SUPPLEMENTARY MATERIAL 1 – UNABRIDGED METHODOLOGY

Trial Design

We conducted a randomised (1:1), parallel group, allocation-concealed, assessor blinded, controlled trial in mixed medical-surgical intensive care units (ICUs) at two hospitals in Australia and in both medical and surgical ICUs at two hospitals in USA. Participants, or as required, the legally authorized representative for the patient were approached to obtain written informed consent. Continuation of informed consent was obtained from the participant once they were capable of comprehending the study. All study procedures were consistent with the Good Clinical Practice Guidelines and the Declaration of Helsinki for the protection of human subjects. This study was overseen by an independent data monitoring committee. The Human Research Ethics Committees/Institutional Review Boards approved the trial protocol at all participating sites (Austin Health Human Research Ethics Committee (HREC/15/76) Australian sites; John Hopkins University Human Subjects Research Institutional Review Board (NA_00093358) and Duke University Institutional Review Board Pro0051021). The trial was initially registered in Australia in May 2012 (ACTRN 12612000528853) as a single site trial. A successful application for peer-reviewed government research funding was made in 2014 that allowed international multicentre involvement and follow-up of participants after hospital discharge. The trial was subsequently registered in Clinicaltrials.gov (NCT02214823) in August 2014.

Inclusion/Exclusion criteria

Patients were eligible for inclusion in the trial if they were mechanically ventilated in an ICU and at least 18 years of age, with sepsis or severe sepsis, expected to require more than 48 hours of mechanical ventilation, and expected to remain in the ICU for a minimum of four days following randomisation. Patients were excluded if they did not meet safety criteria to commence exercise within 72 hours of meeting inclusion criteria; had a primary neurological diagnosis; had a lower limb amputation or malignancies; were unable to perform outcome measures due to pre-morbid physical, intellectual or cognitive impairment or were non English speaking; were not expected to survive ICU; were pregnant; had a body mass index > 40; had an external fixation device or superficial metal in the lower limb; had open wounds or skin abrasions at electrode application points; had a pacemaker without an underlying rhythm and were transferred from another ICU after >2 days of consecutive mechanical ventilation. Participants were not considered eligible for cognitive outcome assessment if they had a score of >3.3 on the Informant Questionnaire on Cognitive Decline in the Elderly (IQ CODE) at baseline prior to randomisation in ICU, obtained from a surrogate; a score of $\geq$10 on the Alcohol Use Disorders Identification Test assessment at baseline or had no fixed address.

Berney S, *et al. Thorax* 2021; 76:656–663. doi: 10.1136/thoraxjnl-2020-215093

We expanded our inclusion criteria to participants with sepsis, severe sepsis or systemic inflammatory response syndrome (SIRS) to improve recruitment and the likelihood of participant survival to primary outcome time points of hospital discharge and six-month follow-up. This change was made on 7th August 2014 when 44 participants had been recruited. On 23rd February 2015, when 52 participants had been recruited, we included a battery of neuro-psychological tests, patients reported outcomes and test of psychological function with follow-up to 12 months. On 4th March 2016, we expanded our criteria to include participants receiving extra corporeal membrane oxygenation (ECMO) for more than 48 hours in lieu of mechanical ventilation, this change did not result in additional participants.

The period of enrolment was from 19th July 2012 to 3rd September 2017 with a 12 month follow up which ended 29th September 2018.

Intervention and control group

Participants were recruited in ICU by a Research Assistants associated with the trial as soon as feasible after meeting inclusion criteria with the aim of the intervention commencing as soon as was feasible subject to meeting intervention safety guidelines (eSupplement Table 1). Individual participants allocated to the intervention group received up to 60 mins/day of intervention per day. Within the intervention group patients, 1 leg was randomly allocated to receive FES cycling with the other leg receiving cycling alone (without FES). The intervention had a goal of being delivered for at least five out of seven day per week. If an intervention session was less than 10 minutes it was considered incomplete and not included in analysis. The intervention was provided in addition to usual care rehabilitation by registered Physiotherapists, experienced in critical care rehabilitation, who were not blinded to the randomisation group. Either the RT300 (Restorative Therapies, Baltimore USA) or a Hasomed RehabStim FES attached to a MOTO Med cycle (RECK-Technik, Germany) were used to deliver the intervention. Face to face and online training (https://restorative-therapies.com/for-clinicians/critical-care-and-icu/clinician-training/) was provided by Restorative Therapies in the use of the RT300 at three sites, the other site was already experienced in using the MOTO Med cycle, with face to face and remote training provided as needed. Trial staff made individual visits to all sites providing training and orientation to the study protocol. Electrical stimulation was provided to the following muscles: rectus femoris; hamstrings; gluteals and gastrocnemius using large size (20cm) electrodes. Electrodes were placed on both legs although muscle stimulation was only provided to the leg randomised to receive FES. The cycling only leg received sham electrical stimulation. Stimulation (milliAmps) was set at 20-30 for all stimulated muscles and was titrated to achieve muscle contraction and patient comfort. Pulse width (microseconds) was set to 250 for average sized legs and was increased to 300 if the leg was oedematous. Frequency (Hertz) was set at 43.5 and increased as

Berney S, *et al. Thorax* 2021; 76:656–663. doi: 10.1136/thoraxjnl-2020-215093

appropriate to 50 to induce a strong muscle contraction. The control speed was set to 35 revolutions/minute. The intervention physiotherapist recorded if a muscle contraction was achieved during each session for every muscle group. FES-cycling was stopped immediately if any of the ceasing criteria were met (eSupp Table 1). Once participants' status returned to initial safety criteria to commence exercise, FES cycling recommenced and continued until target time was achieved or until the patient did not tolerate the intervention (e.g. request to stop intervention). Patients received FES cycling until discharge from ICU or until up to 28 days (that included a minimum of 20 intervention sessions), whichever came first. Following this time point, physiotherapy intervention was left to the discretion of the clinical team.

Adherence to the trial protocol was assessed using a standardised form at monthly online meetings at start up and subsequently by a random review of 10 cases by trial personnel and site visit reviews of documentation. All data were reviewed by the Chief Investigator and if concerning trends were identified they were raised directly with the site investigator. Adherence data were also presented at an annual trial management meeting.

Usual care occurred in accordance to unit protocols by experienced critical care physiotherapists. The aim of usual care was to maximise physical function (sitting over the side of the bed, marching in place, standing and walking). Treatment duration and intensity were titrated to individual patients at the discretion of the treating physiotherapist. Usual care was reported for each patient while in the ICU and mapped to the ICU Mobility Scale (IMS)[1].

Both arms of the study recorded time spent actively engaging in either FES cycling or usual care rehabilitation. This did not include time for set up, periods of rest or recovery, any passive activity (e.g. passive range of movement or passive transfer to a chair) or cleaning of equipment after the rehabilitation session finished.

Following discharge from ICU, both groups received usual physiotherapy on the hospital ward that aimed to facilitate discharge and maximise strength, function and independence. This usual care physiotherapy also was mapped to the IMS.

Trial outcomes

Primary Outcome – Quadriceps strength

The primary outcome was quadriceps muscle strength measured using a hand held dynamometer at hospital discharge. Participants were measured in supine with the test leg flexed over a standardised foam roll with the other leg resting straight on the bed. The force plate of the hand held dynamometer was applied perpendicular to the tibia just above the talocrural joint. A make test protocol was

Berney S, *et al. Thorax* 2021; 76:656–663. doi: 10.1136/thoraxjnl-2020-215093

employed to achieve isometric knee extension contraction. The examiner held the dynamometer stationary while the subject exerted maximal force against it. A break test was not used as these give higher values as it incorporates eccentric contraction. Participants were instructed to straighten their knee but not to move maximally against the force plate and examiner. Standardised instructions were given "I want you to straighten/kick up build it up (1 second), build it up (2seconds), harder, harder, harder (5 seconds), and relax. The examiner matched the resistance of the participant (2 seconds to generate a maximum force and then 4 seconds as hard as possible). Participants were given one practice try against the examiners hand and then one practice against the dynamometer. Participants were given 30-60 seconds rest between each repetition. The best of three attempts was used for analysis. Muscle force was measured in Newtons and multiplied by individual participants leg length (Tibilae Mediale to Sphyrion) measured in metres to achieve maximal isometric torque (MIT) measured in Newton metres (Nm). Leg length was measured in triplicate and then averaged. For analysis purposes, in both the intervention and control groups the highest of three strength measurements was used for each leg (i.e. for intervention, the leg randomized to FES-cycling and leg randomized to cycling without FES and in control the highest of three measures irrespective of leg)

Primary Outcome - Cognitive Impairment

A battery of validated and standardized performance tests of the most relevant cognitive domains previously used in survivors of critical illness [2] was performed, with the primary outcome being the presence of cognitive impairment at six months. This was defined as having either one cognitive test within the battery with a "score at least 2 standard deviations (SDs) below population norms (i.e., bottom 2.5%) or at least two tests with a score equal to or greater than 1.5 SDs below norms (i.e., bottom 6.7% for both tests) [3]. The following cognitive domains were evaluated with the test battery: executive function, evaluated via the Hayling Sentence Completion Test scaled score (range, 1 to 10; higher is better) [4]; language, evaluated via a test of verbal fluency using letters F, A, S, total score (higher score is better – mean score used in analysis) [5]; immediate and delayed memory, evaluated via the Logical Memory I and Logical Memory II age-adjusted scaled scores (range, 1 to 19; higher is better) from the Wechsler Memory Scale-Third Edition [6, 7]; verbal reasoning and concept formation, evaluated via the Similarities age-adjusted scaled score (range, 1 to 19; higher is better) from the Wechsler Adult Intelligence Scale-Third Edition [7]; and attention and working memory evaluated via the Digit Span age-adjusted scaled score (range, 1 to 19; higher is better) from the Wechsler Adult Intelligence Scale-Third Edition [6, 7]" [8].

Secondary outcomes

Berney S, *et al. Thorax* 2021; 76:656–663. doi: 10.1136/thoraxjnl-2020-215093

Secondary outcomes were as follows: Incidence and duration of delirium; muscle strength: muscle force (N) measured using a hand held dynamometer (HHD); manual muscle testing using the Medical Research Council sum score (range 0-60; <48 indicating intensive care acquired weakness (ICU-AW)[9, 10]; hand grip strength measured using hand grip dynamometry [11]; physical function: physical function in ICU test scored PFITs (range 0-10 higher scored indicating better function) [12]; the Functional Status Score for the ICU (FSS-ICU) (range from 0-35 higher scores indicating better physical function) [13]; the short physical performance battery (SPPB) (range 0-12 higher scores indicating better function)[14] and six minute walk test (6MWT) measured in metres[15].

Self-reported function was measured using the Katz Index of Independence in Activities of Daily Living (ADLs) and the Lawton's Instrumental Activities of Daily Living (IADLs). The Katz ADL index scores range from 2 or less - severe functional impairment, 4 - moderate impairment to 6 - full independence in bathing, dressing, toileting, transferring to a chair, continence and eating[16]. The Lawton IADL scale ranges from 0 (low function, dependent) to 8 (high function, independent)[16].

Symptoms of anxiety and depression were screened using the Hospital Anxiety and Depression scale (HADS) (the two subscale scores range from 0-21 with higher scores indicating greater psychological distress)[17]. The presence of post-traumatic stress disorder was screened using the Impact of Event Scale - Revised (IES-R) which has 22 items rated on a four-point scale. Items were averaged to generate a total mean score with a threshold score of ≥1.6 defining clinically significant symptoms[18]. Health related quality of life (HRQoL) was measured using the 36 item Short Form Health survey questionnaire (SF-36 version 2) and the EQ-5D-5L. The SF36v2 consists of 36 items generating a health profile of eight subscale scores aggregated into two summary scores: Physical Component Summary (PCS) and Mental Component Summary (MCS). Scores range from 0 to 100 for each of the 8 subscales, and can be scored using standardized values (average of 50, standard deviation of 10) which are compared to norm based scores. Higher scores indicate better quality of life[19]. The SF36v2 is commonly used in ICU survivors with acceptability, reliability, and validity in this population[20]. The EQ-5D-5L consist of a descriptive system that covers five dimensions: mobility, self care, usual activities, pain/discomfort, and anxiety/depression. The visual analogue scale asks respondents to score their overall health from 0 (worst imaginable health state to 100 (best imaginable health state)[21].

Ultrasound imaging

The cross-sectional area of the quadriceps muscle was measured in supine using two-dimensional ultrasound (US) Mylab 25Gold Esaote Italy. The rectus femoris was imaged with the Esoate LA523, 12 MHz linear probe at a 5cm depth or Sono-Site M-MSK using a HFL-50 15-6MHZ probe. Participants lay

supine with their knees in extension and muscles relaxed with legs in neutral rotation. A water-soluble transmission gel was applied to the ultrasound (US) head to allow acoustic contact without depressing the dermal surface. The scanning head was applied perpendicular to the long axis of the thigh on its anterior surface two thirds of the distance from the anterior superior iliac spine to the superior border of the patella as previously reported.[22] The point was marked for consistency of location of the sound head. The gain, compression and songraphic device setting were kept constant between participants. Three separate images of the rectus femoris were captured. The cross-sectional sonographic view was selected as it has previously been shown to be more sensitive to changes in muscle echotexture in a critically ill population.[23] Images were analysed using Image-J software (National Institute of Health, Bethesda, MD, USA).

Sample size

A sample size of 40 participants per group was calculated based on published data[24] from a prior ICU RCT of cycling in the ICU demonstrating a mean difference in quadriceps strength of 20 Newton (N), a within-group SD = 31 N assuming, and assuming 80% power and $\alpha<0.05$. For the cognitive impairment outcome, a sample size of 46 participants per group was calculated based a between-group difference in proportions of 0.26 (0.36 in control and 0.10 in intervention) [8] evaluated using the same test battery. Attrition from mortality and loss to follow-up over 6-months was estimated at 40%[25] yielding a final sample size of 154 (72 participants per group).

Randomisation and Blinding

Participants were randomized using a random number generator in a 1:1 ratio using permuted random block allocation to either intervention or usual care. Randomisation was stratified by hospital site. Computer generator randomization lists were prepared by a Research Assistant (RA) unrelated to this trial in random block sizes. The RA prepared up to four sets of sealed, opaque and sequentially numbered envelopes for each site. When each patient was enrolled in the study, an Administrative assistant, unrelated to the study opened the envelope that was next in sequence.

Safety

Serious adverse events were defined as any untoward clinical events that were fatal or immediately life threatening; permanently disabling or severely incapacitating; required prolonged inpatient hospitalization or important medical events that may not have resulted in death or be life threatening, or require prolonged hospitalization may be considered serious adverse events when, based on appropriate medical judgment, they jeopardized the patient and required medical or surgical intervention to prevent one of the serious outcomes listed above.

Berney S, *et al. Thorax* 2021; 76:656–663. doi: 10.1136/thoraxjnl-2020-215093

References

1. Hodgson CL, Berney S, Haines K, Harrold M, Young P, Buhr H, Needham DM, (2013) Development of a mobility scale for use in a multicentre Australia and New Zealand: Trial of Early Activity and Mobilisation in ICU. American Journal of Respiratory & Critical Care Medicine 187: A1323
2. Christie JD, Biester RC, Taichman DB, Shull WH, Jr., Hansen-Flaschen J, Shea JA, Hopkins RO, (2006) Formation and validation of a telephone battery to assess cognitive function in acute respiratory distress syndrome survivors. J Crit Care 21: 125-132
3. Hopkins RO, Weaver LK, Collingridge D, Parkinson RB, Chan KJ, Orme JF, Jr., (2005) Two-Year Cognitive, Emotional, and Quality-of-Life Outcomes in Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med 171: 340-347
4. Burgess PW, Shallice T (2007) The Hayling and Brixton Tests. Thames Valley Test Company, London
5. Lezak MD (1995) Neuropsychological assessment. Oxford University Press, New York
6. Heaton RK, Miller SW, Taylor JR, Grant I (2004) Comprehensive norms for an expanded Halstead–Reitan Battery: demographically adjusted neuropsychological norms for African American and Caucasian adults. In: Editor (ed)^(eds) Book Comprehensive norms for an expanded Halstead–Reitan Battery: demographically adjusted neuropsychological norms for African American and Caucasian adults., City, pp.
7. Wechsler D (1997) Wechsler Memory Scale. Psychology Corporatio San Antonio Texas
8. Needham DM, Dinglas VD, Morris PE, Jackson JC, Hough CL, Mendez-Tellez PA, Wozniak AW, Colantuoni E, Ely EW, Rice TW, Hopkins RO, Network NNA, (2013) Physical and Cognitive Performance of Patients with Acute Lung Injury 1 Year after Initial Trophic versus Full Enteral Feeding. EDEN Trial Follow-up. Am J Respir Crit Care Med 188: 567-576
9. De Jonghe B, Sharshar T, Lefaucheur J, Authier F, Durand-Zaleski I, Boussarsar M, Cerf C, Renaud E, Mesrati F, Carlet J, (2002) Paresis Acquired in the Intensive Care Unit: A Prospective Multicenter Study. The Journal of the American Medical Association 288: 2859-2867
10. Fan E, Ciesla ND, Truong AD, Bhoopathi V, Zeger SL, Needham DM, (2010) Inter-rater reliability of manual muscle strength testing in ICU survivors and simulated patients. Intensive Care Med 36: 1038-1043
11. Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S, (1985) Grip and pinch strength: normative data for adults. Archives of Physical Medicine and Rehabilitation 66: 69-74
12. Denehy L, de Morton NA, Skinner EH, 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
13. Huang M, Chan KS, Zanni JM, Parry SM, Neto SG, Neto JA, da Silva VZ, Kho ME, Needham DM, (2016) Functional Status Score for the ICU: An International Clinimetric Analysis of Validity, Responsiveness, and Minimal Important Difference. Crit Care Med 44: e1155-e1164
14. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB, (1994) A short performance physical battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. Journal of Gerontology 49: M85-M94
15. Holland AE, Spruit MA, Troosters T, Puhan MA, Pepin V, Saey D, McCormack MC, Carlin BW, Sciurba FC, Pitta F, Wanger J, MacIntyre N, Kaminsky DA, Culver BH, Revill SM, Hernandes NA, Andrianopoulos V, Camillo CA, Mitchell KE, Lee AL, Hill CJ, Singh SJ, (2014) An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. Eur Respir J 44: 1428-1446

16. Applegate W, Blass J, Williams T, (1990) Instruments for the functional assessment of older patients. New England Journal of Medicine 322: 1207-1214
17. Bjelland I, Dahl AA, Haug TT, Neckelmann D, (2002) The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 52: 69-77
18. Bienvenu OJ, Williams JB, Yang A, Hopkins RO, Needham DM, (2013) Posttraumatic stress disorder in survivors of acute lung injury: evaluating the Impact of Event Scale-Revised. Chest 144: 24-31
19. Ware J, Kosinski M, Dewey J (2000) How to score Version 2 of the SF-36 Health Survey. Quality Metric Inc, Lincoln
20. Chrispin PS, Scotton H, Rogers J, Lloyd D, Ridley SA, (1997) Short form 36 in the intensive care unit: assessment of acceptability, reliability and validity of the questionnaire. Anaesthesia 52: 15-23
21. Group EQ, (1990) EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. Health Policy 16: 199-208
22. Seymour JM, Ward K, Sidhu PS, Puthucheary Z, Steier J, Jolley CJ, Rafferty G, Polkey MI, Moxham J, (2009) Ultrasound measurement of rectus femoris cross-sectional area and the relationship with quadriceps strength in COPD. Thorax 64: 418-423
23. Cartwright MS, Kwayisi G, Griffin LP, Sarwal A, Walker FO, Harris JM, Berry MJ, Chahal PS, Morris P, (2013) Quantitative neuromuscular ultrasound in the intensive care unit. Muscle Nerve 47: 255-299
24. Burtin C, Clerckx B, Robbeets C, Ferdinande P, Langer D, Troosters T, Hermans G, Decramer M, Gosselink R, (2009) Early exercise in critically ill patients enhances short-term functional recovery. Crit Care Med 37
25. Denehy L, Skinner EH, Edbrooke L, Haines K, Warrillow S, Hawthorne G, Gough K, Hoorn SV, Morris ME, Berney S, (2013) Exercise rehabilitation for patients with critical illness: a randomized controlled trial with 12 months of follow-up. Crit Care 17