ICU-acquired weakness, and delirium but also to decrease hospital and ICU stay [1,2]. There is a lack of a reliable and valid ICU-specific tool to assess mobility status during critical illness [3].

Outcomes reported in investigations of early mobility activities are primarily length of ICU and hospital stay, duration of mechanical ventilation, muscle strength, function, and mortality [4]. However, none of these outcomes are a specific and sensitive assessment of patient mobility status. Scales and tools commonly used to measure functional status, a concept related to mobility, are the Barthel Index [5], Functional Independence Measure [6], Physical Function ICU Test (PFIT) [7], and Functional Status Score for Intensive Care Unit (FSS-ICU) [3]. None of the above tools, however, included any specific ICU-related items that may impact functional mobility.

Establishing the interrater reliability of a tool is essential to the reproducibility of measurements and the accuracy of inferences drawn from such measurements [8]. The Perme ICU Mobility Score is a novel tool, which was developed for patients in any ICU setting and also with any diagnosis. The Perme ICU Mobility Score consists of 15 items grouped in 7 categories: mental status, potential mobility barriers, functional strength, bed mobility, transfers, gait, and endurance. The category "potential mobility barriers" includes ICU environment-specific items that can potentially interfere with progressive mobility [9]. The Perme ICU Mobility Score, presented in the Appendix, ranges from 0 to 32, and the scale uses a maximum range of 2 to 4 points for each of the 15 items. A high total score indicates few mobility barriers and minimal assistance required for mobility activities, and a low score indicates more barriers to mobility and greater assistance needed for mobility.

The purpose of this study was to test the interrater reliability of the Perme ICU Mobility Score for patients admitted to the cardiovascular ICU (CVICU). This setting was chosen because of the high acuity of patients, steady census, and an established culture and practice of early and progressive mobility.

☆ All of the authors report that they have no actual or potential personal or financial conflicts of interest in the subject matter of this report.

* Corresponding author. Houston Methodist Hospital, Medical Center, Department of Rehabilitation Services, 6565 Fannin, M1-024, Houston, Texas 77030, USA. Tel.: +1 713 441 2675; fax: +1 713 441 0143.
E-mail addresses: ricardo.nawa@usp.br (R.K. Nawa), calettvin@houstonmethodist.org (C. Lettvin), chris.winkelman@case.edu (C. Winkelman), prbevora@fmrp.usp.br (P.R.B. Evora), cperme@houstonmethodist.org (C. Perme).

0883-9441/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcrc.2014.01.019

2. Methods

This was a prospective observational study. Two members of the research team simultaneously collected the Perme ICU Mobility Scores. Data for this prospective observational study were collected between February 2013 and March 2013. Twenty consecutive patients who were admitted to the 40-bed CVICU at Houston Methodist Hospital, Houston, TX, were enrolled in this study. The inclusion criteria were patients older than 18 years who were referred to physical therapy and who met criteria to participate in physical therapy according to the CVICU guidelines as proposed by Perme and Chandrashekar [10]. A sample size of 19 subjects was determined to be necessary to achieve an intraclass correlation (ICC) of 0.9, using a 1-sided α value of .05 and power of 0.80 [11]. A sample size of at least 19 subjects, using the same values for α and power, was also calculated as adequate to determine κ values of 0.85 or greater for 2 raters [12].

There were no exclusion criteria. The Houston Methodist Hospital Institutional Review Board approved this study, and informed consent was waived, as this was a prospective observational study that conformed to current practices in the CVICU.

3. Procedure

The raters selected to participate in this study did not contribute in the initial development of the tool, and before the initiation of this study, they were trained in how to obtain the Perme ICU Mobility Score. Rater "A" has 11 years of clinical experience and is board certified by American Board of Physical Therapy Specialties as a cardiovascular and pulmonary clinical specialist. Rater "B" is also a physical therapist and has more than 5 years of clinical experience in addition to participating in a 1-year ICU-focused clinical fellowship. The 2 raters collected the score after the initial physical therapy evaluation. Rater "A" assessed the patient, and the rater "B" observed the entire process but did not physically touch the patient as part of the assessment. Both raters completed the score sheet immediately after finishing the physical therapy evaluation. The score sheets were completed separately without any communication between the raters in an attempt to avoid bias. The total time spent to fill out the score sheet by both raters was collected to estimate the time taken for the completion of the tool. The rater "B" was responsible for collecting the nonidentifiable patient characteristics; Acute Physiology and Chronic Health Evaluation (APACHE) II, ICU and hospital length of stay, and discharge location.

4. Statistical analysis

Data are presented using descriptive analysis including median and interquartile range (IQR) for continuous data and percentages for categorical data. The ICC was used to quantify relative interrater reliability. It provides limited information on measurement error, which means that a high ICC does not necessarily indicate a low possibility of measurement error [13,14].

The Cohen κ coefficient was also used to determine the agreement between observers. It measures the degree of agreement between the scores and includes a correction for any agreement, which may occur by chance [15,16]. The strength of agreement associated with the κ statistics was classified according to Landis and Koch [15]. Bland-Altman plots were also planned to compare interrater agreement. All statistical analyses were performed using STATA version 11 software (StataCorp, College Station, TX).

5. Results

The 20 subjects enrolled in the study had a median (IQR) of 64.5 (20-86) years, and the characteristics are summarized in Table 1. At the time of the initial physical therapy evaluation, the mean Perme ICU Mobility Score was 21.5 (SD, 7.10), 15% of the patients were on mechanical ventilation, and 65% required intravenous drips. Fifty percent of the patients were able to ambulate on the first physical therapy session. Of these patients, 4 patients were able to ambulate more than 100 ft in 2 minutes.

Table 1
Patient characteristics

Characteristics	Median/percentage, n = 20
Age (y) median (IQR)	64.5 (20-86)
Sex	
Male	60%
Female	40%
BMI (kg/m^2)	25.15 (13.3-40.9)
LOS, median (IQR)	
ICU (d)	4 (1-42)
Hospital (d)	18 (4-73)
APACHE II score, median (IQR)	16.5 (7-30)
Surgical procedures/diagnosis	
Lung transplant	7 (35%)
Coronary artery bypass	4 (20%)
Pneumonia	2 (10%)
LVAD	1 (5%)
Other	6 (30%)
Discharge location	
Home	9 (45%)
LTAC	5 (25%)
SNF	3 (15%)
Other	3 (15%)

BMI indicates body mass index; LOS, length of stay; LVAD, left ventricular assist device; LTAC, long-term acute care; SNF, skilled nursing facility; Other, hospice, nursing home.

The weighted Cohen κ coefficient ranged from 0.60 to 1.0, and the percentage of agreement ranged from 80% to 100% for each item as detailed in Table 2. Interrater reliability (Cohen κ coefficient) was 1.0 for items 2, 3, 5, 6, 7, 8, 13, and 15. Item 9 ($\kappa = 0.6000$) had the lowest κ value and agreement. The Bland-Altman plot for the agreement between the observers is illustrated in Fig. 1 and supports the close agreement. The interrater reliability, estimated from ICC, was 0.98, and results from this calculation are in Table 3. An item-by-item with the total score for both raters is presented in Table 4.

The average time to fill out the Perme ICU Mobility Score was 1 minute and 12 seconds. The time to complete the score was obtained from a mean of 40 assessments by 2 observers for the 20 patients enrolled in the study.

6. Discussion

This report is an initial step in determining the consistency of scoring in a new tool to test the interrater reliability of the Perme ICU

Table 2
The agreement between 2 observers

Perme ICU Mobility Score item	n = 20		
	Agreement (%)	Expected (%)	κ
1. Alertness	95.00	71.00	0.8276
2. Follow commands	100.00	100.00	1.0000
3. Mechanical ventilation	100.00	74.50	1.0000
4. Pain	90.00	50.00	0.8000
5. Lines and tubes	100.00	90.50	1.0000
6. Drips	100.00	54.50	1.0000
7. Legs	100.00	66.50	1.0000
8. Arms	100.00	81.50	1.0000
9. Supine to sit	80.00	50.00	0.6000
10. Static sitting balance	95.00	81.50	0.7297
11. Sit to stand	85.00	45.25	0.7260
12. Static standing balance	90.00	53.00	0.7872
13. Transfer from bed to chair	100.00	51.50	1.0000
14. Gait	95.00	47.50	0.9048
15. Endurance	100.00	34.00	1.0000

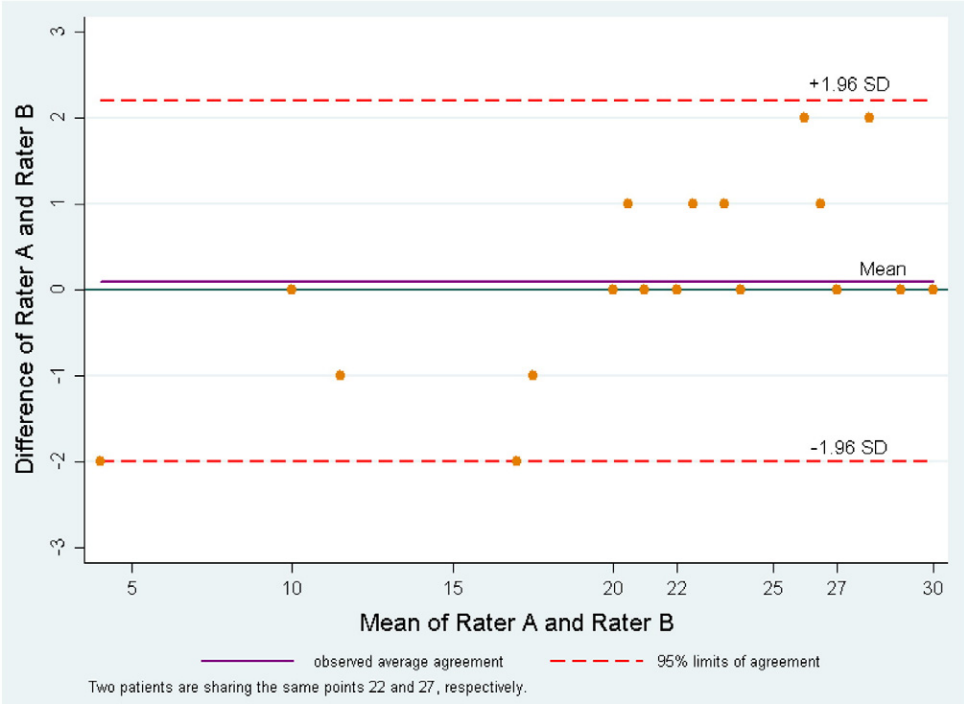

Fig. 1. Bland-Altman plot for the agreement between the observers.

Mobility Score for patients admitted to the CVICU. As calculated a priori, the sample of 20 subjects provided sufficient power (>0.9) to evaluate the tool for interrater reliability. However, we recognize that a larger and more heterogeneous sample will be required for additional psychometric evaluation before recommending wider adoption. Results supported the Perme ICU Mobility Score as a reliable tool in this small sample size of patients who were admitted to the CVICU. Testing the interrater reliability of a tool is an initial step in tool psychometrics, estimating how consistently different observers arrive at the same score [17]. According to Landis and Koch guidelines [14], the Perme ICU Mobility Score showed an excellent interrater agreement estimated from ICC of 0.98801, and the narrow confidence interval of 0.97743 to 0.99859 further supports interrater reliability for the overall score. The Bland-Altman plot demonstrates that the 2 observations are consistently within 2 points of each other, a 6% variation of the 32-point total score and less than 1 SD from the mean (SD, 6.7; Table 3).

The κ coefficient ranged from moderate ($\kappa = 0.6$) to perfect agreement ($\kappa = 1.0$). Of the 15 items in the Perme ICU Mobility Score, 8 had perfect κ coefficients [14]. An additional 6 items had substantial to almost perfect agreement and acceptable κ values. These items were static sitting balance (bed mobility category), sit to stand, and static standing balance (transfers category). The supine-to-sit item in the bed mobility category was the 1 item with a moderate agreement (80%) and κ value (0.6). The 3 items in the transfer category had substantial agreement with a κ value ranging from 0.72 to 0.78. The explanation of the lower agreement and κ values in both bed mobility and transfers categories reflects the difference between "hands on" and "hands off" approaches. Only 1 physical therapist (rater "A") provided direct hands-on assessment during the sessions. The other physical therapist (rater "B") observed the entire process and did not have the opportunity to physically touch the patient. As a result, the second rater was not able to accurately judge the degree of assistance required by the patient during vertical movements and may not have been consistently able to discern the difference between moderate assistance (2 points) and minimum assistance or supervision only (3 points). Nonetheless, a κ value of 0.6 is acceptable for early tool development [14].

Table 3
Intraclass correlation coefficient of interrater reliability

ICC	Asy. SE	95% confidence interval	
0.98801	0.00540	0.97743	0.99859
Estimated SD of patient effect			6.73258
Estimated SD within patient			0.7416198
Estimated interrater reliability of a patient mean (evaluated at n = 2.00)			0.99397

The high κ values for the items can also be associated with the relatively narrow selection of choices in scoring each item. According to Fricke et al [18], a 7-point scale tends to produce a greater variance than a 4-point scale. The scale used in the Perme ICU Mobility Score uses a maximum range of 2 to 4 points for each item. A small selection of 4 choices for the rater creates an increase in the κ statistics [19,20]. However, we justify the limited choices in scoring each item as important for clinicians—a smaller range in responses does not require extensive training before use and reduces the time needed to use the tool. We suggest that the limited choices are useful not only clinically but also capture key features of patient conditions that will provide trend data for researchers about whether mobility progresses, remains the same, or declines over time. Future testing of trend data is needed to validate this aspect of the Perme ICU Mobility Score.

The Perme ICU Mobility Score provides a total score of a patient's mobility status in a specific moment in time. The clinical importance of this ICU-specific tool is its focus on the physical performance for patients requiring intensive care. These patients can have multiple medical problems possibly needing sedation along with tubes and lines that can limit both independent and assisted mobility activities. It is important to note that the functional strength assessment is focused on the patient's ability to move their limbs against gravity. The inclusion of the endurance test measuring the distance walked in 2 minutes provides additional insight about the patient's potential for higher level of activities. Under item 15, if the patient is capable of walking but is unable to do so due to lack of resources or a multitude

Table 4
Item-by-item with total Perme ICU Mobility Score for both raters

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total score
Mean	1.85	1.05	0.925	0.65	0.175	0.5	1.825	2.05	2.625	2.975	2.6	2.7	1.325	1.85	1.375	21.475
SD	0.38	0	0.36	0.50	0.22	0.48	0.74	0.48	1.03	0.85	1.07	1.08	1.40	1.52	1.20	6.71
Median	2	1	1	1	0	0	2	2	3	3	3	3	0	1.5	0.5	22

of barriers, the rater should score it a zero. The authors do not know any other available tool that is specfic to the ICU setting that also measures patient's endurance status.

The FSS-ICU was described by Zanni et al [3] to be used as an outcome measurement in ICUs. The FSS-ICU demonstrated validity in that it was able to to discriminate among discharge settings and successive changes in function for patients in Long-term acute care hospital (LTACH) setting. However, the FSS-ICU has not been tested in acutely ill ICU patients as was the Perme Mobility Score [21]. The PFIT is an outcome measure used to prescribe and evaluate exercise and mobilization of patients in the ICU. Skinner et al [35] reported the results of the PFIT in 12 patients and demonstrated good interrater reliability. Two therapists concurrently observed 10 PFIT tests, which consisted of 2 components: marching on the spot and bilateral shoulder flexion. The PFIT test required participants to stand from a standardized chair. In contrast, the Perme ICU Mobility Score included in-bed as well as walking activities. Unlike the PFIT, the Perme ICU Mobility Score allowed for a measure of function earlier in the disease process when a patient may be unable to sit in a chair and also later when a patient may be able to walk. Another aspect to be considered is the fact that the PFIT is a tool for measurement of exercise capacity, as opposed to the Perme ICU Mobility Score, which was developed specifically to measure functional mobility status.

The Chelsea Critical Care Physical Assessment Tool (CPAx) is another tool to measure physical mobility in the general adult critical care population. The components of the CPAx are respiratory function, cough, moving within the bed, supine to sitting on the edge of the bed, dynamic sitting, sit to stand, standing balance, transferring from bed to chair, stepping, and grip strength [22]. In contrast, the Perme ICU Mobility Score does not require a dynamometer to test grip strength. In addition, the Perme ICU Mobility Score may be easier to use as it does not require additional equipment and respiratory assessment. In the United States, the physical therapist does not typically evaluate respiratory function or cough and, thus, may need additional training to use the CPAx or require input from the respiratory therapist and nursing staff.

One of the advantages of the Perme ICU Mobility Score is the relative ease of use because it takes less than 2 minutes to complete the entire scoring form. Measuring patient mobility status with the Perme ICU Mobility Score is a practical approach that reflects common and accepted parameters related to safe patient mobility in ICU. It can be quickly completed at the end of physical therapy interventions. This preliminary report suggests that the tool can be reliably scored following training.

The Perme ICU Mobility Score was developed to be used in any ICU. However, findings from this study cannot be generalized to other critically ill populations. The sample size in this study is limited to patients in CVICU at 1 institution. Further testing is needed to support reliability among a greater variety of ICU patients, settings, and tool users. Future studies using the Perme ICU Mobility Score can establish linkages with prognostic indicators (eg, determine if a low score is associated with low functional status preadmission), compare the scores in similar or different diagnosis or conditions (anticipating that a more severe diagnosis, APACHE score, or number of therapeutic interventions are associated with lower mobility at admission), and determine whether trend Perme ICU Mobility Scores predict patient recovery (eg, when the score improves by a certain factor, the patient is ready for the next step in progressive mobility or an upward trend score is associated with either a dependent or independent discharge location following ICU).

7. Conclusion

The Perme ICU Mobility Score demonstrated interrater reliability for the CVICU population, which is an important first step in this tool's development process. Measurement of mobility status has important implications for ICU clinical practice and research. Researchers will benefit from a specific tool that reflects mobility status of patients in the ICU rather than tools designed to assess mobility in other settings. A reliable and valid ICU-specific tool to assess mobility status during critical illness will help clinicians document progression, setbacks, and quality improvement measures in response to mobility initiatives.

Acknowledgments

The Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Houston Methodist Hospital Multidisciplinary Research Grant Program supported this study. The authors thank Jiaqiong Xu, PhD, for statistical analysis contribution. We also thank Catherine Currier-Buckingham and Sharon Cormier for administrative support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jcrc.2014.01.019.

References

[1] Hodgson CL, Berney S, Harrold M, et al. Clinical review: early patient mobilization in the ICU. Crit Care 2013;17:207–13.
[2] Barr J, Fraser GL, Puntillo K, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013;41:263–306.
[3] Zanni JM, Korupolu R, Fan E, et al. Rehabilitation therapy and outcomes in acute respiratory failure: an observational pilot project. J Crit Care 2010;25:254–62.
[4] Hanekom SD, Faure M, Coetzee A. Outcomes research in the ICU: an aid in defining the role of physiotherapy. Physiother Theory Pract 2007;23:125–35.
[5] Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J 1965;14:61–5.
[6] Keith RA, Granger CV, Hamilton BB, et al. The functional independence measure: a new tool for rehabilitation. Adv Clin Rehabil 1987;1:6–18.
[7] Skinner EH, Berney S, Warrillow S, Denehy L. Development of a physical function outcome measure (PFIT) and a pilot exercise training protocol for use in intensive care. Crit Care Resusc 2009;11:110–5.
[8] Sim J, Arnell P. Measurement validity in physical therapy research. Phys Ther 1993;73:102–10.
[9] Perme C, Nawa RK, Winkelman C, et al. The development of a new outcome tool to assess functional mobility in critically ill patients: The Perme Intensive Care Unit Mobility Score. Methodist Debakey Cardiovasc J. (Issue 10, Number 1. in press).
[10] Perme C, Chandrashekar R. Early mobility and walking program for patients in intensive care units: creating a standard of care. Am J Crit Care 2009;18:212–21.
[11] Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998;17:101–10.
[12] Flack VF, Afifi AA, Lachenbruch PA, et al. Sample size determinations for the two rater Kappa statistic. Psychometrika 1988;53:321–5.
[13] Bland JM, Altman DG. Measurement error and correlation coefficients. BMJ 1996;313:41–2.
[14] Harris KD, Heer DM, Roy TC, et al. Reliability of a measurement of neck flexor muscle endurance. Phys Ther 2005;85:1349–55.
[15] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[16] Kottner J, Gajewski BJ, Streiner DL. Guidelines for reporting reliability and agreement studies (GRRAS). Int J Nurs Stud 2011;48:659–60.
[17] Ko J, Kim M. Reliability and responsiveness of the Gross Motor Function Measure-88 in children with cerebral palsy. Phys Ther 2013;93:393–400.

[18] Fricke J, Unsworth C, Worrell D. Reliability of the functional independence measure with occupational therapists. Aust Occup Ther J 1993;40:7–15.
[19] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
[20] Brosseau L, Laferrère L, Couroux N, et al. Intra- and inter-rater reliability and factorial validity studies of the Physiotherapy Functional Mobility Profile (PFMP) in acute care patients. Physiother Theory Pract 1998;15:147–54.
[21] Thrush A, Rozek M, Dekerlegand JL. The clinical utility of the functional status score for the intensive care unit (FSS-ICU) at a long-term acute care hospital: a prospective cohort study. Phys Ther 2012;92:1536–45.
[22] Corner EJ, Wood H, Englebretsen C, Thomas A, Grant RL, Nikoletou D, Soni N. The Chelsea critical care physical assessment tool (CPAx): validation of an innovative new tool to measure physical morbidity in the general adult critical care population; an observational proof-of-concept pilot study. Physiotherapy 2013;99:33–41.