Manuscript title

Cross-cultural adaptation and inter-rater reliability of the Swedish version of the Chelsea Critical Care Assessment Tool (CPAX-Swe) in critically ill patients

Running title

Adaptation and inter-rater reliability of CPAx-Swe

Article category

Assessment procedures

Abstract

Objectives: To translate and culturally adapt the Chelsea Critical Care Physical Assessment Tool into Swedish and to test the interrater reliability of the Swedish version in critically ill patients.

Design: Observational study

Methods: Translation and cross-cultural adaptation was performed in line with international recommendations, including forward and backward translation and expert round table discussions. The inter-rater reliability of the Chelsea Critical Care Physical Assessment Tool - Swedish was then explored in 50 critically ill adult patients, pragmatically recruited, in a University Hospital clinical setting.

Reliability was calculated using intraclass correlation coefficient for aggregated scores and quadratic weighted Cohen's kappa analysis for individual items

Results: The expert round table discussion group agreed that the translation was a satisfactory equivalent to the original version and applicable for use within the clinical setting. Reliability of aggregated scores and individual items were very good (intraclass correlation coefficient of 0.97 and quadric weighted kappa values ranging from 0.88 to 0.98). The measurement error for aggregated scores was low, with a standard error of measurement of 1.79, smallest detectable change of 4.95, and limits of agreement of 5.20 and -4.76. The percentage agreement for individual items ranged from 64% to 88%.

Conclusion: The Chelsea Critical Care Physical Assessment Tool - Swedish was found applicable and appropriate for assessment of functioning in critically ill patients in an acute

setting in Sweden, and it displayed high inter-rater reliability. This implies that the Swedish version can be used as assessment tool within intensive care and acute wards in Sweden.

Contribution of the paper:

Key messages:

- CPAx-Swe is considered reliable for use by physiotherapist in intensive care settings in Sweden.
- The CPAx-Swe is feasible for use within clinical practice thanks to its simplicity and strong clinical relevance.

New knowledge:

- The CPAx-Swe is the first translation of a suitable outcome measure for the use within critically ill patients in Sweden.
- The results of this study confirm the findings of earlier research about the reliability of CPAx.

Keywords: critical care, psychometrics, physiotherapy, rehabilitation, outcome measure

Introduction

In Sweden, as in most developed countries, physiotherapists are part of the multiprofessional intensive care unit team. Physiotherapy interventions in adult critically ill patients include multimodality respiratory physiotherapy, early progressive mobilization and physical activity to prevent and treat respiratory conditions, physical deconditioning, and neuromuscular and musculoskeletal complications [1-4]. To evaluate the effect of such interventions, outcome measures with adequate psychometric properties are needed.

Several outcome measures have been developed for use within the intensive care setting, e.g. the Chelsea Critical Care Physical Assessment Tool (CPAx) [5] the Physical Function in Intensive Care Test [6], the Functional Status Score for the Intensive Care [7] and the Perme mobility scale [8]. According to a systematic review of 26 different outcome measures [9], the CPAx and the Physical Function in Intensive Care Test demonstrated the strongest psychometric properties, however the latter has a significant floor effect.

The CPAx was developed as a bedside assessment tool for the critical care population and has demonstrated validity, reliability and responsiveness [5,10,11]. Ten items (respiratory function, cough, moving within bed, supine to sitting on the edge of bed, dynamic sitting, standing balance, sit to stand, transferring from bed to chair, stepping and grip strength) are rated on a 6-point Guttman-Scale from complete dependency (0) to independency (5). An aggregated score can be calculated (0-50) and higher scores indicate a better functioning/independency. An eLearning package (https://cpax.helmlms.com/login) is available and has proved to be an effective and useful way to deliver standardised education and facilitate clinical implication [12]. Before using the CPAx in a Swedish context, a cross-cultural adaptation and reliability testing after translation was needed. Thus, the objectives of this study were to translate and culturally adapt the CPAx into Swedish and to test the interrater reliability of the Swedish version (CPAx-Swe) in critically ill patients.

Methods

The translation and cross-cultural adaptation were performed with reference to published guidelines [13,14]. After permission from the original author, the CPAx was translated to Swedish by an independent professional translator (nationality English) fluent in Swedish and English. The Swedish translation was then discussed by an expert committee in a round table discussion consisting of five experienced physiotherapists working within intensive care unit, who made several cultural adaptations. The adjusted version was translated back to English by an independent bilingual (nationality English) physiotherapist. The back translation was approved by the original author of the CPAx. A pilot test was completed using two examiners working at the intensive care unit but not involved in the translation process. This was to check that CPAx-Swe was appropriate and applicable for use in acute Swedish healthcare. No adjustments were deemed necessary after the pilot test.

Published guidelines for reporting reliability studies [15] were used for the inter-rater reliability part of the study, and the study was approved by the ethical review board in Stockholm, registration number: 2017/679-31/4.

Patients

Adult patients were pragmatically recruited from the general and the cardiothoracic intensive care units, and the acute/high dependency wards at the Karolinska University Hospital, Stockholm, Sweden, from November 2016 to April 2018. There were no specific exclusion criteria to allow CPAx-Swe to be tested for potential ceiling- and floor effects. A sample size of 50 patients was considered adequate for inter-rater reliability testing [16].

Raters

Twelve physiotherapists aged 23 - 60 with clinical experience ranging from 7 months to 18 years, working at the Karolinska University Hospital during the study-period, participated as raters. All completed the eLearning package (https://cpax.helmlms.com/login) before data collection. The eLearning package was completed in English.

Procedures

Assessments were completed in pairs. Two physiotherapists assessed the same patient at the same time. One physiotherapist led the assessment while the other observed or assisted if needed. Each physiotherapist scored the patient independently of each other without discussion and was blinded to the score allocated. The scores were recorded in separate CPAx-Swe forms. Also noted were the patient's age, sex and diagnosis as well as who was the lead physiotherapist. This process took approximately 30 minutes.

Statistical analysis

Descriptive statistics were used to present data, i.e. mean, standard deviation (sd), median, interquartile range (IQR), frequency and percentage. Analyses of the measurement properties reliability and measurement error [17] were used to evaluate inter-rater reliability of the CPAx-Swe. Reliability analyses consisted of intraclass correlation coefficient (ICC) analysis as described below for aggregated scores and quadratic weighted Cohen's kappa analysis for individual items [18]. Reliability was considered very good if $ICC_{agreement}$ was >0.80 and if weighted kappa values were >0.75 [18]. The standard error of measurement (SEM), smallest detectable change (SDC) and limits of agreement, as described below, were calculated as a parameters of measurement error for aggregated scores [18]. Percentage agreement was used as a parameter for measurement error for individual items [18]. The CPAx-Swe aggregated score data were visualised in Bland–Altman plots to check for systematic bias, outliers or heteroscedasticity, i.e. whether the differences depend on the magnitude of the mean (Figure 1).

$ICC_{agreement} = \sigma^2_p / \sigma^2_p + \sigma^2_o + \sigma^2_{residual}$

σ^2_p = variance due to systematic difference between 'true' scores of patients

σ^2_o = variance due to systematic difference between raters

$\sigma^2_{residual}$ = residual variance

$SEM_{agreement} = \sqrt{(\sigma^2_o + \sigma^2_{residual})}$

$SDC = \sqrt{2} \times 1.96 \times SEM$

Limits of agreement were defined as $\overline{d} \pm 1.96 \times SD_{diff}$ where $\overline{d}$ = mean difference between raters and SD_{diff} = the standard deviation of the differences.

All analyses were performed using SPSS software (version 24).

Results

Translation and cross-cultural adaptation

The Swedish version of the CPAx was found to be a satisfactory equivalent to the original version and was accepted by the original author. The pilot test showed that the instrument was applicable for clinical setting. Minor adjustments in wording were made during the translation process due to linguistic and terminological differences, for instance one piece of equipment (yanker suction) is not used in Swedish hospitals and was therefore alternative wording was required. See Table 1 for an example of wording adjustments for item "cough". The expert committee identified some unclarities, but after contacting the original author these were solved and consensus was reached on the translated version, i.e. the CPAx-Swe (see Appendix).

Inter-rater reliability

Fifty patients (15 women, 35 men) with a mean (sd) age 56.8 (18.9) ranging from 18 to 88 years participated. The majority (n=39, 78%) were recruited from the general intensive care unit, while four (8%) and seven (14%) patients were from the cardiothoracic intensive care unit and the acute/high dependency wards, respectively. The diagnoses were as follows: medicine (n=10), trauma (n=10), surgery (n=9), respiratory (n=8), cardiothoracic (n=8) and haematology (n=4).

Descriptive statistics of assessments and results from interrater reliability tests are presented in Tables 2 and 3. Reliability of aggregated scores and individual items was very good, $ICC_{agreement}$ = 0.97 and quadratic weighted kappa ranging from 0.88 to 0.98, respectively. The measurement error for aggregated scores was low (SEM=1.79 and SDC=4.95). The limits of agreements are displayed in Figure 1. The mean difference between raters was 0.22 and the standard deviation of the differences was 2.54, thus the 95% limits of agreement were 5.20 and -4.76). The percentage agreement, as a parameter for measurement error, for individual items ranged from 64% to 88%.

Discussion

The aim of the present study was to translate and culturally adapt the CPAx into Swedish and then test the inter-rater reliability of the CPAx-Swe. After expert round table discussions, the CPAx-Swe was found to be a satisfactory equivalent to the original version. The inter-rater reliability was considered very good with high reliability and small measurement errors.

The translation and adaption process followed published guidelines. Some deviations from the guidelines were necessary to the translation process due to financial constraints. In this study only one translator was used for the forward translation, and one for the back translation resulting in the potential for bias [13,14]. The compliance to this standardized process vouches for the CPAx-Swe being highly comparable with the original version. In addition, the use of an expert round table discussion and prior pilot testing in the clinical setting to check for understanding and acceptability provide further credibility. A strength was the involvement of the original author (Evelyn Corner) throughout the process. As was the case in the translation and adaptation of the CPAx into Danish [19].

Inter-rater reliability was assessed by analyses of the measurement properties reliability and measurement error as proposed by the international initiative COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) [17,18]. Reliability is according to COSMIN defined as “the proportion of the total variance in the measurements which is because of ‘true’ differences among patients”, and measurement error is defined as “the systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured” [17]. Thus, reliability gives information on how well patients can

be distinguished from each other, and measurement error on how close scores are for repeated measurements, i.e. the agreement between measurements.

The reliability results in the present study were considered very good and correspond to earlier studies by Corner et al [5,12]. The clinical implication is that the CPAx-Swe seems to be a useful assessment tool for evaluation of function of the intensive care unit-patient. The measurement error of the for aggregated scores was considered low, implying that the CPAx-Swe can be used for evaluative purposes. The SDC was 5 points and the limits of agreements were approximately ± 5 points. Thus, if a patient is assessed by different raters, a change of at least 5 points is indicative of an improvement or deterioration. The measurement errors for individual items ranged from 64 to 88 percentage agreement. Although agreement between raters was generally high, it was somewhat lower on the items cough, moving within the bed and stepping. Disagreement in ratings on the cough item may be due to subjectivity in what is considered self-clearing and whether suction is considered high or low in the throat. That some CPAx items tend to yield lower agreement between raters due to subjective interpretation of e.g. “minimal” versus “moderate” assistance has also previously been highlighted [12].

Limitations of this study include that this was a single-center study. To compensate for this, different ward environments were used, i.e. the general and the cardiothoracic intensive care units, and the acute/high dependency wards. Further, only physiotherapists served as raters and the results can therefore not be generalised to other health professionals. That inter-rater reliability of the CPAx-Swe was chosen to study instead of intra-rater/test-retest reliability, was due to the difficulties in performing the latter studies on patients in intensive care since their conditions can vary considerably even over a short period of time. Another possible limitation was the procedure of having assessments completed in pairs where the leading assessor and the constellation of the pairs shifted on each occasion. This made it difficult to discover any systematic between-raters error. The selected procedure is, however, more like the clinical reality of how physiotherapists work in an intensive care setting. Furthermore, assessing patients at different time points is problematic due to patient fatigue and the potential for rapid developments in their clinical picture.

That all raters completed the CPAx eLearning package ensured that they had the same background information about the assessment tool before commencing the study, and

enhanced a common understanding for using the CPAx-Swe. No issues were raised regarding understanding the eLearning package as all participating raters were considered fluent in both English and Swedish. The inter-rater reliability was found to be very good regardless of the variation in age and clinical experience among the raters. This suggests that the CPAx-Swe can be used by other physiotherapists working in acute wards with critically ill patients. The assessment tool can be considered easy to use in the clinical setting due to the minimal use of equipment and the short time required for assessment.

In conclusion, the CPAx-Swe was found applicable and appropriate for assessment of functioning in critically ill patients in an acute setting in Sweden, and it displayed high inter-rater reliability. This implies that CPAx-Swe can be used as assessment tool within intensive care and acute wards in Sweden.

Ethical approval

The study was approved by the ethical review board in Stockholm, registration number: 2017/679-31/4.

Acknowledgements

The authors thank Tim Crosfield and Suzanne Torén for help with the translation process.

References

1. Gosselink R, Bott J, Johnson M, Dean E, Nava S, Norrenberg M, et al. Physiotherapy for adult patients with critical illness: recommendations of the European Respiratory Society and European Society of Intensive Care Medicine Task Force on Physiotherapy for Critically Ill Patients. Intensive Care Med. 2008;34(7):1188-99.
2. Stiller K. Physiotherapy in intensive care: an updated systematic review. Chest. 2013;144(3):825-47.
3. Sommers J, Engelbert RH, Dettling-Ihnenfeldt D, Gosselink R, Spronk PE, Nollet F, et al. Physiotherapy in the intensive care unit: an evidence-based, expert driven, practical statement and rehabilitation recommendations. Clin Rehabil. 2015;29(11):1051-63.
4. Arias-Fernandez P, Romero-Martin M, Gomez-Salgado J, Fernandez-Garcia D. Rehabilitation and early mobilization in the critical patient: systematic review. J Phys Ther Sci. 2018;30(9):1193-201.
5. Corner EJ, Wood H, Englebretsen C, Thomas A, Grant RL, Nikoletou D, et al. 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(1):33-41.

6. 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(2):110-5.
7. 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(12):1536-45.
8. Nawa RK, Lettvin C, Winkelman C, Evora PR, Perme C. Initial interrater reliability for a novel measure of patient mobility in a cardiovascular intensive care unit. J Crit Care. 2014;29(3):475.e1-5.
9. Parry SM, Granger CL, Berney S, Jones J, Beach L, El-Ansary D, et al. Assessment of impairment and activity limitations in the critically ill: a systematic review of measurement instruments and their clinimetric properties. Intensive Care Med. 2015;41(5):744-62.
10. Corner EJ, Soni N, Handy JM, Brett SJ. Construct validity of the Chelsea critical care physical assessment tool: an observational study of recovery from critical illness. Crit Care. 2014;18(2):R55.
11. Corner EJ, Hichens LV, Attrill KM, Vizcaychipi MP, Brett SJ, Handy JM. The responsiveness of the Chelsea Critical Care Physical Assessment tool in measuring functional recovery in the burns critical care population: an observational study. Burns. 2015;41(2):241-7.
12. Corner EJ, Handy JM, Brett SJ. eLearning to facilitate the education and implementation of the Chelsea Critical Care Physical Assessment: a novel measure of function in critical illness. BMJ Open. 2016;6(4):e010614.
13. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186-91.
14. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94-104.
15. Kottner J, Streiner DL. The difference between reliability and agreement. J Clin Epidemiol. 64. United States2011. p. 701-2; author reply 2.
16. Altman D. Practical Statistics for Medical Research. London: Taylor & Francis Ltd; 1991. Page 402.
17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
18. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide: Cambridge University Press; 2011.
19. Astrup K, Corner EJ, Hansen MG, Petersen AK. Translation and cross-cultural adaptation of the Chelsea Critical Care Physical Assessment tool into Danish. Physiother Theory Pract. 2018:1-8.

Figure 1. A Bland-Altman plot of data from the assessments of the 50 critically ill patients by the pair of raters. The difference between the lead and observer physiotherapist's aggregated scores of the Swedish Chelsea Critical Care Physical Assessment Tool (CPAx-Swe) plotted against the mean of the raters' aggregated scores. The mean difference is marked with a solid line and the 95% limits of agreements with dashed lines.

Table 1
Adjustment in wording, terminology and grammar during the translation and expert round table discussion for item “Cough”

	Original version of CPAx	**The back-translated synthesized English version**	**The final version of CPAx (presented as the synthesized English version after the expert round table discussion)**
2. Cough			
Level 0	Absent cough, maybe fully sedated or paralysed	Absent cough. Maybe fully sedated/muscle relaxed	Absent cough. Maybe fully sedated/muscle relaxed
Level 1	Cough stimulated on deep suctioning only	Cough is stimulated in connection to deep suctioning	Cough is stimulated only on deep suctioning
Level 2	Weak ineffective voluntary cough unable to clear independently (e.g requires deep suction)	Weak ineffective cough. The patient can´t clear airways independently. Need deeper suctioning	Spontaneous, weak cough, cannot clear mucus independently, needs deep suctioning.
Level 3	Weak, partially effective cough, sometimes able to clear secretions (e.g requires Yankauer suctioning)	Weak mostly ineffective cough. Can sometimes clear secretions. Requires suctioning in the mouth.	Weak, partially effective cough. Can sometimes clear secretions (e.g requires suctioning in upper airways)
Level 4	Effective cough, clearing secretions with airways clearance techniques.	Effective cough-can clear secretions with breathing exercises.	Effective cough. Clear secretions with airway clearance techniques.

Level 5	Consistent effective voluntary cough, clearing secretions independently	Effective cough, clearing secretions independently	Effective cough, clearing secretions independently

Table 2. Descriptive statistics and inter-rater reliability results of aggregated scores of the Swedish Chelsea Critical Care Physical Assessment Tool (CPAx-Swe).

	Lead rater mean (sd) min-max	Observer rater mean (sd) min-max	ICC (95%CI)	SEM	SDC
CPAx-Swe score	24.4 (10.6) 3-47	24.2 (10.9) 3-47	0.97 (0.95-0.98)	1.79	4.95

sd: standard deviation, min: minimum value, max: maximum value, ICC: intraclass correlation coefficient, 95% CI: 95 % confidence interval, SEM: standard error of measurement, SDC: smallest detectable change

Table 3. Descriptive statistics and inter-rater reliability results of individual items of the Swedish Chelsea Critical Care Physical Assessment Tool (CPAx-Swe).

CPAx-Swe items	Lead rater median (IQR)	Observer rater median (IQR)	kappa values*	agreement
Respiratory function	3.50 (3-4)	3 (3-4.25)	0.95	88%
Cough	4 (2.75-5)	4 (3-5)	0.88	64%
Moving within the bed	2 (1-4)	3 (1-4)	0.86	66%
Supine to sitting on the edge of bed	1 (1-2)	1.5 (1-2)	0.93	74%
Dynamic sitting	4 (3-5)	4 (3-4.25)	0.92	84%
Standing balance	2 (1-3)	2 (1-3)	0.92	78%
Sit to stand	1 (1-2)	1 (1-2)	0.90	74%
Transferring from bed to chair	2 (1-4)	2 (1-4)	0.94	84%
Stepping	1 (0-3)	1 (0-3)	0.88	64%
Grip strength	1 (0.75-3)	1.5 (0-3)	0.98	86%

IQR: interquartile range * quadratic weighted kappa values