INTRODUCTION

An increasing number of patients are surviving critical illness due to advances in medical care (Graf et al., 2005). However both the critical illness itself and the iatrogenic effects of its management, such as enforced immobilization, sedation, mechanical ventilation and physical inactivity, can result in severe and rapid peripheral and respiratory muscle wasting (Latronico and Bolton, 2011; Puthucheary et al., 2013). This is referred to as 'Intensive Care Unit-Acquired Weakness' (ICU-AW). ICU-AW affects around 43% (IQR 9-86%) of critically ill patients (Appleton, Kinsella and Quasim, 2015; Vanhorebeek, Latronico and Van den Berghe, 2020) and is linked to presence of sepsis and multi-organ failure (Fan et al., 2014). The rapid and substantial loss of muscle mass and reduced muscle strength that occurs during the ICU stay can result in prolonged weaning from mechanical ventilation, physical disability and impaired activities of daily living (ADL) (Herridge et al., 2011; Vanhorebeek, Latronico and Van den Berghe, 2020; Visser et al., 2002).

Early physiotherapy for patients in the ICU is essential to minimize the physical consequences of critical illness (Anekwe, Biswas, Bussières and Spahija, 2020; Schaller et al., 2016; Schweickert et al., 2009) and improve long-term outcomes and survival (Iwashyna, Ely, Smith and Langa, 2010; Needham et al., 2012).

Assessing and monitoring physical function is essential to be able to monitor progress thereby helping to focus the patient care, supporting the treatment plan and ensuring continuity of care from the ICU to the ward (Häggström and Bäckström, 2014; Rosa et al., 2016). Several measurement instruments have been developed to assess and monitor physical function in ICU patients in a standardized way (e.g., Physical Functional in ICU Test-scored; Functional Status Score for the ICU; Perme Mobility Scale and The Chelsea Critical Care Physical Assessment tool(CPAx)) (Corner, Soni, Handy and Brett, 2014; Parry et al., 2015; Perme, Nawa, Winkelman and Masud, 2014). The CPAx tool is unique in that it incorporates assessment of respiratory function and the ability to cough as well as functional muscle testing, thereby monitoring the effects of ICU-AW on both peripheral and respiratory muscles. These two items separate the CPAx from other ICU-specific measurement instruments (Parry et al., 2015; Parry, Huang and Needham, 2017).

It is important that measurement instruments have good clinimetric properties such as acceptable reliability and responsiveness. Reliability reflect the consistency of a measurement method (Mokkink, Terwee, Patrick, et al., 2010). A component of this is measurement error, which tests how similar the results of the repeated measurements are, and allows quantification of the systematic and random error of a score that is not attributed to true change in the construct to be measured (Mokkink, Terwee, Gibbons, et al., 2010). Responsiveness is defined as the ability of an instrument to accurately detect change over time (Mokkink, Terwee, Knol, et al., 2010).

The original (English) version of the CPAx has shown good inter-rater reliability (ICC 0.988 to 0.996), validity, responsiveness and a limited floor and ceiling effect in trauma and general ICU (Corner, Handy and Brett, 2016; Corner, Soni, Handy and Brett, 2014; Parry et al., 2015). The CPAx has undergone translation and cross-cultural adaptation from English to Danish including evaluation of face validity of the Danish version of the CPAx (called CPAx-D) (Astrup, Corner, Hansen and Petersen, 2020). Whether the CPAx-D is reliable and responsive to change remains to be investigated. Therefore, the objective of this study was to evaluate the inter-rater reliability and the responsiveness of the CPAx-D in a population of critically ill patients.

METHODS

The study was performed in accordance to **CO**nsensus-based **S**tandards for the selection of health **M**easurement **In**struments (COSMIN) (Mokkink, Terwee, Gibbons, et al., 2010). The study was conducted at the Department of Physiotherapy and Occupational Therapy. An ethical application was submitted however considered unnecessary as the study did not involve changes to usual care The study was approved by the Data Protection Agency (Reference number 681665).

The CPAx Tool

The CPAx consists of 10 domains (respiratory, cough, moving within the 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) which are rated on a 6-point scale from complete dependency (score 0) to independency (score 5) (Corner et al., 2013). The total sum score ranges from 0-50, with a higher score indicating a better physical function.

Participants

Critically ill patients were recruited from three different ICUs, representing a large variation in terms of diagnosis. Inclusion criteria were: 1) adult patients (age 18 and above); and 2) patients considered clinically stable and suitable to receive physiotherapy treatment. Exclusion criteria were: 1) acute neurological diagnoses (e.g., Guillain-Barré syndrome, cerebral hemorrhage or other diseases with acute CNS involvement); and 2) patients unable to speak or understand Danish. Patients with acute neurological diagnoses were not included because the original (English) version of the CPAx was validated in ICU patients without acute neurological diseases, other than ICU-AW.

The following demographic data were extracted from the medical records: sex, age, body mass index (BMI), number of comorbidities, use of mobility aid prior to hospitalization, reason for ICU admission, number of days with mechanical ventilation, length of hospital admission before the ICU and the length of the ICU stay.

Raters

These raters were seven physiotherapists, who routinely treated patients in the ICU (2-15 years of clinical experience in the ICU). Prior to the study, the raters completed the English E-leaning program (Corner, Handy and Brett, 2016), followed by a short training period to familiarize themselves with the CPAx. The raters were calibrated by assessing at least 13 patients in the ICU with the CPAx-D and discussing the assessments with a CPAx experienced supervisor. During the process of completing the E-learning course and the calibration period, the CPAx-D underwent a few adjustments for a clearer understanding of the content. These adaptations were approved by the original developer of the CPAx tool, E.J. Corner, before the use of the final version of CPAx-D in this study (Appendix 1). After the calibration period all seven raters completed two pilot tests in order to practice the standardized reliability test procedure.

Inter-Rater Reliability

Each of the patients were assessed by two of the seven raters on the CPAx-D. To do this the raters observed a physiotherapy session performed by a physiotherapist independent of the project who guided the patients through all 10 items of the CPAx-D. Meanwhile, the two raters present in the room during the treatment session, individually assessed the patient's ability to perform these 10 items on the CPAx-D, without any involvement in the treatment or discussion between raters. Both raters were blinded to the assessment of the other rater. The session lasted for approximately 30-40 minutes.

Responsiveness

Responsiveness was investigated according to the COSMIN guideline (Angst, 2011; de Vet, Bouter, Bezemer and Beurskens, 2001) using the construct approach. Overall, it seems reasonable to assume that the patients' condition will improve considerably from the point of ICU admission to the point of ward transfer. The study group hypothesized that the change in the total CPAx-D score from early admission to leaving the ICU will show large Effect Size (ES) and Standardized Response Mean (SRM) (≥0.8) (Cohen, 1988).

For the responsiveness analysis two assessments at baseline and follow-up were needed. The baseline assessment was collected at an early stage during ICU admission as part of the inter-rater reliability testing, using the score of one of the raters. The follow-up assessment was completed by one of the two inter-reliability raters who had performed the baseline assessment, before the patient was transferred from the ICU to the general ward or shortly after arriving at the general ward (+/- one day).

All patients involved in the inter-rater reliability test were eligible for investigating responsiveness, except patients that were: 1) moved from the ICU to a regular ward within 24 hours after the inter-rater reliability assessment; 2) moved to the regular ward for terminal or palliative care; 3) transferred to another hospital before being follow-up tested or 4) because of death.

<u>Statistical Analysis</u>

A sample size of at least 50 is recommended for inter-rater reliability testing (Mokkink, Terwee, Patrick, et al., 2010). Descriptive statistics were used to present the characteristics of the study population. Normal distributed data were described by the mean and standard deviation (SD), otherwise by median and interquartile range or percentage.

The difference in total CPAx-D score between rater 1 and 2 was analyzed with a paired t-test. Reliability of the total CPAx-D score was investigated using the intraclass correlation coefficient (ICC) model 2.1 with 95% confidence intervals (CI), and a quadratic weighted kappa for the 10 items (de Vet, Terwee, Mokkink and Knol, 2011). ICC and Kappa values between 0.75-0.90 indicate good reliability and ICC and Kappa values ≥0.90 were considered as excellent reliability (Koo and Li, 2016).

Measurement error of the total CPAx-D score was assessed with standard error of measurement (SEM) and minimal detectable change (MDC), and percentage agreement for the 10 items. SEM was calculated as SEM = SD/ √2. Next, SEM was converted into MDC (MDC = 1.96 x √2 x SEM).

A Bland-Altman plot of the total CPAx score was made including 95% limits of agreement (LOA) (de Vet, Terwee, Mokkink and Knol, 2011).

Responsiveness was assessed using ES and SRM with values between 0.5 to 0.8 considered moderate and ≥0.8 considered large (Cohen, 1988). Responsiveness was evaluated by testing the hypothesis that ES and SRM was ≥ 0.8. Possible floor and ceiling effects were also examined using a 15% cut off. The alpha was set at .05 values. Statistical analyses were conducted with STATA 16.1 software (STATA Corp, College Station).

RESULTS

A total of 66 patients were included in the reliability study with 24 of these included in the responsiveness assessment. The characteristics of the study population are presented in Table 1. The mean was 66 years, 65% were men, mean BMI was 27 (SD 5.6), 94% had one comorbidity, 68% had 3 or more comorbidities and 32% needed an mobility aid to hospital admission.

<u>Inter-Rater Reliability</u>

The range of the total CPAx-D score at baseline was 4-44 points, and the range of the CPAx-D scores among the 24 follow-up tests was 10-49 points. There was no significant difference between raters (p=0.81). The ICC was 0.996 (95% CI: 0.993; 0.997), SEM was 0.72 point and MDC 2.0 points (Table 2).

The Bland-Altman plot revealed no signs of heteroscedacity and LOA were +2.0/- 2.0 points (Figure 1). The quadratic weighted kappa on the 10 items individually ranged between 0.914 and 0.995 and the agreement between 97.9% and 99.9% (Table 3).

Responsiveness

The mean difference in CPAx-D score between the baseline and follow-up test was 9.8 points (95% CI 6.2; 13.5) ($P<0.0001$). ES was 1.2 and SRM was 1.1. which was in accordance with the hypothesis, that the change in the total CPAx-D score from early admission to leaving the ICU would show a large ES and SRM (≥0.8).

Floor and Ceiling Effect

None of the 66 included patients scored zero or fifty points on the total CPAx-D score on either assessments. This means there was no ceiling effect or floor effect of the total CPAx score.

DISCUSSION

The objective of this study was to evaluate the inter-rater reliability and the responsiveness of the CPAx-D in a population of critically ill patients in the ICU.

Excellent inter-rater reliability was found both for the total score (ICC = 0.996) and all ten individual items (Kappa = 0.914-0.995). The measurement error in terms of MDC was 2.0 points equal to 8.1% of the mean score of the two raters, which is considered acceptable for individual assessment in CPAx.

Our results are consistent with those found in two other studies investigating reliability. A study of the original CPAx tool demonstrated ICC values ranging from 0.996 to 0.988 (Corner, Handy and Brett, 2016). However, in this study the CPAx assessments were based on videotaped sessions. Another study investigating the Swedish CPAx used the same method as in our study and found results comparable to ours (ICC= 0.97 and quadratic weighted Kappa 0.86-0.98) although the quadratic weighted Kappa values in our study were a bit higher than in the Swedish study (Holdar et al., 2021). This difference might be due to a different training and calibration procedure of the raters.

The results of the change score from baseline to follow-up showed an ES of 1.2 points and a SRM of 1.1. This result is in accordance with the predefined hypothesis which indicated that the CPAx-D was responsive to measure a change of the expected magnitude from early during ICU admission to the time being transferred to a regular ward.

For comparison, a feasibility study investigated the ES of the CPAx in a complex Neurorehabilitation Unit (Wilson-Barry, Spencer and Haworth, 2019), and found an ES of the CPAx of 1.02 which is similar to our result. However, these studies should be compared cautiously due to the difference in patient groups.

The range of the total scores from 4 to 49 points showed that no floor or celling effect was present in our population. Furthermore, the range of scores recorded in this study suggest that the full spectrum of the CPAx scores in all 10 domains were used, indicating that the CPAx is sensitive to the full range of function from the weakest, most passive and unstable patients in the ICU to the patients able to independently mobilize without assistance. A previous study of floor and ceiling effects of the CPAx in an ICU population described a limited floor effect (3.2%) and ceiling effect (0.8%) (Corner, Soni, Handy and Brett, 2014), which supports the efficacy of CPAx during the overall ICU admission.

Limitations of the Study

The present study has some limitations. First, our results can only be generalized within physiotherapists and not necessarily to other health professionals at the ICU. We only included physiotherapists as raters in this reliability study, because the different items of the CPAx-D are focusing on aspects of physical function that are included in the regular assessment and treatment done by the physiotherapists working within the ICU.

Finally, the sample size for responsiveness was small, including only 24 patients. The 42 patients were excluded from the follow-up assessment in line with the exclusion criteria i.e., due to transfer to the regular ward within 24 hours after the baseline assessment, transfer to another hospital or death. Nevertheless, baseline characteristics of patients excluded from the follow-up assessment did not differ from the patients that were included in the responsiveness analysis.

Strengths of the Study

First; random variability between test scores is often caused by subjective evaluations of the raters. In this study we attempted to prevent biases and inaccuracy between the raters by having all seven raters completing a training period. This period consisted of taking the English E-learning course, gaining experience with the CPAx-D during a calibration period and finally completing two pilot tests followed by discussion with a supervisor before participating in the reliability test procedure.

These steps were applied to ensure that the raters had the same level of understanding and experience when applying the CPAx-D. The rationale is, that these steps should also be applied before implementing the CPAx tool in clinical practice to ensure consistency.

Secondly; the raters were physiotherapists who had ample experience with daily treating patients in the ICU. This choice was made to reflect usual clinical practice of the ICU setting, where physiotherapists need to be trained and have some clinical experience before treating patients.

Another strength of CPAx is the ease of use, as the assessment can be done as part of the usual physiotherapeutic intervention with the patient. The assessment itself only requires the usual equipment for mobilization and a dynamometer to test the grip strength. Subsequently, it takes less than 5 minutes to complete the CPAx form.

Perspective and Further Research

The aim is for the CPAx-D to support the interdisciplinary goal setting for ICU patients by reaching different milestones towards independent respiratory function, ability to cough effectively and achieve physical independence, as well as optimizing the written documentation for the benefit of the interdisciplinary collaboration.

Having a core set of measurement instruments to assess physical functioning and treatment effect in patients in the ICU as well as during the overall hospital admission is important. Having just one measurement instrument to cover the entire hospitalization period would be ideal, but may not be possible because of the large variations in physical functioning from early ICU admission until hospital discharge. The CPAx-D could also be used to explore patient recovery trajectories from the ICU to hospital discharge.

CONCLUSION

The CPAx-D showed excellent inter-rater reliability and responsiveness. No floor or ceiling effect was present in the study population. This makes CPAx-D suitable for use in any ICU population both in clinical practice and research.

ACKNOWLEDGMENTS

The authors are grateful to the physiotherapists at (Blinded) the department of Physiotherapy and Occupational Therapy who contributed with their expertise and time to this study. This work was supported by (blinded).

Declaration of Interest

The authors report no conflict of interest.

REFERENCES

Journal article

Anekwe DE, Biswas S, Bussières A, Spahija J 2020 Early rehabilitation reduces the likelihood of developing intensive care unit-acquired weakness: A systematic review and meta-analysis. Physiotherapy 107: 1-10.

Angst F 2011 The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Medical Research Methodology 11: 152.

Appleton RT, Kinsella J, Quasim T 2015 The incidence of intensive care unit-acquired weakness syndromes: A systematic review. Journal of the Intensive Care Society 16: 126-136.

Astrup K, Corner EJ, Hansen MG, Petersen AK 2020 Translation and cross-cultural adaptation of the Chelsea Critical Care Physical Assessment tool into Danish. Physiotherapy Theory and Practice 36: 1027-1034.

Corner EJ, Handy JM, Brett SJ 2016 eLearning to facilitate the education and implementation of the Chelsea Critical Care Physical Assessment: A novel measure of function in critical illness. BMJ Open 6: e010614.

Corner EJ, Soni N, Handy JM, Brett SJ 2014 Construct validity of the Chelsea critical care physical assessment tool: An observational study of recovery from critical illness. Critical Care (London, England) 18: R55.

Corner EJ, Wood H, Englebretsen C, Thomas A, Grant RL, Nikoletou D, Soni N 2013 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 99: 33-41.

de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ 2001 Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. International Journal of Technology Assessment in Health Care 17: 479-487.

Fan E, Cheek F, Chlan L, Gosselink R, Hart N, Herridge MS, Hopkins RO, Hough CL, Kress JP, Latronico N, et al. 2014 An official American Thoracic Society Clinical Practice guideline: The diagnosis of intensive care unit-acquired weakness in adults. American Journal of Respiratory and Critical Care Medicine 190: 1437-1446.

Graf J, Wagner J, Graf C, Koch KC, Janssens U 2005 Five-year survival, quality of life, and individual costs of 303 consecutive medical intensive care patients - A cost-utility analysis. Critical Care Medicine 33: 547-555.

Herridge MS, Tansey CM, Matté A, Tomlinson G, Diaz-Granados N, Cooper A, Guest CB, Mazer CD, Mehta S, Stewart TE, et al. 2011 Functional disability 5 years after acute respiratory distress syndrome. New England Journal of Medicine 364: 1293-1304.

Holdar U, Eriksson F, Siesage K, Corner EJ, Ledström V, Svensson-Raskh A, Kierkegaard M 2021 Cross-cultural adaptation and inter-rater reliability of the Swedish version of the Chelsea critical care assessment tool (CPAX-Swe) in critically ill patients. Disability and Rehabilitation 43: 1600-1604.

Häggström M, Bäckström B 2014 Organizing Safe Transitions from Intensive Care. Nursing Research and Practice 2014: 175314.

Iwashyna TJ, Ely EW, Smith DM, Langa KM 2010 Long-term cognitive impairment and functional disability among survivors of severe sepsis. JAMA 304: 1787-1794.

Koo TK, Li MY 2016 A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine 15: 155-163.

Latronico N, Bolton CF 2011 Critical illness polyneuropathy and myopathy: A major cause of muscle weakness and paralysis. The Lancet Neurology 10: 931-941.

Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, Knol DL, Bouter LM, de Vet HC 2010 Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Medical Research Methodology 10: 82.

Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J 2010 The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Medical Research Methodology 10: 22.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC 2010 The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of Clinical Epidemiology 63: 737-745.

Needham DM, Davidson J, Cohen H, Hopkins RO, Weinert C, Wunsch H, Zawistowski C, Bemis-Dougherty A, Berney SC, Bienvenu OJ, et al. 2012 Improving long-term outcomes after discharge from intensive care unit: Report from a stakeholders' conference. Critical Care Medicine 40: 502-509.

Parry SM, Granger CL, Berney S, Jones J, Beach L, El-Ansary D, Koopman R, Denehy L 2015 Assessment of impairment and activity limitations in the critically ill: A systematic review of measurement instruments and their clinimetric properties. Intensive Care Medicine 41: 744-762.

Parry SM, Huang M, Needham DM 2017 Evaluating physical functioning in critical care: Considerations for clinical practice and research. Critical Care 21.

Perme C, Nawa RK, Winkelman C, Masud F 2014 A tool to assess mobility status in critically ill patients: The Perme Intensive Care Unit Mobility Score. Methodist Debakey Cardiovascular Journal 10: 41-49.

Puthucheary Z, Rawal J, McPhail M, Connolly B, Ratnayake G, Chan P, Hopkinson N, Padhke R, Dew T, Sidhu P, et al. 2013 Acute Skeletal Muscle Wasting in Critical Illness. JAMA: the journal of the American Medical Association 310: 1591-1600.

Rosa RG, Maccari JG, Cremonese RV, Tonietto TF, Cremonese RV, Teixeira C 2016 The impact of critical care transition programs on outcomes after intensive care unit (ICU) discharge: Can we get there from here? Journal of Thoracic Disease 8: 1374-1376.

Schaller SJ, Anstey M, Blobner M, Edrich T, Grabitz SD, Gradwohl-Matis I, Heim M, Houle T, Kurth T, Latronico N, et al. 2016 Early, goal-directed mobilisation in the surgical intensive care unit: A randomised controlled trial. The Lancet 388: 1377-1388.

Schweickert WD, Pohlman MC, Pohlman AS, Nigos C, Pawlik AJ, Esbrook CL, Spears L, Miller M, Franczyk M, Deprizio D, et al. 2009 Early physical and occupational therapy in mechanically ventilated, critically ill patients: A randomised controlled trial. The Lancet 373: 1874-1882.

Vanhorebeek I, Latronico N, Van den Berghe G 2020 ICU-acquired weakness. Intensive Care Medicine 46: 637-653.

Visser M, Kritchevsky SB, Goodpaster BH, Newman AB, Nevitt M, Stamm E, Harris TB 2002 Leg muscle mass and composition in relation to lower extremity performance in men and

women aged 70 to 79: The health, aging and body composition study. Journal of the American Geriatrics Society 50: 897-904.
Wilson-Barry E, Spencer S, Haworth J 2019 Feasibility for the use of the Chelsea Critical Care Physical Assessment tool in a complex Neurorehabilitation Unit. Physiotherapy 105: e99-e100.

Book

Cohen J 1988 Statistical Power Analysis for the Behavioral Sciences. New York.

de Vet HC, Terwee CB, Mokkink LB, Knol DL 2011 Measurement in Medicine - A practical guide. Cambridge University Press.