Intensive Care Med (2015) 41:744–762
DOI 10.1007/s00134-015-3672-x

SYSTEMATIC REVIEW

Selina M. Parry
Catherine L. Granger
Sue Berney
Jennifer Jones
Lisa Beach
Doa El-Ansary
René Koopman
Linda Denehy

Assessment of impairment and activity limitations in the critically ill: a systematic review of measurement instruments and their clinimetric properties

Received: 24 October 2014
Accepted: 16 January 2015
Published online: 5 February 2015

Take home message: 33 different measures were identified, and only 20 have published clinimetric properties.

Electronic supplementary material
The online version of this article (doi:10.1007/s00134-015-3672-x) contains supplementary material, which is available to authorized users.

S. M. Parry (✉) · C. L. Granger · D. El-Ansary · L. Denehy
Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Level 7 Alan Gilbert Building, Parkville, Melbourne, VIC 3010, Australia
e-mail: selina.parry@unimelb.edu.au
Tel.: +61-3-83444171

C. L. Granger · L. Beach
Department of Physiotherapy, Melbourne Health, Melbourne, VIC, Australia

C. L. Granger · S. Berney · L. Denehy
Institute for Breathing and Sleep, Melbourne, VIC, Australia

S. Berney · J. Jones
Department of Physiotherapy, Austin Health, Melbourne, VIC, Australia

R. Koopman
Department of Physiology, The University of Melbourne, Melbourne, VIC, Australia

Abstract *Purpose:* To identify measures used to evaluate the broad constructs of functional impairment and limitations in the critically ill across the continuum of recovery, and to evaluate, synthesise and compare the clinimetric properties of the measures identified. *Methods:* A systematic review of articles was carried out using the databases Medline (1950–2014), CINAHL (1982–2014), EMBASE (1980–2014), Cochrane Library (2014) and Scopus (1960–2014). Additional studies were identified by searching personal files. Eligibility criteria for selection: Search 1: studies which assessed muscle mass, strength or function using objective non-laboratory measures; Search 2: studies which evaluated a clinimetric property (reliability, measurement error, validity or responsiveness) for one of the measures identified in search one. Two independent reviewers assessed articles for inclusion and assessed risk of bias using the consensus-based standards for selection of health status measurement instruments checklist. *Results:* Thirty-three measures were identified; however, only 20 had established clinimetric properties. Ultrasonography, dynamometry, physical function in intensive care test scored and the Chelsea critical care physical assessment tool performed the strongest for the measurement of impairment of body systems (muscle mass and strength) and activity limitations (physical function), respectively. *Conclusions:* There is considerable variability in the type of measures utilized to measure physical impairments and limitations in survivors of critical illness. Future work should identify a core set of standardized measures, which can be utilized across the continuum of critical illness recovery embedded within the International Classification of Functioning framework. This will enable improved comparisons between future studies, which in turn will assist in identifying the most effective treatment strategies to ameliorate the devastating longer-term outcomes of a critical illness.

Keywords Muscle strength · Physical function · Critical care · Intensive care · Outcome measurement · Muscle mass

Abbreviations

6MWT	Six-minute walk test
ADL	Activities of daily living
COSMIN	Consensus-based standards for selection of health status measurement instruments
CPAx	Chelsea critical care physiotherapy assessment tool
FSS-ICU	Functional status score for the intensive care

ICC	Intra-class correlation coefficient	IMS	Intensive care unit mobility scale	PFIT-s	Physical function in intensive care test scored
ICF	International Classification of Functioning	LOS	Length of stay	PRISMA	Preferred reporting for systematic reviews and meta-analyses
ICU	Intensive care unit	MMT	Manual muscle strength testing	SF-36	Short Form-36
ICU-AW	Intensive care unit acquired weakness	MRC-SS	Medical Research Council sum-score		

Introduction

The prolonged diminution in muscle strength and function are concerning disabilities following critical illness [1]. Increasingly, interventions aimed at preventing and minimising these impairments are the focus of research studies [2]. Muscle wasting occurs early and rapidly in the intensive care unit (ICU) setting [3]. A conceptual framework, the International Classification of Functioning (ICF), encompasses three core domains: impairment, activity limitations and participation restriction [4, 5], and has been proposed as a model in which measures can be organised. There are three commonly used study endpoints that sit within this model: muscle mass, strength (body systems) and function (activity limitations).

In muscle mass, strength and function are highly interconnected entities. Muscle mass is a passive non-volitional outcome which enables quantification of muscle morphology, and may relate to measurement of muscle strength and the development of intensive care unit-acquired weakness (ICU-AW) [1]. Muscle strength provides greater detail on the patient's level of impairment, as it is a dynamic measure. At the top of the hierarchy is function, which is the most patient-centred outcome and provides information on activity limitation within the ICF framework. The measurement of function is complex, containing information about task completion (cognition), coordination, processing of visual information and central motor drive, and the activation of signalling pathways from the motor cortex to the muscle [6]. Measurement of strength and function requires patients to be alert and able to cooperate with testing. This is in contrast to measurement of muscle mass, which can be quantified, using non-volitional methods such as ultrasonography.

When selecting the most appropriate measure to evaluate efficacy and change over time, clinicians and researchers need to consider whether the clinimetric properties of the measure of interest have been established. Reliability determines the ability of an instrument to obtain accurate results, which are free from measurement error when the instrument is repeated by multiple assessors (inter-rater reliability) or longitudinally (intra-rater reliability or test–retest reliability) [7, 8]. Validity determines the ability of an instrument to measure what it is intended to measure, i.e. how well an instrument obtains data, as hypothesised, when compared to an instrument measuring a similar construct (construct validity–hypotheses testing); how well an instrument performs in comparison to the "gold standard" measure (criterion–concurrent validity); and how well data from an instrument predicts a future score or outcome (criterion–predictive validity) [7, 8]. Responsiveness refers to the ability of an instrument to detect a true change in the score obtained which is statistically or clinically meaningful over time [8]. There are two main methods used to determine the minimal important difference (MID): a distribution-based method and an anchor-based method [9]. The anchor-based method takes into account the patient perception of change using anchors such as much worse and much better in a scale such as the global rating of change scale [9]. Measures developed for one setting or patient population should only be extrapolated with caution [8]. In ICU, the environment, patient alertness, sedation, delirium and severity of illness, time, resources and expertise are factors which influence the choice of measure [10], as well as the clinimetric properties of the measures [11].

To date, there has been no systematic, comprehensive evaluation and synthesis of measures used to assess muscle mass, strength and function in the critically ill across the continuum of recovery including examination of the clinimetric properties of these measures. There have been two published systematic reviews addressing use of outcome measures in the critically ill [12, 13]. However, these reviews are either focused on one specific aspect of clinimetric evaluation (e.g. only reliability [12]), or one type of outcome of interest (e.g. only physical function [13]).

Therefore, the objectives of this review were to:

- identify measures which are used to evaluate muscle mass, strength and function in the critically ill population, at any point along the trajectory of critical illness recovery (including in ICU, hospital, and post-hospitalisation settings) and
- evaluate, synthesise and compare the clinimetric properties of the measures identified.

The consensus-based standards for the selection of health status measurement instruments (COSMIN) and the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines were followed (http://www.prisma-statement.org).

Methods

Protocol

The review was registered on PROSPERO (CRD4201 400893). The search for this systematic review was conducted in two parts: part 1 involved the identification of measures, which have been used to evaluate muscle mass, strength and function in the critically ill. This first search allowed a list of measures to be generated. Part 2 involved a second search conducted to identify papers examining the clinimetric properties of measures identified in part 1.

Part 1: identification of measures

Five electronic databases were searched by one reviewer using a systematic, comprehensive and reproducible search strategy [Electronic Supplementary Material (ESM) Table E1]. Electronic databases were accessed via The University of Melbourne, Australia library with the last search run on 17 October 2014. Two independent reviewers determined eligibility against pre-determined criteria (Table 1). A list of measures, was generated from the results of part 1.

Part 2: clinimetric properties of measures

Five electronic databases were searched by one reviewer (S.P.) (Fig. 1) with the last search run on 17 October 2014. The search filter adopted (ESM Table E1) was based on guidelines provided by Terwee and colleagues [14] for systematic reviews examining the clinimetric properties of measures. The study selection and data extraction followed the same methodology as described for part 1. Two independent reviewers (S.P., C.G.) used the COSMIN checklist, a validated tool, to evaluate the risk of bias of the included studies from part 2 [15]. Each study was evaluated on the relevant item(s) of the COSMIN checklist (reliability; measurement error; hypotheses testing; criterion validity and responsiveness). An overall quality score for each item was obtained by using the lowest score recorded [15]. The agreement between reviewers was estimated using percentage agreement and the kappa statistic [16].

Results

Part 1: identification of measures

A list of 33 measures was generated (muscle mass $n = 3$, strength $n = 4$ and function $n = 26$) (ESM Figure E1; Fig. 1). Percentage of agreement for title and abstract was 96 % ($\kappa = 0.90$) and for full-text was 93 % ($\kappa = 0.86$).

Part 2: clinimetric properties of measures

Study selection and study characteristics

A total of 47 articles were included (Fig. 1). Percentage of agreement for title and abstracts was 91 % ($\kappa = 0.82$) and for full-text was 92 % ($\kappa = 0.80$). The characteristics of included studies are summarized in Table 2.

Table 1 Study eligibility for inclusion in systematic review (parts 1 and 2)

Characteristics	Inclusion	Exclusion
Design	Quantitative study designs RCTs, pseudo-RCTs, cohort studies, case–control studies or case series as per NHMRC classification	Studies not published in a peer-reviewed journal, descriptive commentary (reviews, editorials, narratives), conference abstracts
Participants	Adults >18 years of age in the ICU setting or survivors of ICU at any point in the continuum of recovery	<6 participants in the study Specialized patient populations such as trauma, stroke, burns, transplant
Intervention	Did not form part of the eligibility criteria	
Outcome measures	Part 1: Objective OM which on face validity aimed to measure muscle mass, strength or function Part 2: Ax of clinimetric properties (reliability, measurement error, validity, responsiveness) of an OM identified in part 1	Part 1: Laboratory measures OM originally designed as questionnaires Part 2: Studies reporting on development of an OM without investigation of the clinimetric properties of the outcome Indirect validity (validation of alternative OM against one of the OM identified in part 1)
Publication	No language restrictions applied to the initial search Part 2: No date restrictions	Part 1: Papers published prior to 2004 excluded

ICU intensive care unit, *NHMRC* National Health and Medical Research Council, *OM* outcome measure, *RCT* randomised controlled trial

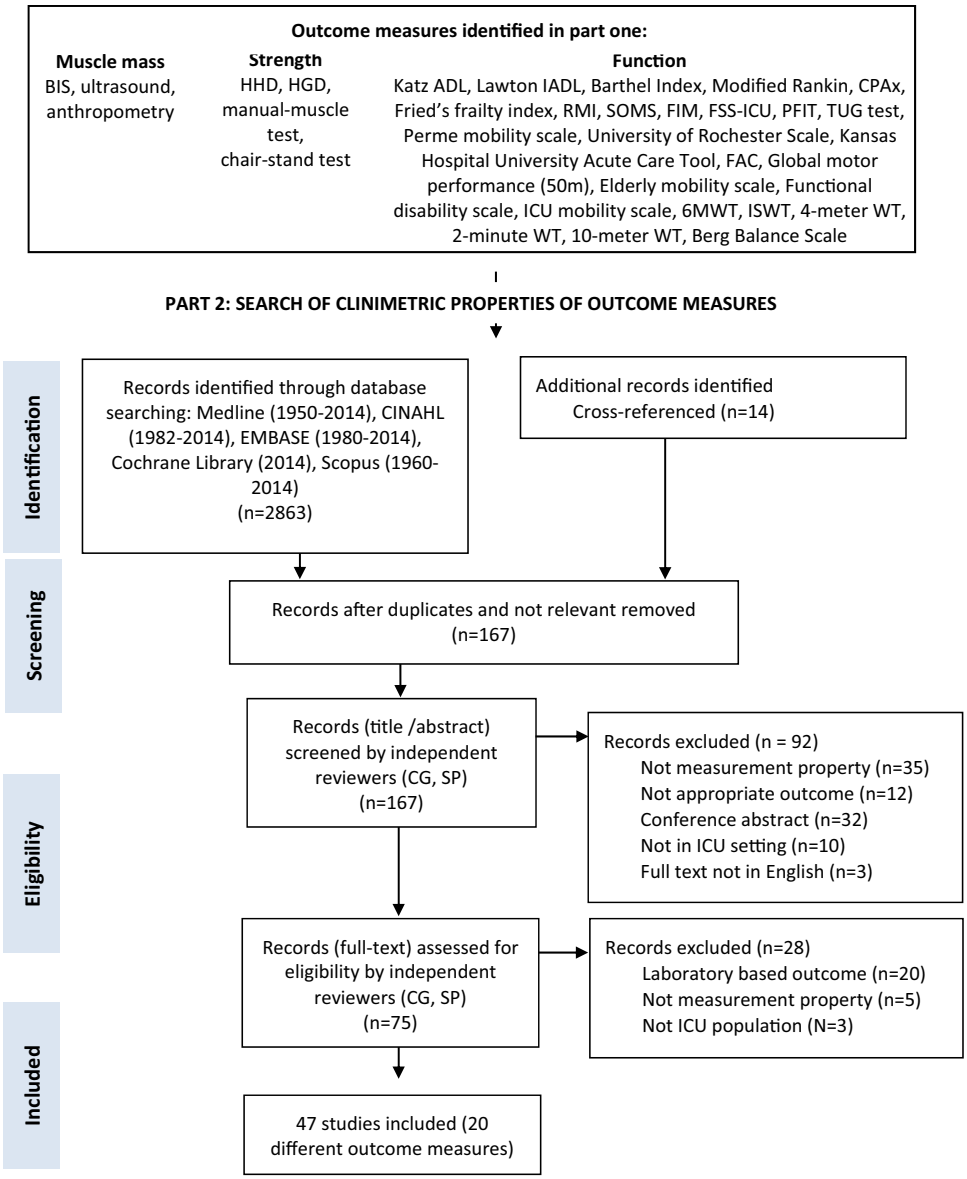

Fig. 1 Flow diagram of clinimetric properties search—part 2. *ADL* activities of daily living, *Ax* assessment, *CINAHL* cumulative index to nursing and allied health literature, *CPAx* Chelsea critical care physical assessment tool, *EMBASE* the Excerpta Medica database, *FAC* functional ambulatory category, *FIM* Functional Independence Measure, *FSS-ICU* functional status score for the intensive care, *IADL* independence in activities of daily living, *ICU* intensive care unit, *ISWT* incremental shuttle walk test, *n* number, *PEDRO* physiotherapy evidence database, *PFIT* physical function in intensive care test, *RMI* Rivermead mobility index, *SOMS* surgical intensive care unit optimal mobilisation score, *TUG* timed-up-and-go-test, *WT* walk test, *6MWT* 6-min walk test

Outcome measures

Clinimetric properties evaluated by studies were: reliability (studies $n = 16$), measurement error ($n = 4$), construct validity (hypothesis testing) ($n = 31$), criterion-predictive validity ($n = 18$) and responsiveness ($n = 11$) (Tables 3, 4, ESM Tables 2–4).

Risk of bias

Percentage agreement for risk of bias assessment between reviewers was 97 % ($\kappa = 0.95$). Overall, studies scored "fair" or "poor" for the measurement properties evaluated (ESM Table E2). The worst scored area amongst studies included was design requirements (sample size and lack of a priori hypotheses).

Study results

Study results are summarised in Table 3 and described in the following sections. Individual study results are presented in ESM Tables E3–5. Overall ultrasonography, dynamometry, physical function in intensive care test scored (PFIT-s) and the Chelsea critical care physical assessment tool (CPAx) performed the best in terms of clinimetric properties (Table 3).

Table 2 Study characteristics of studies included in part 2

Author, year, location	*n*	Gender M/F	Age Mean ± SD or median [IQR]	Outcome measure	Setting	Severity of illness APACHE II mean ± SD or median [IQR]	ICU LOS Median [IQR]	Timepoint/s of assessment for outcome of interest
Muscle mass								
Baldwin 2014 Australia [17]	16	9/7	62 ± 17	US, BIS, HHD, HGD	ICU	20 [12–44]	NR	Awakening and 1 day later
Grimm 2013 Germany [18]	28	25/3	70 [62–75]	US	ICU	23 ± 7	35 ± 71	D2–5 and D14 (after onset of severe sepsis or septic shock)
Savalle 2012 France [19]	49	30/29	66 ± 16	BIS, circumference	Medical/surgical ICU	NR	3 (range 2–13)	ICU admission
Baldwin 2012 Australia [20]	17	10/7	78 (range 30–87)	BIS	Medical/surgical ICU	20 ± 5	NR	Random cross-section (ICU LOS at time of first test) median 13 (range 5–21) with 2nd test repeated 2 days later
Gruther 2008 Austria [21]	118	88/30	Mean age 55	US, circumference	ICU	NR	NR	($n = 17$) Study 1: baseline and D28 ($n = 101$) Study 2: random cross-section one time-point
Reid 2004 United Kingdom [22]	50	26/24	56 (range 19–79)	US, circumference	ICU	17 (range 2–43)	NR	Serial measurements every 1–3 days for between 5 and 39 (median 7) days
Faisy 2000 France [23]	51	42/9	69 ± 11	BIS, circumference	Respiratory ICU	NR	NR	Random cross-section
Frankenfield 1999 United States [24]	46	31/15	Mean (SEM) M: 44 ± 4 F: 58 ± 6	BIS	Surgical ICU	NR	NR	Random cross-section
Morais 1998 Brazil [25]	50	24/26	57 ± 8	BIS	Surgical ICU	NR	6 ± 7	ICU admission and ICU DC
Phang 1996 Canada [26]	45	27/18	58 ± 17	BIS	ICU	18 ± 7	NR	Baseline and 7 days
Campbell 1995 United Kingdom [27]	9	8/1	63 (range 58–68)	US, circumference	ICU	NR	NR	Every 1–4 days for between 5 and 11 days min of 5 measures
Robert 1993 United States [28]	33	16/17	56 ± 18	BIS	Medical/surgical ICU	NR	NR	Baseline and ICU DC
Muscle strength								
Yosef-Brauner 2014 Israel [29]	18	7/11	Gp1: 62 ± 12 Gp2: 52 ± 18	MRC	ICU	Gp1: 21 ± 5 Gp2: 19 ± 9	NR	Baseline, 48–72 h later and ICU DC
Fan 2013 United States [30]	520	292/228	52 [42–63]	Circumference, MRC HGD, 6MWT	ICU	26 [20–33]	13 [7–21]	Hosp DC, 3, 6, 12 and 24 months after onset of ALI

Table 2 continued

Author, year, location	*n*	Gender M/F	Age Mean ± SD or median [IQR]	Outcome measure	Setting	Severity of illness APACHE II mean ± SD or median [IQR]	ICU LOS Median [IQR]	Timepoint/s of assessment for outcome of interest
Connolly 2013 United Kingdom [10]; study 1 (inter-observer agreement)	20	12/8	68 [52–75]	MRC	Medical/surgical ICU	20 [16–24]	34 [26–58]	Awakening
Connolly 2013 United Kingdom [10]; study 2 (clinical prediction)	94	64/30	66 [55–76]	MRC	Medical/surgical ICU	17 [15–22]	11 [6–25]	Awakening and 7 days post-awakening
Lee 2012 United States [31]	104	59/45	61 ± 18	MRC, HGD	Surgical ICU	15 ± 9	5 [3–10]	Awakening
Hermans 2012 Belgium [62]; study 1 (MRC-SS)	75	38/37	59 [52–71]	MRC	Medical/surgical ICU	NR	22 [15–30]	Random cross-section
Hermans 2012 Belgium [62]; study 2 (HGD)	46	27/19	48 [47–68]	HGD	Medical/surgical ICU	NR	15 [9–32]	Random cross-section
Baldwin 2013 Australia [32]	17	10/7	78 [46–82]	MRC, HGD, HHD (quads and biceps)	ICU	20 ± 5	18 [12–21]	Awakening
Vanpee 2011 Belgium [33]	51	32/19	64 [53–72]	HHD	Medical/surgical ICU	NR	NR	Awakening
Hough 2011 United States [34]	30	21/9	49 ± 15	MRC	ICU for 10 pts Ward for 20 pts	NR	NR	Awakening
Fan 2010 United States [35]	19	NR	NR	MRC	$n = 9$ ICU $n = 10$ simulated	NR	NR	Random cross-section
Brunello 2010 Switzerland [36]	39	28/11	67 ± 14	MRC, Barthel	ICU	NR	12 ± 7	Neuro exam daily upon awakening until ICU DC, D28, and 6 months post-ICU DC
Sharshar 2009 France [37]	115	75/40	65 [52–77]	MRC	Medical and surgical ICUs	NR	23 [16–35]	Days 7, 14, 21 and 30 after awakening in patients who were not discharged from ICU and who had MRC < 48
Ali 2008 United States [38]	136	65/71	58 ± 16	MRC, HGD	Medical ICU	NR	NR	Awakening
Function								
Corner 2014 United Kingdom [39]	499	NR	62 ± 18	CPAx	Medical/surgical ICU	16 [10–20]	12 ± 16	Assessed at least 3× per week (including ICU discharge)
Denehy 2014 Australia and United States [40]	177	114/63	60 [49–72]	6MWT, TUG, Berg balance scale, chair stand test	Medical and surgical ICU	19 [16–23]	8 [6–14]	3 months post-ICU DC

Table 2 continued

Author, year, location	*n*	Gender M/F	Age Mean ± SD or median [IQR]	Outcome measure	Setting	Severity of illness APACHE II mean ± SD or median [IQR]	ICU LOS Median [IQR]	Timepoint/s of assessment for outcome of interest
Tripathy 2014 India [41]	109	80/29	75 ± 8	Katz ADL	Follow-up post-ICU admission	19 ± 7	7 ± 3	Baseline, 28 days post-admission and 12 months post-hosp DC
Nawa 2014 United States [42]	20	12/8	65 [20–86]	Perme mobility scale	Cardiovascular ICU	17 [7–30]	4 [1–42]	At time of initial PT evaluative
Nordon-Craft 2014 United States [43]	51	32/19	51 ± 16	PFIT, MRC, grip	ICU	18 ± 6	20 [12–26]	Awakening
Hodgson 2014 Australia [44]	100	38/62	58 ± 17	ICU mobility scale	Medical/surgical ICU	19 ± 7	NR	Random cross-section in ICU
Baldwin 2014 United States [45]	22	15/7	77 ± 9	Fried's frailty index	Post-medical ICU on the ward	27 ± 10	5 [3–9]	Hosp DC, 1 and 6 months post-hosp DC
Denehy 2013 Australia [46] (clinimetric sample)	116	70/46	59 ± 15	PFIT	ICU and hosp setting	19 ± 6	7 [6–11]	D5 of ICU adm and ICU DC
Corner 2013 United Kingdom [47]	33	25/8	67 [51–75]	CPAx	ICU	20 ± 6	NR	Random cross-section
Abd-El-Gawad 2013 Egypt [48]	65	42/23	70 ± 11	Katz ADL	Geriatric and chest ICU	NR	11 ± 8	Preadmission score (1 month prior)
Thrush 2012 United States [49]	101	62/39	70 [61–78]	FSS-ICU	LTACH	NR	24 [19–43]	LTACH admission and discharge (and every 2 weeks in between)
Kasotakis 2012 United States [50]	113	66/47	60 ± 18	SOMS	Surgical ICU	16 ± 7	5 ± 5	Daily in the morning
Allison 2012 Australia [51]	173	104/69	57 ± 16	6MWT	Post-hosp DC	19 ± 10	9 ± 8	Week 1, 8 and 26 after hosp DC
Vest 2011 United States (1 month cohort) [52]	110	49/61	73 ± 8	Katz ADL	Post-hosp DC (MICU cohort)	21 ± 6	NR	1 month and 1 year after ICU DC
Daubin 2011 France [53]	100	65/35	79 ± 3	Katz ADL	Medical ICU	24 [18–30]	NR	Preadmission (1 month prior) and 3 months post-ICU admission
Clini 2011 Italy [54]	77	46/31	75 ± 7	Katz ADL, FIM, strength	Weaning unit	12 ± 4	Mean (range) 51 (12–115) days	Mean (SD) 24 (3) days before admission to weaning unit Measures at baseline and post-intervention period
Skinner 2009 Australia [55]	12	7:5	57 ± 13	PFIT	Medical/surgical ICU and weaning unit	Range (10–25)	NR	Random cross-section
Sacanella 2009 Spain [56]	230	140/90	75 ± 6	Barthel and Lawton IADL	Medical ICU	20 ± 6	12 ± 12	ICU admission

Table 2 continued

Author, year, location	*n*	Gender M/F	Age Mean ± SD or median [IQR]	Outcome measure	Setting	Severity of illness APACHE II mean ± SD or median [IQR]	ICU LOS Median [IQR]	Timepoint/s of assessment for outcome of interest
Van der Schaaf 2008 The Netherlands [57]	69	43/26	60 [49–71]	Barthel, HGD, FAC, MRC	ICU	16 [12–20]	7 [5–17]	3–5 days post-ICU DC
Swafford 2008 United States [58]	50	NR	Mean age 73 (range 40–96)	KHU scale, FIM	ICU	NR	NR	First and last PT visits in acute hospital
Chiang 2006 Taiwan [59]	32	24/8	Median [IQR] Control: 79 [73–83] Rehab: 75 [63–80]	HHD, Barthel and FIM	Respiratory weaning unit	NR	NR	Admission to weaning unit, week 3 and week 6
Bo 2003 Italy [60]	659	352/307	77 ± 8	Katz ADL and Lawton IADL	Medical ICU	13 ± 5	7 ± 6	Admission (2 weeks pre-admission)
Dardaine 2001 France [61]	116	65/51	77 ± 5	Katz ADL, circum	Medical/surgical ICU	NR	Median 11	Admission (pre-admission 3 months prior)

adm admission, *ADL* activities of daily living, *ALI* acute lung injury, *APACHE II* acute physiology and chronic health evaluation 2 score, *BIS* bioimpedance spectroscopy, *CPAx* Chelsea critical care assessment tool, *D* day, *DC* discharge, *F* female, *FAC* functional ambulation category, *FIM* Functional Independence Measure, *FSS-ICU* functional status score for the intensive care, *Gp* group, *HGD* handgrip dynamometry, *HHD* handheld dynamometry, *hosp* hospital, *h* hours, *IADL* independence in activities of daily living, *ICU* intensive care unit, *IQR* interquartile range, *KHU* Kansas Hospital University acute care tool, *LOS* length of stay, *LTACH* long-term acute care hospital, *M* male, *min* minimum, *MICU* medical ICU, *MRC* Medical Research Council, *n* number, *Neuro* neurological, *NR* not reported, *OM* outcome measure, *PFIT* physical function in intensive care, *PT* physiotherapist, *pts* patients, *quads* quadriceps, *SD* standard deviation, *SEM* standard error of the mean, *SOMS* surgical optimal mobility scale, *US* ultrasound, *6MWT* 6-min walk test

Table 3 Results: synthesis of evidence regarding clinimetric properties (comparison of outcome measures)

Outcome measure	Reliability, measurement error and responsiveness	Construct validity	Criterion predictive validity
Muscle mass			
Bioimpedance spectroscopy	1 study, excellent intra-session and test–retest reliability for SFB7 device and poor for In Body S20 device [20] 1 study, small to moderate error margins for reactance and resistance using SFB7 device [19]	4 studies, no correlation with biceps/forearm muscle thickness [17], 7-day energy balance or fluid balance [26]; fair correlation with plasma albumin, prealbumin levels and resting metabolic rate [23, 24]; moderate to good correlation with oxygen consumption; protein intake and quadriceps thickness [17, 24, 28]; good to excellent correlation with energy intake [28]	2 studies, BIS defined malnutrition found to be predictive of higher ICU mortality [23] and BIS parameters (resistance, reactance and body impedance co-efficient) not predictive of in-hospital mortality [25]
Ultrasound	1 study, excellent intra-rater reliability for measurement of muscle thickness [17] 1 study excellent intra- and inter-rater reliability for Heckmatt qualitative analysis of muscle echointensity [18] 1 study, non-significant change in mean echointensity score over a 10-day period in the ICU (D4–D14) [18]; 1 study, muscle thickness measurement sensitive to change over time in the ICU setting [22]	5 studies, muscle thickness: no relationship with daily energy balance [22]; moderate to excellent correlation with muscle strength and FFM derived from skinfold thickness [17, 27]; negative correlation with ICU LOS [21] US echogenicity: fair to good correlation with fasciculation scoring on US [18]	X
Circumference	Two studies, no consistent change in circumference pattern over ICU admission period [22, 27]	3 studies, moderate correlation between peripheral circumferences [19], FFM from skin-fold thickness [27]; and strong correlation with HGD [30]	1 study, normal MAC on admission predictive of higher mortality [23] 1 study, MAC not predictive of ICU mortality, MAC under 10th percentile predictive of 6-month mortality almost 3.5 times more likely to decease [61]
Muscle strength			
Hand-held dynamometry	4 studies, excellent intra and inter-rater reliability (except 1 study good inter-rater reliability only for biceps HHD) [17, 32, 33, 59] Moderate to large SEM and MDD for biceps and quadriceps HHD [32]	X	X
Hand-grip dynamometry	3 studies, excellent intra and inter-rater reliability [17, 32, 62] Moderate SEM for HGD, and large MDD [32]	3 studies, no correlation with ICU LOS, hospital LOS or MV duration [63]; good to excellent correlation with MRC and MIP [30, 63]; good test performance for diagnosis of ICU-AW with high sensitivity, specificity, NPV and moderate PPV [38], significant relative reduction in ICU and hospital-free days [38]	2 studies, conflicting findings for prediction of mortality [38, 63]

Table 3 continued

Outcome measure	Reliability, measurement error and responsiveness	Construct validity	Criterion predictive validity
Medical Research Council score	5 studies, excellent inter-rater reliability for overall MRC score in critically ill [10, 34, 35, 38, 62] 2 studies, poor to excellent inter-rater reliability for individual muscle group scores in critically ill [10, 34] 4 studies, slight to substantial inter-rater agreement for Dx of ICU-AW in ICU [10, 34, 35, 62], and almost perfect inter-rater reliability for Dx of ICU-AW in ward setting [34]	8 studies, fair correlation with Barthel and Elderly Mobility scale [10]; moderate to excellent correlation with HGD, and MIP [29, 30] [10, 63], lower MRC score or Dx of ICU-AW is associated with longer ICU and hospital LOS [10, 36, 38, 63]	4 studies demonstrated MRC score or diagnosis of ICU-AW was predictive of ICU and hospital mortality [36–38, 63] 1 study demonstrated MRC score not associated with ICU or hospital mortality [10]
Chair-stand test	X	X	X
Function			
6MWT	1 study, small measurement error overall but differences between individuals for the second test < 15 % [51] floor effect of 3.9 % at 3 months post-ICU discharge [40]	1 study, moderate to good correlation with SF-36 physical function domain [51] 1 study excellent correlation with other functional measures: TUG test, Berg balance scale, SF-36 PF domain and moderate relationship to 5xSTS [40]	X
ISWT	X	X	X
4-m walk test	X	X	X
2-min walk test	X	X	X
10-m walk test	X	X	X
Katz ADL	X	Significant correlation between Katz ADL and SF-36 PF and MF domains at 1-month [52], and no to poor correlation for FIM, muscle strength [54] and 12-month SF-36 PF and MF domain scores [52]	6 studies, with majority of studies ($n = 3$) supportive of predictive ability in relation to short-term mortality [41, 48, 60]; predictive of increased MV duration [48]; 3 studies, not predictive of long-term mortality > 3–6 months [41, 53, 61]
Lawton IADL	X	X	1 study, predictive of post-hospital and cumulative mortality [56]; conflicting results regarding prediction of in-hospital mortality [56, 60]
Barthel index	1 study, moderate to excellent responsiveness observed in a weaning facility in response to a 6-week rehab programme [59]	2 studies, fair correlation with MV duration; moderate correlation with HGD; excellent correlation with FAC [57]; moderate to good correlation with respiratory muscle strength, and peripheral muscle strength and ventilator free time in a weaning facility [59]	1 study, predictive of cumulative mortality but not in-hospital or post-hospital mortality [56]
ICU mobility scale	1 study, excellent inter-rater reliability for overall score [44]	X	X
SOMS	1 study, substantial inter-rater agreement for overall score [50]	1 study, fair correlation with hospital LOS; moderate to good correlation with ICU LOS and HGD [50]	1 study, predictive of in-hospital mortality [50]
Perme mobility scale	1 study, excellent inter-rater reliability for overall score; good to excellent for individual items [42]	X	X

Table 3 continued

Outcome measure	Reliability, measurement error and responsiveness	Construct validity	Criterion predictive validity
PFIT	1 study, excellent inter-rater reliability for individual items with little measurement error [55] Significant mean difference from baseline to post-weaning for individual items of the PFIT scale [55] Large effect size index within the ICU setting, with established MCID of 1.5 points out of 10 [64] 1 study (Australia), floor effect of 21.5 % (n = 31/144) at awakening; ceiling effect at ICU DC of 22.2 % (n = 26/117) [64] 1 study (United States), floor effect of 32 % (n = 11/34) at awakening; floor and ceiling effect at ICU DC of 5 % (n = 2/39); large test responsiveness in United States study ES = 1.14 between baseline and ICU DC [43]	2 studies, fair to excellent correlation with 6MWT and MRC score; excellent correlation with grip strength, and good correlation with TUGT [43, 64]	2 studies, higher admission PFIT predictive of higher MRC and reduced likelihood of DC to rehabilitation or LTACH, conflicting findings for predicting discharge to home [43, 64] ICU DC MRC-SS cut point of 41.5 out of 60 predictive of patient's ability to perform the standing components of the PFIT-s [43]
CPAx	1 study, excellent inter-rater reliability for overall and individual items [47] Ceiling effect 0.8 %, n = 4/499 in ICU, and floor effect of 3.2 % (n = 16/499) [39]	1 study, moderate to good correlation with MV days, peak cough flow, AusTOMs, MRC; and good to excellent correlation with SF-36 physical function domain and no relationship with SF-36 mental function domain [47] 1 study, moderate to strong correlation between ICU discharge CPAx score and hospital discharge destination [39]	X
FSS-ICU	1 study, small effect size index for LTACH setting, significant responsiveness to change over time from admission to discharge from LTACH unit [49]	1 study, significant differences between discharge FSS-ICU scores, higher scores correlated with discharge home [49]	X
FIM	X	1 study, moderate to strong correlation between change in FIM score in a weaning facility and changes in respiratory and peripheral muscle strength, and ventilator free time [59]	X
Fried's frailty index	X	X	1 study, predictive of lower Katz ADL score at 1-month and higher-6-month mortality [45]
Modified Rankin	X	X	X
RMI	X	X	X
TUG test	Floor effect of 2.3 % at 3 months post-ICU discharge [40]	1 study, excellent correlation between TUG test and other functional measures: 6MWT, 5xSTS and Berg Balance Scale, moderate correlation to SF-36 PF domain [40]	X
University of Rochester scale	X	X	X
Kansas Hospital University acute care tool	X	1 study, strong correlation between change in KUH scores and change in transfer and gait FIM scores in the acute hospital [58]	X
FAC	X	X	X

Table 3 continued

Outcome measure	Reliability, measurement error and responsiveness	Construct validity	Criterion predictive validity
Global motor performance (50 m)	X	X	X
Elderly mobility scale	X	X	X
Functional disability scale	X	X	X
Berg balance scale	1 study ceiling effect of 46 % at 3 months post-ICU discharge [40]	X	X

ADL activities of daily living, *AusTOMs* Australian Therapy Outcome Measures, *CPAx* Chelsea critical care physical assessment tool, *DC* discharge, *Dx* diagnosis, *FAC* functional ambulation categories, *FIM* Functional Independence Measure, *FFM* fat-free mass, *FSS-ICU* functional status score for the intensive care, *HGD* handgrip dynamometry, *HHD* hand-held dynamometry, *IADL* instrumental activities of daily living, *ICU* intensive care unit, *ICU-AW* intensive care unit acquired weakness, *ISWT* incremental shuttle walk test, *LOS* length of stay, *LTACH* long term acute care hospital, *m* meters, *MAC* mid-arm circumference, *MCID* minimal clinical important difference, *MF* mental function, *MIP* maximum inspiratory pressure, *MMD* minimal detectable difference, *MRC* Medical Research Council, *MV* mechanical ventilation, *NPV* negative predictive value, *OR* odds ratio, *PFIT* physical function in intensive care test, *PPV* positive predictive value, *RMI* Rivermead mobility index, *SEM* standard error of measurement, *SF-36* Short Form 36 health survey, *SOMS* surgical optimal mobility scale, *TUG* timed up and go test, *US* ultrasound, *6MWT* 6-min walk test

Muscle mass

In the ICU, muscle mass was evaluated using three different approaches: anthropometry, bioimpedance spectroscopy (BIS) and ultrasonography (Fig. 1; Table 3). The reliability and measurement error of anthropometry has not been examined in individuals with critical illness. Circumference measures of limb size were not sensitive to change over time [22, 27] (Table 3; ESM Table E5). BIS had high intra-session and test–retest reliability when using the SFB7 Bioimped device (ICC > 0.94) [20]. A moderate to excellent relationship was established between BIS and quadriceps thickness ($r^2 = 0.61$, $p \leq 0.001$) [17]. There is conflicting evidence for the predictive ability of both anthropometry and BIS in relation to mortality (Table 3; ESM Table E4).

There was excellent intra-rater reliability for measurement of muscle thickness (ICCs ≥ 0.98) and echogenicity (ICC > 0.90) using ultrasonography [17, 18] (Table 3). There was a fair to moderate correlation between upper limb muscle thickness and strength ($r^2 = 0.43$–0.52, $p < 0.01$) and little correlation for quadriceps thickness and strength ($r^2 = 0.22$, $p = 0.07$) [17] (Table 3). Muscle thickness was negatively correlated with ICU length of stay (LOS) ($p < 0.001$) [21] (Table 3; ESM Table E4). The criterion predictive validity of ultrasonography has not been examined but ultrasonography was sensitive to changes in muscle thickness (with a reduction of 1.6–6 % per day in quadriceps thickness) over the ICU admission in two studies [22, 27] (Table 3; ESM Table E5).

Muscle strength

Strength has been evaluated using handgrip dynamometry (assessment of grip strength only), hand-held dynamometry (used to assess all other muscle groups), manual-muscle strength testing (MMT) and the chair-stand test (Fig. 1). The clinimetric properties of only three of these tests (hand-held dynamometry, handgrip dynamometry, MMT) have been evaluated (Table 3; ESM Tables E3–5), with reliability and measurement error the most extensively evaluated clinimetric constructs (Table 3, ESM Table E3).

Manual muscle strength using the Medical Research Council sum-score (MRC-SS) is the most commonly utilised measure for evaluating strength in the ICU setting (Table 2). Whilst there is excellent inter-rater reliability for overall MRC-SS [10, 34, 62], the inter-rater reliability for individual muscle group scores ranges from poor to excellent. Agreement for diagnosis of ICU-AW (< 48 out of 60) is inconsistent ranging from slight to substantial agreement in the ICU [10, 34, 62] and almost perfect in the ward setting [34]. The MRC-SS has a fair correlation with functional outcomes (Barthel index and elderly

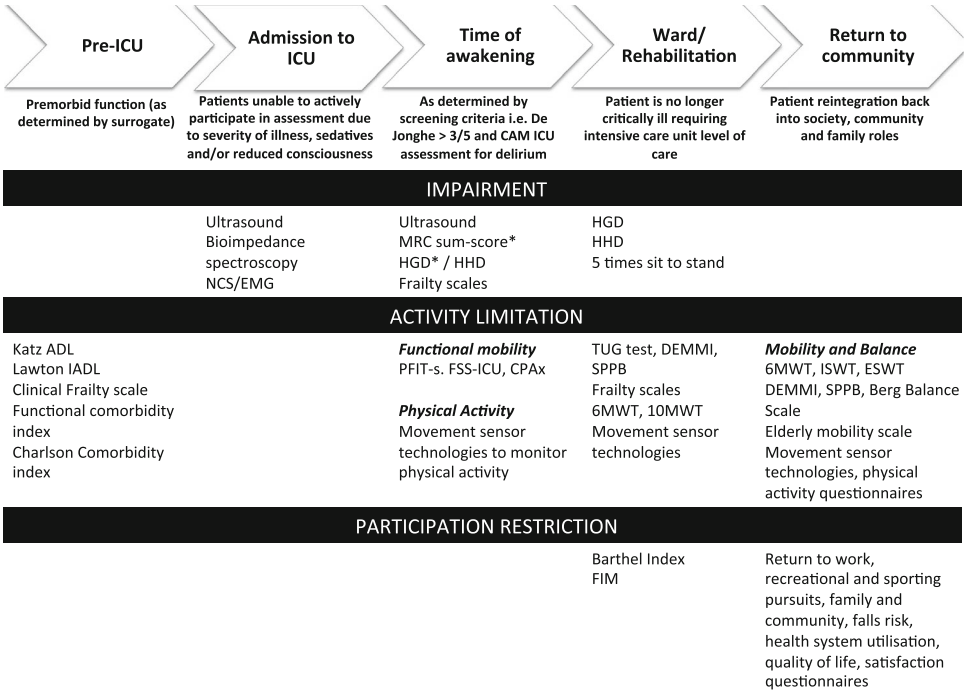

Fig. 2 Suggested schematic guide to mapping of outcome measures within the ICF framework. *ADL* activities of daily living, *CPAx* Chelsea critical care physical assessment tool, *DEMMI* De Morton mobility index, *EMG* electromyography, *ESWT* endurance shuttle walk test, *FSS-ICU* functional status score for the intensive care, *HGD* handgrip dynamometry, *HHD* hand-held dynamometry, *IADL* independence in activities of daily living, *ICU* intensive care unit, *IMS* ICU mobility scale, *MRC* Medical Research Council, *NCS* nerve conduction study, *PFIT-s* physical function in independence test scored, *SPPB* short physical performance battery, *TUG* timed-up-and-go-test, *6MWT* 6-min walk test. *Asterisk* methods for clinically diagnosing the presence of ICU-AW on awakening

mobility scale) in terms of criterion validity; and there is conflicting evidence in the relationship of MMT with ICU and hospital LOS (Table 3). There is inconclusive evidence to determine if MMT can predict short- or long-term mortality (Table 3; ESM Table E4).

There is good to excellent intra- and inter-rater reliability for handgrip (ICC range 0.92–0.97) and hand-held dynamometry (ICC range 0.76–0.96). Measurement error for dynamometry was reported to be between 1.9 and 2.8 kg in one study [32]; however, external validation of these findings needs to be undertaken. Construct validity has only been reported for handgrip [30, 31], and is yet to be established for hand-held dynamometry (Table 3; ESM Table E4). Whilst good test performance has been described for the handgrip cut-off values developed for diagnosing ICU-AW [38], no external validation of these values has been undertaken.

Function

Evaluated using 26 different measures (Fig. 1), of which only 12 have been examined in terms of their clinimetric properties (Table 3; ESM Tables E3–E5). Six measures have been specifically developed for use in the ICU setting: CPAx [39, 47], PFIT-s [64], Perme mobility scale [42], ICU mobility scale [44], surgical intensive care unit optimal mobility scale (SOMS) [50] and the functional status score for the intensive care (FSS-ICU) [65]. Excellent reliability has been established for all these measures except the FSS-ICU tool (Table 3; ESM Table E3).

Construct and criterion predictive validity is established for the PFIT-s, CPAx and SOMS (Table 3; ESM Table E3). The PFIT-s tool has been validated in two independent patient settings in two different continents where both patient management and physiotherapy services differ [43, 64]. It had a fair to excellent correlation with the 6-min walk test (6MWT), MRC-SS, and timed-up-and-go test [64]. Additionally, higher awakening PFIT-s were predictive of higher MRC-SS at ICU discharge [43, 64], and discharge to home [64]. The PFIT-s exhibits both floor and ceiling effects of around 20 %, and an MID has been established of 1.5 points out of 10 [2]. The CPAx at ICU discharge was able to discriminate between patient discharge destinations [39]. The SOMS was predictive of in-hospital mortality and had a moderate to good correlation with ICU LOS and handgrip strength [50]. The clinimetric properties of the 6MWT has been examined in one study post-hospital discharge and demonstrated that patients could walk significantly longer on the repeat 6MWT [51]. It is important to note that this

testing was performed in a home-based setting. No clinimetric evaluation of the 6MWT in-hospital has been undertaken to date. There was a moderate to good correlation between 6MWT and the Short Form-36 (SF-36) physical function domain [40, 51]. There is an excellent correlation between 6MWT and timed-up-and-go test at 3 months post-ICU discharge [40]. No criterion predictive validity has been examined for the 6MWT.

The Katz activities of daily living (ADL) was the most widely utilized measure assessing function identified in part 1. The Katz ADL has not been examined in terms of reliability and measurement error specifically within the ICU setting. The Katz ADL has construct validity with SF-36 physical and mental function domains at 1-month [52], but no correlation with the Functional Independence Measure score or 12-month SF-36 scores [52, 54]. The Katz ADL is reported to be predictive of short-term mortality [41, 48, 60] but not longer-term mortality (3–6 months) [41, 53, 61].

Discussion

This systematic review focused on three commonly assessed endpoints used in critical illness: representing body system impairments and functional limitations (-muscle mass, strength and function). Thirty-three different measures were identified; however, only 20 have published clinimetric properties. Ultrasonography, dynamometry, PFIT-s and the CPAx performed the strongest for the measurement instruments for muscle mass, strength and function, respectively.

Based on this review, whilst anthropometry (circumference) is a simple method and easily obtainable, it should not be considered a primary end-point in clinical and research practice. It is not sensitive to change over time, due to other variables such as adiposity, oedema and hydration status affecting circumference measurement, particularly in the ICU setting [21, 66]. Non-ICU studies have demonstrated anthropometry is unreliable and under-represents muscle wasting [67].

Bioimpedance spectroscopy enables bedside quantification of body water and mass compartments including fat-free and fat mass measurements [20]. Whilst prediction equations and algorithms have been developed for some populations [68], it is recommended that raw data be utilized in ICU as no specific reference equation has been developed. There are also challenges with using BIS which need to be taken into consideration such as cost and factors which can affect impedance measurements such as fluid status and ability to obtain accurate height and weight measurements in ICU [20]. The responsiveness of BIS and what constitutes a clinically meaningful change in scores is unknown. However, because it is non-invasive, quick to use and non-volitional, further research is warranted.

The findings of this review indicate that ultrasonography has high responsiveness and excellent intra-rater reliability for measurement of muscle thickness [17] and echogenicity using the Heckmatt approach to quantify muscle echotexture changes [18]. The association between measures of muscle thickness and strength were only fair to good in one study [17]. This is in contrast to the findings in non-ICU studies where ultrasonography was shown to have strong construct validity with measures of strength [69] and has been correlated with architectural changes which occur at a cellular level (as identified by invasive muscle biopsy) [70]. The cause of muscle wasting in ICU is likely multi-factorial. However, it is generally accepted that immobilization and inflammatory stimuli are important contributing factors to the development of ICU-AW [71]. Muscles are adaptive and respond to changes in loading and inflammation in different ways depending on their composition. The response of a specific skeletal muscle will, among other factors, depend on muscle fibre composition and differing contractile properties, which may contribute to specific task and muscle dysfunction. As an example, a study within individuals with COPD demonstrated significant weakness in the quadriceps musculature and preservation of strength in the adductor pollicis muscle [72]. Therefore, it is important that we examine which muscles may be most sensitive in enabling early diagnosis of future functional impairments. There was limited association between strength and thickness measures on ultrasonography. Muscle thickness does not contain information about the neuromuscular conducting properties or dysfunction of the contractile apparatus, and is a two-dimensional representation of muscle size. It is therefore possible that muscle thickness may under-estimate the loss of strength in patients (in contrast to cross-sectional area which may be more sensitive) and may not enable detection of changes in the quality of the muscle and nerve, which can be affected in ICU-AW.

Ultrasonography is demonstrated to have predictive utility for survival in neuromuscular diseases [73]. Although reliability and validity has been demonstrated regardless of expertise level for image acquisition using ultrasonography in a non-ICU study [74], it is important that assessors follow a standardised methodology. A recent study in the ICU demonstrated excellent reliability regardless of expertise level for the analysis of echogenicity [75].

It is important to consider the timing of measurements/treatments particularly when comparing different regimens of muscle preserving interventions. The rate of muscle loss in ICU patients follows a logarithmic curve; as a consequence, patients will experience a higher absolute rate of muscle atrophy in early compared to later phases of their ICU stay [21]. In accordance with this, a delayed measurement may fail to identify the initial muscle loss. This is termed lead-time bias, which may

therefore be an important confounder when examining treatment efficacy. It is important that timing of measurements is reported within future studies. Further research is required to determine if ultrasonography correlates to measures of strength and function and to determine if it has predictive utility in identifying individuals at risk of ICU-AW.

Manual-muscle strength testing is the most commonly utilized measure across the recovery continuum [76]. It has been used both as a diagnostic tool for identifying the presence of ICU-AW and to quantify strength [77]. Whilst excellent inter-rater reliability has been established for overall MRC-SS [10, 34, 35, 62], there is variability in terms of findings for individual muscle groups (poor to excellent reliability) and agreement for the dichotomization of the presence or absence of ICU-AW (slight to substantial in the ICU setting) [10, 34, 62]. Although the majority of studies have demonstrated MRC-SS to be predictive of ICU and hospital mortality [36–38, 63], one study found no relationship between MRC-SS and mortality [10]. The inconsistency in reliability and validity findings between studies may relate to variability in the screening methods used to determine the appropriateness and timing of testing. The task results are dependent on the patient's level of consciousness and mental status. In the ICU setting, many patients are intermittently unable to cooperate because of a reduced level of consciousness/ability to understand due to the critical illness itself or due to the administration of sedative medications. Day-to-day variation may reflect fluctuations in motivation, attention or cognitive dysfunction rather than an increase in muscle dysfunction. This is true of all volitional measurement whether it is in the ICU or in the community. Other reasons for inconsistency in reliability and validity findings include: muscles examined, testing technique (isometric and through range) [77] and the statistical analyses used. To improve measurement accuracy it is important to use standardised phrasing and strong encouragement. Dynamometry has been shown to be a more sensitive method to quantify changes in strength over time particularly once a patient has anti-gravity strength [12]. Normative values have been published for both handgrip and hand-held dynamometry [78, 79]. Handgrip dynamometry is quick, simple and requires minimal training to use, and cut-off values for the diagnosis of ICU-AW have been developed [38]. Further examination of the clinimetric properties are warranted, as well as standardisation of testing methodology including screening to facilitate generalisability across different trials.

Measurement of function is a primary endpoint in many research studies; however, the measures utilised vary. Twenty-six different measures have been used in research trials to date with less than half having one or more established clinimetric properties reported specifically in individuals with critical illness. The PFIT-s and CPAx tools are the most robust function measures with established reliability, validity and responsiveness. However, because of the volitional nature of the tests, there are floor effects in critically ill patients that are greater for the PFIT-s than the CPAx. An MID of 1.5 out of 10 has been established statistically for the PFIT-s tool. An MRC-SS of 41.5 out of 60 had excellent sensitivity and specificity (>80 %) for predicting whether an individual will be able to perform sit-to-stand and marching components of the PFIT-s at ICU discharge [43]. The magnitude of change in muscle performance that represents a clinically meaningful change to the patient has not been calculated in relation to the MID. This is also true for all measures currently published.

The ICU environment is a challenging setting in which to develop a core set of measures to evaluate changes in function. Inflammatory, metabolic and electrolyte changes can all influence muscle function. In critically ill patients, all these parameters are subject to large day-to-day variations. This heterogeneity renders stable study conditions practically impossible and responses unpredictable. This is in conjunction with fluctuations in patient's ability to follow commands and perform volitional testing. Due to this heterogeneity, it is important that all contributing factors to the development of muscle weakness are documented and reported to ensure accurate interpretation of results and increase comparability between future studies.

Adler and colleagues in their systematic review noted that a key endpoint in studies was the time to achieve milestones and the distance ambulated [80]. These measures are not objective and have no established clinimetric properties or evidence of responsiveness over time. It is important that clinicians and researchers use standardized measures to evaluate functional recovery. Data are most commonly being extrapolated from the gerontology or neurological populations. This is evident, e.g., in the use of the Barthel and Functional Independence Measure outcomes both of which have established clinimetric properties in non-critically ill patient populations [81]. There may be key differences particularly in the early stages of the critical illness including: alertness, delirium and sedation that can affect a patient's performance. Further, the ceiling effects of these measures have not been documented in the critically ill.

The ICF framework provides a scaffold in which clinicians and researchers can use outcome measures appropriate for different stages of recovery to capture changes in the patient's level of impairment, activity limitations and participation restrictions. Please refer to Fig. 2 for a potential framework in which measures could be mapped across the continuum from admission to return to home fitting within the ICF framework including suggested measures which warrant investigation in the critically ill such as the de Morton mobility index [82]

which has been utilized in geriatric populations. It is important to consider in the longer-term the patients' ability to achieve a safe community level of ambulation to be able to cross at traffic lights, travel on transportation and to have the physical capacity to perform day-to-day activities such as carrying shopping or walking. This review has focused on the body systems and functional activity limitations outcomes, which can be utilissed to evaluate a patient's level of recovery across the continuum of care. However, it is also important to consider cognitive, mental and psychological outcomes which can be mapped across the continuum, and which are sensitive to detect changes in the patient's recovery.

Limitations

There is the potential for publication bias due to exclusion of non-English articles. There is also the possibility that studies with negative clinimetric findings may not have been published. For risk of bias, the majority of the included studies scored lowest for "inadequate sample size" although they may have statistically justified a smaller sample size than is considered appropriate based on the COSMIN checklist.

Conclusions

Ultrasonography, dynamometry, PFIT-s and CPAx demonstrated the strongest clinimetric properties. Further research into this area, including identification of a core set of measures which can be utilised across the continuum of recovery fitting within the ICF framework. This will enable greater generalizability of findings between studies to determine efficacy of interventions. Furthermore, using the ICF model will direct measurement of tests with similar constructs to be used within each of the classification categories so that the right test is used for the outcome of interest at the most appropriate time-point [40].

Conflicts of interest This research has been undertaken by Ms Parry (primary author) as part of her doctoral qualification with the support of a National Health and Medical Research Council Dora Lush Scholarship (#103923) and previously the Stella Mary Langford scholarship. Pat Cosh grant funding enabled this research to be undertaken. R.K. is currently in receipt of a CR Roper Fellowship and S.B. is a recipient of a National Health and Medical Research Council Fellowship. The remaining authors have disclosed that they do not have any potential conflicts of interest.

References

1. Hough CL (2013) Improving physical function during and after critical care. Curr Opin Crit Care 19(5):488–495
2. Denehy L, Skinner H, Edbrooke L, Haines K, Warrillow S, Hawthorne G, Gough K, Vander Hoorn S, Morris M, Berney S (2013) Exercise rehabilitation for patients with critical illness: a randomized controlled trial with 12 months follow up. Crit Care 17(4):R156
3. 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 310(15):1591–1600
4. Iwashyna T (2012) Trajectories of recovery and dysfunction after acute illness, with implications for clinical trial design. Am J Respir Crit Care Med 186(4):302–304
5. World Health Organisation (2001) International Classification of Functioning, Disability and Health. WHO, Geneva
6. Sherwood L (2004) Human physiology—from cells to systems, 5th edn. Thomson Brooks/Cole Learning, Belmont
7. de Vet H, Terwee C, Mokkink L, Knol D (2011) Measurement in medicine. Cambridge University Press, Cambridge
8. Portney L, Watkins M (2009) Foundations of clinical research, applications to practice, 3rd edn. Appleton and Lange, Connecticut
9. Guyatt G et al (2002) Methods to explain the clinical significance of health status measures. Mayo Clin Proc 77(4):371–383
10. Connolly B, Jones G, Curtis A, Murphy P, Douiri A, Hopkinson N, Polkey M, Moxham J, Hart N (2013) Clinical predictive value of manual muscle strength testing during critical illness: an observational cohort study. Crit Care 17(5):R229
11. Fitzpatrick R, Davey C, Buxton M, Jones D (1998) Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 2(14):1–74
12. Vanpee G, Hermans G, Segers J, Gosselink R (2014) Assessment of limb muscle strength in critically ill patients: a systematic review. Crit Care Med 42(3):701–711
13. Tipping C, Young P, Romero L, Saxena M, Dulhunty J, Hodgson C (2012) A systematic review of measurements of physical function in critically ill adults. Crit Care Resusc 14:302–311
14. Terwee C, Jansma E, Riphagen I, de Vet H (2009) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 18(8):1115–1123
15. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, Bouter L, de Vet H (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 19:539–549
16. Sim J, Wright C (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 85(3):257–268
17. Baldwin C, Bersten A (2014) Alterations in respiratory and limb muscle strength and size in patients with sepsis who are mechanically ventilated. Phys Ther 94(1):68–82

18. Grimm A, Teschner U, Porzelius C, Ludewig K, Zielske J, Witte O, Brunkhorst F, Axer H (2013) Muscle ultrasound for early assessment of critical illness neuromyopathy in severe sepsis. Crit Care 17(5):R227
19. Savalle M, Gillaizeau F, Maruani G, Puymirat E, Bellenfant F, Houillier P, Fagon J, Faisy C (2012) Assessment of body cell mass at bedside in critically ill patients. Am J Physiol Endocrinol Metab 303(3):E389–E396
20. Baldwin C, Paratz J, Bersten A (2012) Body composition analysis in critically ill survivors: a comparison of bioelectrical impedance spectroscopy devices. J Parenter Enteral Nutr 36(3):306–315
21. Gruther W, Benesch T, Zorn C, Paternostro-Sluga T, Quittan M, Fialka-Moser V, Spiss C, Kainberger F, Crevenna R (2008) Muscle wasting in intensive care patients: ultrasound observation of the M. quadriceps femoris muscle layer. J Rehabil Med 40(3):185–189
22. Reid C, Campbell I, Little R (2004) Muscle wasting and energy balance in critical illness. Clin Nutr 23(2):273–280
23. Faisy C, Rabbat A, Kouchakji B, Laaban J (2000) Bioelectrical impedance analysis in estimating nutritional status and outcome of patients with chronic obstructive pulmonary disease and acute respiratory failure. Intensive Care Med 26(5):518–525
24. Frankenfield D, Cooney R, Smith J, Rowe W (1999) Bioelectrical impedance plethysmographic analysis of body composition in critically injured and healthy subjects. Am J Clin Nutr 69(3):426–431
25. Morais A, Comarella A, Pitanga K, Faintuch J (1998) Interest of conventional clinical, biochemical and bioimpedance measurements as indicators of mortality risk in critical patients. Rev Hosp Clin 53(4):176–180
26. Phang P, Aeberhardt L (1996) Effect of nutritional support on routine nutrition assessment parameters and body composition in intensive care unit patients. Can J Surg 39(3):212–219
27. Campbell I, Watt T, Withers D, England R, Sukumar S, Keegan MA, Faragher B, Martin DF (1995) Muscle thickness, measured with ultrasound, may be an indicator of lean tissue wasting in multiple organ failure in the presence of edema. Am J Clin Nutr 62(3):533–539
28. Robert S, Zarowitz BJ, Hyzy R, Eichenhorn M, Peterson EL, Popovich J (1993) Bioelectrical impedance assessment of nutritional status in critically ill patients. Am J Clin Nutr 57(6):840–844
29. Yosef-Brauner O, Adi N, Shahar T, Yehezkel E, Carmeli E (2014) Effect of physical therapy on muscle strength, respiratory muscles and functional parameters in patients with intensive care unit-acquired weakness. Clin Respir J 9(1):1–6. doi. 10.1111/crj.12091
30. Fan E, Dowdy D, Colantuoni E, Mendez-Tellez P, Sevransky J, Shanholtz C, Himmelfarb C, Desai S, Ciesla N, Herridge M et al (2014) Physical complications in acute lung injury survivors: a 2-year longitudinal prospective study. Crit Care Med 42(4):849–859
31. Lee CM, Fan E (2012) ICU-acquired weakness: what is preventing its rehabilitation in critically ill patients? BMC Med 10:115
32. Baldwin C, Paratz J, Bersten A (2013) Muscle strength assessment in critically ill patients with handheld dynamometry: an investigation of reliability, minimal detectable change, and time to peak force generation. J Crit Care 28(1):77–86
33. Vanpee G, Segers J, Van Mechelen H, Wouters P, Van den Berghe G, Hermans G, Gosselink R (2011) The interobserver agreement of handheld dynamometry for muscle strength assessment in critically ill patients. Crit Care Med 39(8):1929–1934
34. Hough C, Lieu B, Caldwell E (2011) Manual muscle strength testing of critically ill patients: feasibility and interobserver agreement. Crit Care 15(1):R43
35. Fan E, Ciesla N, Truong A, Bhoopathi V, Zeger S, Needham D (2010) Inter-rater reliability of manual muscle strength testing in ICU survivors and simulated patients. Intensive Care Med 36(6):1038–1043
36. Brunello A, Haenggi M, Wigger O, Porta F, Takala J, Jakob S (2010) Usefulness of a clinical diagnosis of ICU-acquired paresis to predict outcome in patients with SIRS and acute respiratory failure. Intensive Care Med 36(1):66–74
37. Sharshar T, Bastuji-Garin S, Stevens R, Durand M, Malissin I, Rodriguez P, Cerf C, Outin H, De Jonghe B (2009) Presence and severity of intensive care unit-acquired paresis at time of awakening are associated with increased intensive care unit and hospital mortality. Crit Care Med 37(12):3047–3053
38. Ali N, O'Brien J, Hoffmann S, Phillips G, Garland A, Finley J, Almoosa K, Hejal R, Wolf K, Lemeshow S et al (2008) Acquired weakness, handgrip strength, and mortality in critically ill patients. Am J Respir Crit Care Med 178(3):261–268
39. Corner E, Soni N, Handy H, Brett S (2014) Construct validity of the Chelsea critical care physical assessment tool: an observational study of recovery from critical illness. Crit Care 18:R55
40. Denehy L, Nordon-Craft A, Edbrooke L, Malone D, Berney S, Schenkman M, Moss M (2014) Outcome measures report different aspects of patient function three months following critical care. Intensive Care Med. doi: 10.1007/s00134-3513-3
41. Tripathy S, Mishra J, Dash S (2014) The geriatric critically ill patient in the developing world—mortality and functional outcome at one year: a prospective single centre study. J Crit Care 29(3):474.e7–474.e13
42. Nawa R, Lettvin C, Winkelman C, Evora P, Perme C (2014) Initial inter-rater reliability for a novel measure of patient mobility in a cardiovascular ICU. J Crit Care 29(3):475.e1–475.e5
43. Nordon-Craft A, Schenkman M, Edbrooke L, Malone D, Moss M, Denehy L (2014) The physical function intensive care test: implementation in survivors of critical illness. Phys Ther 94(10):1–9
44. Hodgson C, Needham D, Haines K, Bailey M, Ward A, Harrold M, Young P, Zanni J, Buhr H, Higgins A et al (2014) Feasibility and inter-rater reliability of the ICU mobility scale. Heart Lung 43(1):19–24
45. Baldwin M, Reid M, Westlake A, Rowe J, Granieri E, Wunsch H, Dam T, Rabinowitz D, Goldstein N, Maurer M et al (2014) The feasibility of measuring frailty to predict disability and mortality in older medical intensive care unit survivors. J Crit Care 29(3):401–408
46. Denehy L, Edbrooke L, Nordon Craft A, Berney S, Malone D, Schenkman M, Moss M (2013) An investigation of physical performance based tests and patient self report physical function in survivors of intensive care (ICU). American Thoracic Society, Philadelphia
47. Corner E, Wood H, Englebretsen C, Thomas A, Grant R, 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(1):33–41
48. Abd-El-Gawad W, Adly N, Salem H (2013) Diagnostic accuracy of activities of daily living in prediction of community-acquired pneumonia outcomes in elderly patients admitted to intensive care units. J Clin Gerontol Geriatr 4(4):123–127

49. Thrush A, Rozek M, Dekerlegand J (2012) 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 92(12):1536–1545
50. Kasotakis G, Schmidt U, Perry D, Grosse-Sundrup M, Benjamin J, Ryan C, Tully S, Hirschberg R, Waak K, Velmahos G et al (2012) The surgical intensive care unit optimal mobility score predicts mortality and length of stay. Crit Care Med 40(4):1122–1128
51. Alison J, Kenny P, King M, McKinley S, Aitken L, Leslie G, Elliott D (2012) Repeatability of the six-minute walk test and relation to physical function in survivors of a critical illness. Phys Ther 92(12):1556–1563
52. Vest M, Murphy T, Araujo K, Pisani M (2011) Disability in activities of daily living, depression, and quality of life among older medical ICU survivors: a prospective cohort study. Health Qual Life Outcomes 9:9
53. Daubin C, Chevalier S, Seguin A, Gaillard C, Valette X, Prevost F, Terzi N, Ramakers M, Parienti J, du Cheyron D et al (2011) Predictors of mortality and short-term physical and cognitive dependence in critically ill persons 75 years and older: a prospective cohort study. Health Qual Life Outcomes 9:35
54. Clini E, Crisafulli E, Antoni F, Beneventi C, Trianni L, Costi S, Fabbri L, Nava S (2011) Functional recovery following physical training in tracheotomized and chronically ventilated patients. Respir Care 56(3):306–313
55. Skinner E, Berney S, Warrillow S, Denehy L (2009) Development of a physical function outcome measure (PFIT) and a pilot exercise training protocol for use in intensive care. Crit Care Resusc 11:110–115
56. Sacanella E, Pérez-Castejón J, Nicolás J, Masanés F, Navarro M, Castro P, López-Soto A (2009) Mortality in healthy elderly patients after ICU admission. Intensive Care Med 35(3):550–555
57. van der Schaaf M, Dettling D, Beelen A, Lucas C, Dongelmans D, Nollet F (2008) Poor functional status immediately after discharge from an intensive care unit. Disabil Rehabil 30(23):1812–1818
58. Swafford B (2008) Validity of Kansas University Hospital physical therapy acute care functional outcomes tool. Acute Care Perspect (Fall):14–18
59. Chiang L, Wang L, Wu C, Wu H, Wu Y (2006) Effects of physical training on functional status in patients with prolonged mechanical ventilation. Phys Ther 86(9):1271–1281
60. Bo M, Massaia M, Raspo S, Bosco F, Cena P, Molaschi M, Fabris F (2003) Predictive factors of in-hospital mortality in older patients admitted to a medical intensive care unit. J Am Geriatr Soc 51(4):529–533
61. Dardaine V, Dequin P, Ripault H, Constans T, Giniès G (2001) Outcome of older patients requiring ventilatory support in intensive care: impact of nutritional status. J Am Geriatr Soc 49(5):564–570
62. Hermans G, Clerckx B, Vanhullebusch T, Segers J, Vanpee G, Robbeets C, Casaer M, Wouters P, Gosselink R, Van Den Berghe G (2012) Interobserver agreement of Medical Research Council sum-score and handgrip strength in the intensive care unit. Muscle Nerve 45(1):18–25
63. Lee J, Waak K, Grosse-Sundrup M, Xue F, Lee J, Chipman D, Ryan C, Bittner E, Schmidt U, Eikermann M (2012) Global muscle strength but not grip strength predicts mortality and length of stay in a general population in a surgical intensive care unit. Phys Ther 92(12):1546–1555
64. Denehy L, de Morton N, Skinner E, Edbrooke L, Haines K, Warrillow S, Berney S (2013) A physical function test for use in the intensive care unit: validity, responsiveness, and predictive utility of the physical function ICU test (scored). Phys Ther 93(12):1636–1645
65. Zanni J, Korupolu R, Fan E, Pradhan P, Janjua K, Palme J, Brower R, Needham D (2010) Rehabilitation therapy and outcomes in acute respiratory failure: an observational pilot project. J Crit Care 25:254–262
66. Green CJ, Campbell IT, McClelland P, Hutton JL, Ahmed MM, Helliwell TR, Wilkes RG, Gilbertson AA, Bone JM (1995) Energy and nitrogen balance and changes in midupper-arm circumference with multiple organ failure. Nutrition 11(6):739–746
67. Young A, Hughes A, Russell P, Parker M, Nichols P (1980) Measurement of quadriceps muscle wasting by ultrasonography. Rheumatol Rehabil XIX(3):141–148
68. Earthman C, Traughber D, Dobratz J, Howell W (2007) Bioimpedance spectroscopy for clinical assessment of fluid distribution and body cell mass. Nutr Clin Pract 22:389–405
69. Strasser E, Draskovits T, Praschak M, Quittan M, Graf A (2013) Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. Age 35:2377–2388
70. Reimers K, Reimers CD, Wagner S, Paetzke I, Pongratz DE (1993) Skeletal muscle sonography: a correlative study of echogenicity and morphology. J Ultrasound Med 12(2):73–77
71. Batt J, Dos Santos CC, Cameron JI, Herridge MS (2013) Intensive care unit-acquired weakness clinical phenotypes and molecular mechanisms. Am J Respir Crit Care Med 187(3):238–246
72. Man W, Soliman M, Nikoletou D, Harris M, Rafferty G, Mustfa N, Polkey M, Moxham J (2003) Non-volitional assessment of skeletal muscle strength in patients with chronic obstructive pulmonary disease. Thorax 58:665–669
73. Arts I, Overeem S, Pillen S, Schelhaas H, Zwarts M (2010) Muscle ultrasonography to predict survival in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 82:552–554
74. Cartwright M, Demar S, Griffin L, Balakrishnan N, Harris J, Walker F (2013) Validity and reliability of nerve and muscle ultrasound. Muscle Nerve 47(4):515–521
75. Sarwal A, Parry S, Berry M, Hsu F-C, Lewis M, Justus N, Morris P, Denehy L, Berney S, Dhar S et al (2014) Inter-observer reliability of quantitative muscle ultrasound analysis in the critically ill population. J Ultrasound Med (in press)
76. Fan E, Cheek F, Chlan L, Gosselink R, Hart N, Herridge M, Hopkins R, Hough C, Kress J, Zochodne D et al (2014) An official American Thoracic Society clinical practice guideline: the diagnosis of intensive care unit-acquired weakness in adults. Am J Respir Crit Care Med 190(12):1437–1446
77. Elliott D, Denehy L, Berney S, Alison J (2011) Assessing physical function and activity for survivors of a critical illness: a review of instruments. Aust Crit Care 24(3):155–166
78. Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S (1985) Grip and pinch strength: normative data for adults. Arch Phys Med Rehabil 66(2):69–74
79. Bohannon RW (1997) Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil 78:26–32

80. Adler J, Malone D (2012) Early mobilization in the intensive care unit: a systematic review. Cardiopulm Phys Ther J 23(1):5–13

81. van der Putten J, Hobart J, Freeman J, Thompson A (1999) Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel index and the Functional Independence Measure. J Neurol Neurosurg Psychiatry 66(4):480–484

82. de Morton N, Brusco N, Wood L, Lawler K, Taylor N (2011) The de Morton mobility index (DEMMI) provides a valid measure for measuring and monitoring the mobility of patients making the transition from hospital to the community: an observational study. J Physiother 57(2):109–116