



Published in final edited form as:

Am J Crit Care. 2012 July ; 21(4): 261–269. doi:10.4037/ajcc2012111.

Patient-Nurse Interrater Reliability and Agreement of the Richards-Campbell Sleep Questionnaire

Biren B. Kamdar, MD, MBA, MHS, Pooja A. Shah, Lauren M. King, RN, MSN, Michelle E. Kho, PhD, PT, Xiaowei Zhou, ScM, Elizabeth Colantuoni, PhD, Nancy A. Collop, MD, and Dale M. Needham, MD, PhD

and **Dale M. Needham** are members of the Outcomes After Critical Illness and Surgery (OACIS) Group, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, Maryland. Needham is also an associate professor in the Department of Physical Medicine and Rehabilitation at Johns Hopkins University. **Pooja A. Shah** is an undergraduate student in the Department of Applied Mathematics and Statistics at Johns Hopkins University. **Lauren M. King** is a nurse in the medical intensive care unit at Johns Hopkins Hospital. **Xiaowei Zhou** and **Elizabeth Colantuoni** are bio-statisticians in the Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. **Nancy A. Collop** is a professor of medicine and director of the sleep center at Emory University in Atlanta, Georgia. **Michelle E. Kho** is a physical therapist in the Department of Physical Medicine and Rehabilitation at Johns Hopkins University.

Abstract

Background—The Richards-Campbell Sleep Questionnaire (RCSQ) is a simple, validated survey instrument for measuring sleep quality in intensive care patients. Although both patients and nurses can complete the RCSQ, interrater reliability and agreement have not been fully evaluated.

Objectives—To evaluate patient-nurse interrater reliability and agreement of the RCSQ in a medical intensive care unit.

Methods—The instrument included 5 RCSQ items plus a rating of nighttime noise, each scored by using a 100-mm visual analogue scale. The mean of the 5 RCSQ items comprised a total score. For 24 days, the night-shift nurses in the medical intensive care unit completed the RCSQ regarding their patients' overnight sleep quality. Upon awakening, all conscious, nondelirious patients completed the RCSQ. Neither nurses nor patients knew the others' ratings. Patient-nurse agreement was evaluated by using mean differences and Bland-Altman plots. Reliability was evaluated by using intraclass correlation coefficients.

Results—Thirty-three patients had a total of 92 paired patient-nurse assessments. For all RCSQ items, nurses' scores were higher (indicating “better” sleep) than patients' scores, with significantly higher ratings for sleep depth (mean [SD], 67 [21] vs 48 [35], $P = .001$), awakenings (68 [21] vs 60 [33], $P = .03$), and total score (68 [19] vs 57 [28], $P = .01$). The Bland-Altman plots also showed that nurses' ratings were generally higher than patients' ratings. Intraclass correlation coefficients of patient-nurse pairs ranged from 0.13 to 0.49 across the survey questions.

Conclusions—Patient-nurse interrater reliability on the RCSQ was “slight” to “moderate,” with nurses tending to overestimate patients' perceived sleep quality.

Sleep in the intensive care unit (ICU) is characterized by frequent arousals and loss of the restorative sleep stages necessary for healing.¹ Despite decades of research focusing on disrupted sleep in the ICU, only recently have outcomes related to ICU-associated sleep loss gained attention, mainly because of the potential association of such sleep loss with ICU delirium.²⁻⁴ Obtaining a reliable and valid measurement of sleep quality is an important barrier to ICU sleep research, because no method of evaluating sleep is currently acceptable and feasible for widespread use in the ICU.^{5,6} Although polysomnography is the reference standard and the most recognized measure of sleep, its requirement for cumbersome machinery, attentive staff, and expert interpretation render it costly and logistically challenging for large-scale ICU studies. Moreover, interpretation of polysomnograms is particularly difficult in critically ill patients because electroencephalographic patterns may be altered by common ICU medications and illnesses such as sepsis, shock, hepatic encephalopathy, and renal failure.^{7,8}

Compared with polysomnography, subjective survey instruments, such as the Richards-Campbell Sleep Questionnaire (RCSQ), have been investigated as potential practical instruments for ICU sleep measurement.⁹⁻¹⁴ Despite the feasibility and low cost of the RCSQ, concerns remain about the appropriateness of patients' self-reported sleep quality while they are sedated and/or delirious. Hence, researchers have investigated whether nurses' ratings are similar to patients' ratings. Two studies^{11,15} of patient-nurse interrater reliability/agreement had favorable results suggesting that RCSQ surveys could be completed by nurses in lieu of patients. However, results of 2 other studies^{5,12} demonstrated poor interrater agreement between patients and nurses. The generalizability and comparability of these existing reliability/agreement studies is limited by the heterogeneity of the populations of patients studied and the use of differing and less-established statistical methods. Hence, our objective was to rigorously evaluate patient-nurse RCSQ interrater reliability/agreement in a population of medical ICU (MICU) patients.

Methods

Setting and Sample

This interrater reliability/agreement evaluation was conducted in the MICU at Johns Hopkins Hospital from June 28 to July 21, 2010, as part of an ongoing larger sleep quality improvement project. The Johns Hopkins MICU is a closed ICU with 16 private rooms and a nurse to patient ratio of 1 to 2. A curtain and sliding glass door can be used to separate each patient's room from the ICU hallway. All rooms are uncarpeted, have a television, and have a small window behind the bed. During the larger quality improvement project, MICU-wide interventions were performed to improve patients' sleep, including environmental improvements (eg, dimming room and hallway lights, closing patients' doors/curtains, minimizing overhead pages, turning off televisions), offering of nonpharmacological sleep aids (eg, eye masks, earplugs, and soothing music), and promotion of a pharmacological guideline for use of sleep aids. Any patient 18 years old or older who was spending 1 full night (9_{PM} to 7_{AM}) in the MICU was eligible for inclusion in this evaluation. Exclusion criteria were as follows: positive delirium screening (using the Confusion Assessment Method for the ICU¹⁶) during the preceding night shift, inability to understand English, major communication barriers (eg, inability to write or point to answers), or moribund status.

RCSQ Instrument

The RCSQ was used to measure sleep quality for eligible MICU patients. Previously validated against polysomnography recordings in a MICU population,¹⁰ the RCSQ is a brief 5-item questionnaire used to evaluate perceived sleep depth, sleep latency (time to fall

asleep), number of awakenings, efficiency (percentage of time awake), and sleep quality (Table 1). Each RCSQ response was recorded on a 100-mm visual-analogue scale, with higher scores representing better sleep and the mean score of these 5 items, known as the “total score,” representing the overall perception of sleep. As done in prior studies using the RCSQ,¹¹ our questionnaire also included a sixth item evaluating perceived nighttime noise (range: 0 mm for “very quiet” to 100 mm for “very noisy”). MICU nurses received in-service educational sessions regarding completion of the RCSQ.

Data Collection

The MICU night-shift nurses were instructed to complete an RCSQ for every MICU patient approximately 30 minutes before completion of their 12-hour shift ending at 7 AM. This RCSQ assessment was based solely on each nurse's perception of their patient's overnight sleep quality. To maximize survey completion, a nurse leader provided a daily reminder to complete the RCSQ. Nursing assignments for patient care occurred independently of our reliability evaluation. To minimize bias, the night shift nurses were unaware that an assessment of interrater reliability/agreement was occurring.

Between 7 AM and 10 AM, an independent member of the MICU sleep project team (B.K., P.S., or L.K.) identified eligible MICU patients for patient-based completion of the RCSQ. Identified patients were approached and given the option of completing the RCSQ by themselves or with assistance from the project team member (eg, if unable to use a writing instrument). Neither the project team member nor the patient knew the nurse's RCSQ response. In addition to RCSQ data, a project team member recorded the patient's age, race, and sex. Patients' comorbid conditions before hospitalization were evaluated by using the Charlson Comorbidity Index¹⁷ and the Functional Comorbidity Index.¹⁸ Acuity of illness at ICU admission was measured by using the Sequential Organ Failure Assessment (SOFA) score.¹⁹ All project team members received training similar to that received by the MICU nurses regarding completion of the RCSQ.

Statistical Analysis

The 24-day duration of this evaluation of inter-rater reliability/agreement was selected to ensure enrollment of at least the minimum sample size of 21 unique patient-nurse RCSQ survey pairs. This sample size could be exceeded if enrollment were greater than expected during the designated period. Similar to prior reliability studies, this 21 patient sample size was calculated by using standard methods for reliability studies, with the intention of testing whether a reliability coefficient of 0.90 exceeded a reliability of 0.80, given 2 raters, a 1-tailed α of 0.05, and power of 80%.²⁰⁻²²

For patients staying in the MICU for 1 night or longer, we designated the first paired nurse-patient RCSQ assessment as the “initial” survey pair. “Repeated” pairs were designated as the initial pair plus any subsequent RCSQ survey pairs for the same patient. We defined parameters of interrater agreement as those measuring the “closeness” of patients' and nurses' questionnaire scores.^{23,24} Agreement was calculated by using the mean difference between patients' and nurses' paired scores and paired Student *t* tests. For patients with repeated sleep surveys, the *P* value for this mean difference measure of agreement was calculated on the basis of a robust variance estimate that accounted for the correlation of within-patient repeated measures.

Interrater agreement was also measured by using a modified Bland-Altman plot, which visually presented differences in paired responses from patients and nurses (*y* axis) in relation to patients' responses (*x* axis)²⁵ for each of the 6 questionnaire items and the total score.

Interrater reliability was defined as the ability to distinguish measurements conducted by the nurse and patient raters, based upon the statistical variability of patients' and nurses' sleep ratings.²⁴

Because sleep ratings were measured on a continuous scale, we calculated interrater reliability by using the intra-class correlation coefficient (ICC), which represents the proportion of total variation that can be explained by differences across patients. Larger ICCs indicate strong clustering of within-patient responses, indicating greater reliability of the nurses' and patients' reports. The ICCs were estimated from the results of random effects models, which included a random effect for patients for the initial pairs and random effects for both patients and nurses for the repeated pairs. Confidence intervals for the ICCs were constructed by using the bootstrap method (with 1000 repeated samples from the data). On the basis of previously published classifications, we qualitatively classified reliability as follows: slight, 0.0 to 0.20; fair, 0.21 to 0.40; moderate, 0.41 to 0.60; substantial, 0.61 to 0.80; and almost perfect, 0.81 to 1.00.²⁶

For statistical calculations, we used *R* version 2.10.1.0 (University of Auckland, New Zealand). Statistical significance was defined by a 2-sided *P* value less than .05. A description of this project was provided to the institutional review board chair at Johns Hopkins Hospital, who deemed it "quality improvement" within the context of the MICU's ongoing efforts to improve sleep quality and not requiring approval of the institutional review board in accordance with the Office for Human Research Protections standards.²⁷ Reporting of this evaluation was done in accordance with the Guidelines for Reporting Reliability and Agreement Studies.²⁸

Results

During a 24-day period, 41 MICU patients met the inclusion criteria for enrollment. Of these 41 patients, 8 were excluded: 1 did not speak English, 1 declined participation, and 6 did not have a nurse-completed RCSQ assessment. We enrolled the remaining 33 patients; the median (interquartile range [IQR]) age for patients was 52 (46-64) years, and 61% were female (Table 2).

The 33 patients had 137 patient-days with completion of 121 patient-based RCSQs and 101 nurse-based RCSQs, with a total of 92 paired patient-nurse assessments. Reasons for the 16 days in which patient-based questionnaires were not completed were as follows: 8 days when the patient was unable to communicate, 2 when the patient was not present in the room, 2 when the patient declined to participate, 1 when the patient had a change in clinical status (moribund), and 3 when the reason was unspecified. The primary reason for the 36 missing questionnaires from nurses was the heavy workload during the shift (eg, new admission or clinical instability of another patient near the end of the shift).

As reported by patients, the mean (SD) overall sleep quality (ie, the RCSQ total score) was 57 (30) on the 33 initial questionnaires and 57 (28) mm for the 92 repeated questionnaires, suggesting a tendency toward favorable (ie, score >50 on a 100-mm visual-analogue scale) sleep quality ratings in the MICU setting (Table 3). Patients' mean scores for each of the 5 individual RCSQ items ranged from 49 to 62 for the 33 initial measurements and 48 to 61 for the 92 repeated measurements, with the "sleep depth" domain having the lowest mean RCSQ score (ie, worst sleep characteristic) and "returning to sleep" having the highest mean RCSQ score (ie, best sleep characteristic). The mean score of 73 for perceived ICU noise was the highest (ie, high score = less noisy) among all 7 sleep survey scores.

Interrater agreement of the 33 initial patient-nurse survey pairings showed that the nurses' mean scores were higher than the patients' mean scores on all 5 RCSQ measures and the

total measure, but this difference was statistically significant only for the sleep depth measurement (Table 3). Very similar results were observed for the 92 repeated surveys. Given the larger sample size, the differences in mean scores for sleep depth, awakenings, and total score were statistically significant (after adjustment for the within-subject correlation in responses). Evaluation of the perceived noise measurement for the 33 initial surveys and 92 repeated paired RCSQ surveys demonstrated no significant difference in the mean difference between patients' and nurses' evaluations.

The modified Bland-Altman plot for all 7 sleep survey measures demonstrated that nurses consistently reported higher ratings than their patients reported. The magnitude of this overestimate by nurses was greater for lower sleep scores by patients (eg, 0-30) than for higher scores by patients (eg, 60-90), possibly because of a ceiling effect of this instrument (ie, maximum possible score was 100). The RCSQ total score measure was representative of this pattern of interrater agreement observed on all Bland-Altman plots (see Figure).

The ICC was slight to moderate for all 7 measures when the initial survey pairs (range, 0.14-0.49) were used and slight to fair when the repeated survey pairs (range, 0.13-0.34) were used (Table 3).

Discussion

We evaluated the patient-versus-nurse interrater reliability/agreement of perceived sleep quality ratings by using the RCSQ for 33 MICU patients who had a total of 92 paired patient-nurse assessments. With the exception of the sleep depth measure, the mean ICU sleep quality scores by patients for all RCSQ measures were balanced toward being favorable, with the highest mean score given for the evaluation of perceived ICU noise (ie, high score=less noisy). In evaluating the interrater reliability/agreement, we found that nurses' sleep quality ratings had slight to moderate reliability compared with patients' ratings, with nurses tending to overestimate their patients' sleep quality.

Our patient-reported RCSQ sleep scores were similar to scores reported in previous ICU studies

(Table 4). For example, the mean (SD) total RCSQ score for the 33 initial patient-completed surveys was 57 (30), consistent with 60 (27) in Richard's original validation study in 70 male ICU patients who were not receiving mechanical ventilation.¹⁰ When compared with a recent trial¹⁴ involving a noise- and light-reduction sleep-promoting intervention in a surgical ICU, each of our patient-reported mean scores (with the exception of sleep depth) fell between the preintervention and postintervention sleep quality scores of that trial. Last, our finding that patient-reported noise ratings tended to be higher (ie, less noisy) than corresponding nurse-reported noise scores was similar to the results of other studies that used the same nighttime noise question.^{11,14}

Our patient-nurse reliability and agreement numbers were similar to those reported in 2 recent reliability/agreement studies.^{5,12} The first study⁵ involved 91 paired assessments in 24 general ICU patients, 79% of whom were receiving mechanical ventilation. That study showed a wide range of differences in patients' and nurses' scores, suggesting poor agreement. The second study,¹² conducted in a surgical ICU, demonstrated substantial disagreement (57 of 101 one-time paired measures) between patients' and nurses' RCSQ scores, with nurses overestimating sleep ratings in 40 (70%) of the 57 discrepant cases. However, our results differed from the results of 2 older and smaller studies, which suggested high reliability/agreement between patients' and nurses' RCSQ ratings. These studies include Campbell's initial research,¹⁵ which showed "no significant difference" between 30 paired patient-nurse sleep ratings, and a study¹¹ of postoperative ICU patients

that showed a high correlation coefficient ($r = 0.869$, $P < .001$) between 13 paired RCSQ assessments.

Differences in methods may explain some differences in results between our project and prior published reports. For example, this evaluation occurred within a preexisting MICU sleep quality improvement project, which involved a nightly nursing checklist of interventions to promote patients' sleep. In view of the existing quality improvement project, it is plausible that nurses may have positively biased their perceived sleep ratings, especially when they were actively performing actions (eg, turning off television and lights) to promote patients' sleep at night. Moreover, the repeated daily completion of the RCSQ by nurses may have created scoring fatigue and reduced nurses' vigilance in completing the RCSQ, as might be suggested by missing nurse surveys and smaller variability (standard deviation) in nurses' scores compared with patients' scores.

In addition, by excluding only those patients who were cognitively or physically unable to complete the sleep questionnaire, we enrolled a heterogeneous sample of ICU patients that had numerous baseline comorbid conditions and a high burden of illness. Because previous studies enrolled patients of lower acuity (eg, no mechanical ventilation, postoperative, long-term ICU, and/or strict exclusion criteria), our patients were systematically sicker and therefore may have experienced worse sleep than was perceived by their nurses. Finally, because we adhered to recently established guidelines for reliability and agreement studies,²⁸ the evaluation and interpretation of our reliability/agreement estimates may have been more conservative than in prior studies, which often used subjective cutoffs or descriptive measures to report patient-nurse inter-rater reliability/agreement.

This project had potential limitations. First, conducting the project within a preexisting sleep quality improvement initiative may limit the generalizability of these survey ratings to other ICUs and may have affected the overall scores of patients and nurses, as previously discussed. However, our scores were similar to the scores reported in prior studies, which may limit this concern. Second, our project setting and population—a single MICU in a tertiary academic hospital with high severity of illness—may also limit the generalizability of the ratings and reliability/agreement results to different ICU settings and populations of patients. However, as noted in Table 4, our patient sleep questionnaire ratings were comparable to those collected in the presence of both control (ie, adlib ICU setting) and experimental (ie, sleep-promoting interventions) settings in various types of ICUs internationally. Third, our results may have been limited by nurse surveys that were missing because of the competing demands of other nursing care requirements. However, this rating of missed assessments was comparable to the only rate (25%) reported in another ICU-based RCSQ reliability study.⁵ Last, the confidence intervals for our reliability estimates were wider than anticipated because our sample size calculation was based on a minimum reliability of 0.8 and an observed reliability of 0.9, estimates that were far higher than our actual reliability calculations. However, by adhering to the sample size reporting guidelines, our analysis provides meaningful data for calculating sample sizes for future research.

In conclusion, patient-nurse interrater reliability/agreement of the RCSQ-based sleep questionnaire in a MICU setting is slight to moderate, with nurses tending to overestimate sleep quality compared with their patients. Based on these findings, studies that use RCSQ-based measures of ICU patients' sleep quality must consider potential proxy-related biases that may affect results. Future similar investigations of patient-nurse reliability of the RCSQ in the ICU should evaluate the effect of different ICU settings and patients' characteristics to evaluate further any potential proxy-related bias.

Acknowledgments

We thank all the MICU nursing and other staff at Johns Hopkins Hospital who participated in this project, Dr Rahul Ravilla for assistance with data collection, and Mr Peter Shaw for assistance with data entry.

FINANCIAL DISCLOSURES

Dr Kamdar is recipient of a Ruth L. Kirschstein NRSA award from the National Institutes of Health (F32 HL104901). Dr Kho is funded by a Canadian Institutes of Health Research Fellowship Award and Bisby Prize.

REFERENCES

1. Kamdar BB, Needham DM, Collop NA. Sleep deprivation in critical illness: its role in physical and psychological recovery. *J Intensive Care Med.* 2012; 27(2):97–111. [PubMed: 21220271]
2. Mistraletti G, Carloni E, Cigada M, et al. Sleep and delirium in the intensive care unit. *Minerva Anesthesiol.* 2008; 74(6):329–333. [PubMed: 18500209]
3. Trompeo, AC.; Vidi, Y.; Ranieri, VM. Sleep and delirium in the critically ill: cause or effect?. In: Vincent, JL., editor. *Yearbook of Intensive Care and Emergency Medicine.* Springer; New York, NY: 2006. p. 719-725.
4. Figueroa-Ramos MI, Arroyo-Novoa CM, Lee KA, Padilla G, Puntillo KA. Sleep and delirium in ICU patients: a review of mechanisms and manifestations. *Intensive Care Med.* 2009; 35(5):781–795. [PubMed: 19165463]
5. Bourne RS, Minelli C, Mills GH, Kandler R. Clinical review—sleep measurement in critical care patients: research and clinical implications. *Crit Care.* 2007; 11(4):226–242. [PubMed: 17764582]
6. Watson PL. Measuring sleep in critically ill patients: beware the pitfalls. *Crit Care.* 2007; 11(1466):159. [PubMed: 17850679]
7. Blume WT. Drug effects on EEG. *J Clin Neurophysiol.* 2006; 23(736):306–311. [PubMed: 16885705]
8. Kaplan PW. The EEG in metabolic encephalopathy and coma. *J Clin Neurophysiol.* 2004; 21(5):307–318. [PubMed: 15592005]
9. Richards K. Techniques for measurement of sleep in critical care. *Focus Crit Care.* 1987; 14(4):34–40. [PubMed: 3650169]
10. Richards KC, O'Sullivan PS, Phillips RL. Measurement of sleep in critically ill patients. *J Nurs Meas.* 2000; 8(2):131–144. [PubMed: 11227580]
11. Frisk U, Nordstrom G. Patients' sleep in an intensive care unit: patients' and nurses' perception. *Intensive Crit Care Nurs.* 2003; 19(6):342–349. [PubMed: 14637294]
12. Nicolás A, Aizpitarte E, Iruarizaga A, Vázquez M, Margall A, Asiain C. Perception of night-time sleep by surgical patients in an intensive care unit. *Nurs Crit Care.* 2008; 13(1):25–33. [PubMed: 18226052]
13. Williamson JW. The effects of ocean sounds on sleep after coronary artery bypass graft surgery. *Am J Crit Care.* 1992; 1(1):91–97. [PubMed: 1307884]
14. Li SY, Wang TJ, Vivienne Wu SF, Liang SY, Tung HH. Efficacy of controlling night-time noise and activities to improve patients' sleep quality in a surgical intensive care unit. *J Clin Nurs.* 2011; 20(3-4):396–407. [PubMed: 21219521]
15. Campbell, C. A Comparison of Patients' and Nurses' Perception of Patient Sleep in the Intensive Care Unit [master's thesis]. Northwestern State University; Natchitoches, LA: 1986.
16. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *JAMA.* 2001; 286(21):2703–2710. [PubMed: 11730446]
17. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987; 40(5):373–383. [PubMed: 3558716]
18. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. *J Clin Epidemiol.* 2005; 58(6):595–602. [PubMed: 15878473]

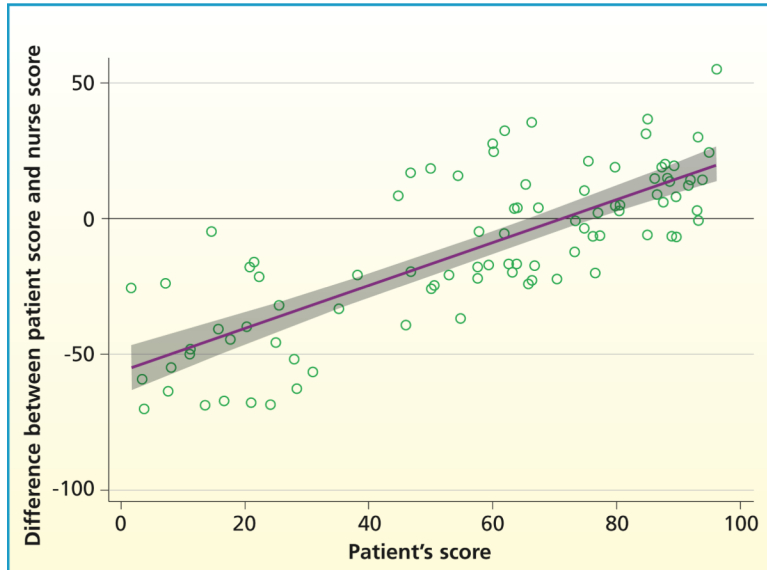
19. Vincent JL, Moreno R, Takala J, et al. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dys-function/failure. *Intensive Care Med.* 1996; 22(7):707–710. [PubMed: 8844239]
20. Stratford PW, Spadoni GF. Sample size estimation for the comparison of competing measures' reliability coefficients. *Physiother Can.* 2003; 55(4):225–229.
21. Donahoe L, McDonald E, Kho ME, MacLennan M, Stratford PW, Cook DJ. Increasing reliability of APACHE II scores in a medical-surgical intensive care unit: a quality improvement study. *Am J Crit Care.* 2009; 18(1):58–64. [PubMed: 19116406]
22. Kho ME, McDonald E, Stratford PW, Cook DJ. Interrater reliability of APACHE II scores for medical-surgical intensive care patients: a prospective blinded study. *Am J Crit Care.* 2007; 16(4): 378–383. [PubMed: 17595370]
23. Costa-Santos C, Bernardes J, Ayres-de-Campos D, Costa A, Costa C. The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. *J Clin Epidemiol.* 2011; 64(3):264–269. [PubMed: 20189765]
24. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006; 59(10):1033–1039. [PubMed: 16980142]
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1(8476):307–310. [PubMed: 2868172]
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1):159–174. [PubMed: 843571]
27. Health & Human Services Office of Human Research Protections. [January 31, 2011] Quality Improvement Activities: Frequently Asked Questions website. <http://answers.hhs.gov/ohrp/categories/1569>.
28. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011; 64(1):96–106. [PubMed: 21130355]

Night nurses completed the RCSQ tool based on their perceptions of the sleep quality for every patient.

Patient-nurse interrater reliability was slight to moderate for all sleep measures.

Nurses tend to overestimate their patients' sleep quality.

Similar to other reports, patients report less noise than do nurses.

**Figure.**

Bland-Altman plot of total scores of 92 Richards-Campbell Sleep Questionnaires shows the relationship between patient score (x axis) and the difference between the patient's score and the nurse's score (y axis). Each circle represents 1 patient-nurse paired questionnaire. The solid diagonal line is a fitted representation of the simple linear regression of patient-nurse score difference as a function of patients' scores, with the gray shaded area representing the 95% confidence interval.

Table 1

Sleep questionnaire

Measure	Question ^a
1. Sleep depth	My sleep last night was: light sleep (0) ... deep sleep (100)
2. Sleep latency	Last night, the first time I got to sleep, I: just never could fall asleep (0) ... fell asleep almost immediately (100)
3. Awakenings	Last night, I was: awake all night long (0) ... awake very little (100)
4. Returning to sleep	Last night, when I woke up or was awakened, I: couldn't get back to sleep (0) ... got back to sleep immediately (100)
5. Sleep quality	I would describe my sleep last night as: a bad night's sleep (0) ... a good night's sleep (100)
6. Noise ^b	I would describe the noise level last night as: very noisy (0) ... very quiet (100)

^aEach question is scored by using a 100-mm visual analog scale in which a higher score is better.

^bQuestion 6 is not a part of the original 5-item Richards-Campbell Sleep Questionnaire (RCSQ), but was included in this project for consistency with other studies that used the RCSQ.¹¹

Table 2

Characteristics of the 33 patients at enrollment

Characteristic	Value
Age, median (IQR), y	52 (46-64)
Female sex, No. (%)	20 (61)
Race, No. (%)	
White	15 (45.5)
Black	15 (45.5)
Other	3 (9)
Charlson Comorbidity Index score, ^a median (IQR)	2.0 (1.0-4.0)
Functional Comorbidity Index score, ^b median (IQR)	2.0 (1.0-3.0)
SOFA acuity of illness score at MICU admission, median (IQR)	6.0 (4.0-9.0)
MICU admission diagnosis category, No. (%)	
Respiratory failure ^c	13 (39)
Gastrointestinal	7 (21)
Sepsis (nonpulmonary origin)	3 (9)
Renal	3 (9)
Cardiovascular	2 (6)
Monitoring/procedure	3 (9)
Other	2 (6)
Received mechanical ventilation during project, No. (%)	10 (33)

Abbreviations: IQR, interquartile range; MICU, medical intensive care unit; SOFA, Sequential Organ Failure Assessment.

^aCharlson Comorbidity Index: a weighted score derived from 19 categories of comorbid conditions, with higher scores reflecting a greater burden of comorbid disease and risk of death.

^bFunctional Comorbidity Index: a score derived from 18 categories of comorbid conditions, with higher scores reflecting a greater burden of comorbid disease and risk of impaired physical function.

^cIncludes pneumonia, acute respiratory distress syndrome, chronic obstructive pulmonary disease, asthma, and pulmonary embolism.

Table 3

Sleep questionnaire: patients' versus nurses' estimates

Sleep questionnaire measure	Estimate, mean (SD)					P ^b	ICC (95% CI) ^c
	Survey ^a	No.	Patient	Nurse	Difference (95% CI)		
1. Sleep depth	Initial	33	49 (38)	63 (20)	-15 (-28 to -2)	.02	0.24 (0.00-0.52)
	Total	92	48 (35)	67 (21)	-19 (-26 to -11)	.001	0.34 (0.15-0.49)
2. Sleep latency	Initial	33	60 (39)	66 (21)	-6 (-20 to 8)	.39	0.21 (0.00-0.47)
	Total	92	60 (36)	68 (21)	-9 (-16 to -1)	.11	0.26 (0.13-0.39)
3. Awakenings	Initial	33	57 (35)	67 (22)	-10 (-22 to 3)	.14	0.22 (0.00-0.49)
	Total	92	60 (33)	68 (21)	-9 (-15 to -2)	.03	0.21 (0.07-0.34)
4. Returning to sleep	Initial	33	62 (35)	71 (20)	-9 (-22 to 4)	.18	0.14 (0.00-0.40)
	Total	92	61 (34)	71 (20)	-10 (-17 to -2)	.08	0.29 (0.15-0.44)
5. Sleep quality	Initial	33	58 (35)	64 (21)	-5 (-19 to 8)	.42	0.17 (0.00-0.45)
	Total	92	59 (33)	67 (21)	-8 (-15 to -1)	.08	0.16 (0.01-0.28)
Total score (average of 1-5)	Initial	33	57 (30)	66 (19)	-9 (-20 to 2)	.10	0.49 (0.26-0.67)
	Total	92	57 (28)	68 (19)	-11 (-17 to -5)	.01	0.28 (0.17-0.39)
6. Noise	Initial	32	73 (30)	68 (18)	5 (-7 to 17)	.39	0.25 (0.00-0.42)
	Total	88	71 (30)	71 (18)	1 (-7 to 8)	.93	0.13 (0.01-0.28)

^aInitial is the first matched patient-proxy sleep questionnaire for each patient (33 paired surveys), whereas "Total" represents 92 matched patient-proxy questionnaires, including any repeated sleep questionnaires performed daily with each patient.

^bP for difference in "Initial" mean scores calculated by using paired t test and for difference in "Total" mean scores calculated by using a clustered regression analysis to adjust for repeat measures.

^cConfidence intervals for the intraclass correlation coefficients (ICCs) were generated by using the bootstrap method with K = 1000 repeated samples.

Table 4
Comparison of studies that used the Richards-Campbell Sleep Questionnaire (RCSQ)

Characteristic	Present project		Richards et al ^{10,d}		Frisk and Nordstrom ^{11,b}		Nicolás et al ^{12,c}		Williamson ^{13,d}		Li et al ^{14,e}	
	MICU	General ICU	General ICU	SICU	SICU	SICU	CSICU	CSICU	CSICU	CSICU	Control	Exp
Setting												
No. of patients' surveys analyzed	33	70	70	31			104	30	30	30	27	28
Patients' sleep ratings, f mean (SD)												
1. Sleep depth	49 (38)	44 (34)	44 (34)	40	51 (26)	51 (26)	51	56	51	51 (28)	51 (28)	64 (22)
2. Sleep latency	60 (39)	66 (30)	66 (30)	48	56 (27)	56 (27)	61	71	61	54 (29)	54 (29)	65 (23)
3. Awakenings	57 (35)	66 (29)	66 (29)	53	42 (24)	42 (24)	60	65	60	51 (26)	51 (26)	65 (16)
4. Returning to sleep	62 (35)	62 (31)	62 (31)	47	56 (26)	56 (26)	56	68	56	54 (30)	54 (30)	66 (20)
5. Sleep quality	58 (35)	64 (34)	64 (34)	39	53 (28)	53 (28)	57	69	57	51 (26)	51 (26)	65 (20)
Total score (average of items 1-5)	57 (30)	60 (27)	60 (27)	46	51 (22)	51 (22)	57	66	57	52 (26)	52 (26)	65 (19)
6. Noise (if applicable)	73 (30)	NA	NA	67	NA	NA	NA	NA	NA	63 (23)	63 (23)	74 (14)

Abbreviations: CSICU, cardiac surgery intensive care unit; Exp, experimental group; ICU, intensive care unit; MICU, medical intensive care unit; NA, not applicable; SICU, surgical intensive care unit.

^aMale ICU patients, 55-79 years old, in stable condition, no mechanical ventilation, in 8-bed US ICU for >48 hours.

^bConscious and alert patients 19-85 years old spending at least 48 hours in 6-bed ICU in Sweden.

^cPatients not receiving mechanical ventilation during a 5-month period in a 16-bed ICU in Spain. RCSQ collected during 2 consecutive nights as part of an initiative to minimize nighttime noise and light in the ICU.

^dPatients in US CSICU after coronary artery bypass graft surgery, stratified into 2 groups: experimental group (played "white noise" ocean sounds each night to promote sleep) and control group (no ocean sounds). RCSQ performed once on the fourth postoperative night.

^ePostoperative patients (2 of 55 were receiving mechanical ventilation) in SICU in Taiwan. Study consisted of 2 phases: 3-month control (ad-lib ICU setting) followed by 3-month experimental period with environmental strategies to promote sleep. RCSQ performed 1 time on 2nd postoperative night.

^fHigher sleep survey scores indicate better perceived sleep quality and less perceived noise in the ICU.