Intensive care unit acquired weakness: measuring recovery from critical illness

EJ Corner

Critical illness can lead to muscle wasting, functional decline and long-term disability. This is well described subjectively in the literature; however, to date, objective measures have proved elusive. This narrative review aims to explore the objective measurement tools available to assess intensive care unit-acquired weakness, to review the psychometric properties of these tools and to show current practice and the obstacles we face in the development of new measurement scales.

Keywords: rehabilitation; critical illness; physiotherapy; assessment tool; outcome assessment

“Evidence, which we have means to strengthen for or against a proposition, is our proper means for attaining truth.”
Florence Nightingale

Introduction

Intensive care unit-acquired weakness (ICU-AW) is a common consequence of critical illness, affecting between 25% and 100% of critical care patients. ICCU-AW is a syndrome encompassing myopathic, neuropathic and atrophic changes, which clinically presents as profound weakness. Its presence can hinder and dictate the course of recovery, leading to protracted weaning from mechanical ventilation, long hospital stays, permanent disability and a reduction in health-related quality of life. Mortality is also higher in those with a diagnosis of ICU-AW; however this may demonstrate a link between ICU-AW and severity of illness, rather than ICU-AW being a direct cause of mortality. Bed rest and immobility, neuromuscular blockade, corticosteroids, poor glycaemic control, hypoalbuminaemia, and sepsis have all been linked with the development of ICU-AW. However, the exact physiological mechanisms are poorly understood.

In March 2009, the UK National Institute for Health and Clinical Excellence (NICE) introduced clinical guideline 83 (CG83) entitled ‘Rehabilitation after Critical Illness.’ This guideline focuses on the optimisation of recovery from critical illness and makes sensible recommendations about the management of ICU-AW. These comprise a cycle of clinical assessments to determine the risk of morbidity, followed by the implementation of patient-specific rehabilitation programmes and goals, administered throughout the critical care stay and up to three months after hospital discharge.

NICE CG83 represents important progress in the recognition and management of ICU-AW and allows a focus beyond just survival to optimal recovery. Regrettably, a crucial finding from the guideline development process is the dearth both of evidence supporting physical rehabilitation strategies and validated assessment tools to screen for, assess and monitor the presence of ICU-AW. This makes uptake and implementation of CG83 in a standardised manner difficult or erratic at best. Appleton et al completed a national survey of rehabilitation in Scottish Intensive Care Units in 2011 to which 96% of lead clinicians and 100% of lead physiotherapists responded. Of these, only 14% of lead clinicians and 30% of physiotherapists had read the guidelines and only 66% of those who read them found them useful. Although the authors do not explore the reasons behind this, they suggest that the lack of validated assessment tools and lack of evidence-based rehabilitation practice is influential in the uptake of NICE CG83.

This narrative review aims to explore the psychometric properties and clinical applicability of the current measurement systems used in the literature to assess for ICU-AW; to see how the psychometric properties of these assessment tools, or lack thereof, impacts upon the validity of interventional studies; to explore what new measures are being developed; and finally, to discuss the obstacles associated with standardising assessment of ICU-AW in a heterogeneous population.

Psychometric testing of measurement scales

There are three key psychometric properties that need to be evaluated when developing any form of assessment tool. These are:

- validity – is it measuring what it is supposed to?
- reliability – can it be administered consistently?
- responsiveness – is it able to reflect clinically meaningful change?

Validity can be broken down further into criterion (concurrent), face, content, construct and predictive validity.
Criterion validity is the comparison of the new score to an established gold standard. Face and content validity are similar, and answer the questions of whether the tool looks right and whether it includes the right components, weighted appropriately. Construct validity is the development of hypotheses that the new scale will compare to measures of the same construct or behave as expected in specific populations. Predictive validity is a tool’s ability to predict a defined outcome.10-12

It is essential that any scoring system, regardless of purpose, is both reliable and valid. If a measure cannot be administered consistently and does not accurately reflect what it is intended to measure, then its results are not only meaningless but even potentially dangerous. Measurement systems that are not fit for purpose will increase the likelihood of errors in interventional studies and paint inaccurate clinical pictures, both of which may have a negative impact on patient management. Responsiveness to change as a psychometric property is more ambiguous. Some argue that this needs independent evaluation; however, if a measure is both valid and reliable then by definition it should also be responsive to change.11

**Problems with established measurement systems**

There is a range of rehabilitation tools available to measure physical disability. The Rehabilitation Measures Database summarises the available assessment tools for use in rehabilitation research and clinical practice, listing a total of 104 measures.12 Almost all of these measures are problematic when applied to the critical care environment. Many of the available tools lack specificity, as none of them were designed for the critical care population. Most of the current measurement tools available have not been tested for validity or reliability, and those that have, either do not possess it or have insufficient data.10 The severity of disability associated with ICU-AW means that many of these measures have a significant ‘floor’ effect, i.e., they are unable to detect clinically meaningful change in patients with lower levels of function. Finally, even if the psychometric properties are reasonable, the measurement tools are often complex and time-consuming to administer.

To explore this in more detail, some of the most commonly adopted measurement tools used in critical care have been reviewed. These are the Medical Research Council (MRC) sumscore, the Rivermead Motor Index (RMI), the Functional Independence Measure (FIM) and the Barthel Index (BI).

Muscle testing using the MRC sumscore is a common diagnostic and assessment tool for ICU-AW, involving the manual grading of muscle strength in 12 of the major muscle groups.13 The MRC score can be a useful tool and is widely used by physiotherapists in all specialist areas. However, this test requires the patient to be co-operative, which is not always achievable, especially when the patient is acutely unwell. Furthermore, components of the MRC score lack objectivity, and it is a poor prognosticator for hospital length of stay and mortality, therefore lacking in predictive validity.14

The RMI was developed for, and psychometrically scrutinised in, the stroke population.15-17 It has demonstrated good reliability and validity in this patient group. It is a quick and useful test for mobile stroke patients and comprises 15 tasks, based on a Guttman scale. The lowest level on this Guttman scale is the ability to roll independently. If a patient can do this they score 1. The next level is the ability to sit on the edge of the bed from supine independently, giving a score of 2. This progresses to a total score of 15, which represents a patient able to run 10 metres without a limp in 4 seconds. Critically ill patients suffering with ICU-AW who are unable to breathe independently would barely get onto this scale. This is not a criticism of the RMI per se, but a caution in its use out of the context in which it was derived.

The FIM was designed for the more general assessment of disability, therefore is potentially more transferable to the critical care setting. The FIM grades a patient on 18 motor and cognitive tasks, giving a score ranging from 18-126. Its psychometric properties have been tested in a variety of patient groups including stroke patients, those in general rehabilitation and in patients with multiple sclerosis. It demonstrates good reliability and validity.12 However it has not been fully evaluated in the critical care population. Furthermore, the FIM takes 45 minutes to administer, requires training to complete and needs input from the whole multidisciplinary team. This may be feasible in a rehabilitation centre, but not in a busy ICU.

The BI is a comprehensive assessment tool assessing independence in 10 activities of daily living (ADLs) such as feeding, dressing and ambulation. The score for each component ranges from either 0-5, 0-10 or 0-15. Reliability is good and validity has been shown in stroke patients with correlation with the FIM, however testing is inconclusive in the critical care population.10,18,19 The BI is quick to administer and is therefore one of the favoured measures for use in the critical care environment. The main concern with the BI is its floor effect. An observational study20 of disability in patients at discharge from ICU found a median score of 2 on the modified 20 point BI, while van der Schaaf and colleagues21 showed that 40% of patients in the week following discharge from critical care were rated as completely dependent. While this might highlight the degree of disability associated with ICU-AW, it also indicates that, as a measure, the BI lacks responsiveness to clinically meaningful change in patients on the ICU.

The current status of these assessments was summarised by Skinner and colleagues using a survey of physiotherapy practice and prescription of exercise in 167 Australian adult ICUs.22 They concluded that, ‘No validated functional outcome measures were used, highlighting that further research is required to enable adequate evaluation of exercise prescription and rehabilitation in ICUs.’

**Impact on interventional research**

Interventional studies require both a well defined intervention and a well defined measure of outcome that is able to discern change and hence quantify the impact of that intervention. In the context of the critical care rehabilitation literature, interventions are usually aimed at improving function, so a valid and reliable tool should be able to detect potentially small changes in functionality. Without such a measure, the quality of these studies is destined to be substandard.
Chaing et al.\textsuperscript{23} completed a randomised controlled trial (RCT) looking at the impact of regular whole body exercise in 39 patients undergoing prolonged mechanical ventilation compared to a control group receiving no exercise. Primary outcome measures were muscle strength, BI and FIM. The results found that the intervention group had greater muscle strength and function at six weeks (p<0.05) and that the control group suffered a decline in muscle strength (p<0.05). An insignificant decline in function (measured by the BI) was also shown in the control group; however, as the BI lacks responsiveness in patients with low levels of function, this may represent a type 2 error. Chaing acknowledged the lack of responsiveness of the BI and the complex administration of the FIM, and favoured the FIM for future research.

Zanotti and colleagues\textsuperscript{24} demonstrated that the use of neuromuscular electrical stimulation (NMES) in weaning respiratory patients showed potential benefits. Patients were randomised to lower limb NMES with active limb mobilisation (ALM) or to ALM alone (control). There was a greater increase in muscle strength in the intervention group (p=0.02) and a reduction in time taken to transfer from bed to chair (10.75 days versus 14.3 days, p=0.0001). One problem with the study was that 'transfer' was not defined precisely, i.e., whether it was with assistance or independently, making this outcome measure subjective, especially as there was no mention of assessor blinding.

Schweickert et al.\textsuperscript{25} undertook a single-blind RCT in the United States investigating early mobilisation in combination with sedation holds versus normal US practice (i.e., sedation holds and mobilisation on prescription). Their outcome measures included functional status quantified by the FIM, unassisted distance walked and number of functionally independent activities of daily living.

They found that a higher proportion of participants returned to functional independence at discharge in the intervention group (p=0.02), which suggests a causal relationship. The pre-admission functional status was established with the BI and return to function was quantified using the FIM. Quantifying function with two different measures at the two time points, neither of which is validated in this population, makes conclusions difficult. The BI was assessed at discharge and a significant difference between groups was found (p=0.05). Its rejection as the primary outcome measure possibly reflects the authors' lack of confidence in the sensitivity of the BI. Furthermore, there was no significant difference in the number of functionally independent ADLs between groups either at hospital or ICU discharge (p=0.15 and p=0.6 respectively); the results are therefore conflicting.

These studies illustrate that the lack of a validated assessment tool is problematic in clinical practice, but also in research. There is a clear and urgent need for a simple and holistic bedside measurement system that is easily taught, responsive to change, reproducible, valid and easily intelligible across all clinical interfaces. This need has been recognised and has prompted researchers to begin to develop a number of critical care-specific measurement tools.

### Development of new critical care specific measurement tools

Skinner et al.\textsuperscript{26} have been developing the Physical Function ICU Test (PFIT) for patients in the ICU. This was produced in a centre for weaning patients with tracheostomies and includes activities such as marching on the spot. While the PFIT has shown responsiveness and reliability, it was not tested for validity and was a small study (n=13). The test focuses on cardiovascular fitness rather than physical function and hence is a somewhat one-dimensional assessment. There are a number of other key components of physical function that need to be addressed in order to holistically reflect physical ability. While it was constructed and trialled in a tertiary ICU, its responsiveness in the bedbound, acutely ill ICU patient is uncertain.

The University of Rochester Acute Care Evaluation (URACE) is also in development.\textsuperscript{27} This is a scoring system that grades independence with bed mobility, transfers (bed to chair), locomotion and stairs. However, as yet it has not been tested for reliability and validity and no patient data has yet been published. It has also been developed with input from clinicians at a single centre and thus its generalisability needs evaluation.

The Association of Chartered Physiotherapists in Respiratory Care list on their website a number of basic scoring systems which have been put together in an effort to address this problem.\textsuperscript{28} These include the Manchester Mobility Scale and the Modified Clinical Outcome Variable Scale (MCOVS).

The Manchester Mobility Score is a simple scoring system that monitors a patient's maximal level of activity on a daily basis.\textsuperscript{29} It is a basic scoring system that undoubtedly would be quick to complete, however the degree of insight into the patient's actual level of function that it provides appears limited and to date there is no published research describing its development or validation.

Published data exists looking at the validity of the Clinical Outcome Variable Scale (COVS), however as yet there is no published research on the modified version.\textsuperscript{29} The COVS has demonstrated reliability and validity in the stroke and spinal injury population, but not the general critical care population and it takes up to 45 minutes to administer.\textsuperscript{12,20}

The Chelsea Critical Care Physical Assessment tool (CPAx) is also a contender.\textsuperscript{30} The CPAx has been developed by the author of this review. It is a numerical and pictorial scoring system that pulls together 10 commonly assessed components of physical function, identified by a focus group of lead critical care physiotherapists. Each component is graded on a Guttman scale from 0-5, giving a total score out of 50. Proof of concept work has demonstrated construct validity and inter-rater reliability in an observational study of 33 patients, and ongoing research suggests predictive validity for hospital outcome. However all of these new scoring systems are in their infancy and although the data is promising, validation is difficult and much more research is essential.

Against this background, it is useful to consider what hinders the development of a simple objective marker of recovery from ICU-AW.
Obstacles in the development of a critical care scoring system

The first paper highlighting the safety issues that should be considered when mobilising the critically ill was by Stiller in 2007. So the concept of early rehabilitation in critical care is a relatively new one and possibly reflects a reduction in mortality rates and an increasing emphasis on long-term quality of life, not just survival. It is no surprise therefore, that the science behind rehabilitation in the critically ill population is in its infancy and consequently, so is the development of a measuring system to monitor physical recovery in the rehabilitation stage.

Although the need has been established, there are still significant obstacles to the development of a rehabilitation measurement tool. Heterogeneity of patients can make categorisation of patients and research more difficult, often requiring larger sample sizes and limiting the generalisability of research results. The lack of a gold standard comparator makes validation difficult and necessitates comparison with scoring systems which may be substandard measures. Acuity of illness and cognitive issues can make the assessment of physical function complicated.

The translation of current rehabilitation measures to an intensive care environment is complex due to the specific needs of patients with ICU-AW. The degree of potentially reversible global weakness seen in ICU-AW is only ever seen elsewhere in Guillain-Barré patients, and means that most current tools are not sensitive enough to detect improvement in physical function in the early stages. Those tools that may be useful, such as the FIM are often complex and hence, their clinical utility reduced. Furthermore, in the UK, the availability of research training and funding available for physiotherapists to carry out research is limited.

Summary and recommendations

Rehabilitation in critical illness is a relatively new concept that is under-researched and consequently lacks a sufficient evidence base. Much of the research that is available is methodologically flawed by the lack of standardised outcome measures. The translation of existing measures into the intensive care environment may be useful; however it is likely that the specific needs of intensive care patients will be missed. Furthermore, the floor effect may result in a lack of responsiveness to clinically meaningful change in the early stages of rehabilitation. Current research looking at the development of scoring systems to address this issue is promising but in its infancy. Additionally, their development is made difficult by the heterogeneity of intensive care patients, lack of gold standard comparators and small patient numbers.

In order to address this problem, a united approach is essential. If scale developers work together, widespread data collection from a large patient population will be possible, which will facilitate development of a generalisable measurement tool with strong psychometric properties. It is essential that any measurement system is simple to understand and use in order to increase its utility. It must demonstrate reliability and validity, which may be best achieved through looking at predictive validity for functional outcome, instead of through comparison to other substandard measures. Ideally such a measure should have a pictorial representation, to allow ease of interpretation and it should facilitate communication between both professionals and service users. In the words of Florence Nightingale:

“The figures must be as clear as a picture – they must tell a story as clearly as does a picture of the crucifixion.”

Declaration

No conflicts of interest are declared and no financial support was received for this manuscript.

References


Evelyn J Corner Senior Respiratory Physiotherapist, Senior NIHR CLAHRC Fellow, Chelsea and Westminster NHS Foundation Trust evelyn.corner@chelseawest.nhs.uk