

Core Outcome Measures for Clinical Research in Acute Respiratory Failure Survivors

An International Modified Delphi Consensus Study

Dale M. Needham^{1,2,3}, Kristin A. Sepulveda^{1,2}, Victor D. Dinglas^{1,2}, Caroline M. Chessare^{1,2}, Lisa Aronson Friedman^{1,2}, Clifton O. Bingham III^{4,5}, and Alison E. Turnbull^{1,2,6}

¹Outcomes After Critical Illness and Surgery Group, ²Division of Pulmonary and Critical Care Medicine, School of Medicine, ³Department of Physical Medicine and Rehabilitation, School of Medicine, ⁴Division of Rheumatology, School of Medicine, ⁵Division of Allergy and Clinical Immunology, School of Medicine, and ⁶Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

Abstract

Rationale: Research evaluating acute respiratory failure (ARF) survivors' outcomes after hospital discharge has substantial heterogeneity in terms of the measurement instruments used, creating barriers to synthesizing study data.

Objectives: To identify a minimum set of core outcome measures that are essential to include in all clinical research studies evaluating ARF survivors after discharge.

Methods: We conducted a three-round modified Delphi consensus process with 77 participants (47% female, 55% outside the United States), including clinical researchers from more than 16 countries across six continents, patients/caregivers, clinicians, and research funders. Participants reviewed standardized information on measure instruments for seven consensus-derived outcomes plus one recommended outcome.

Measurements and Main Results: Response rates were 91 to 97% across the three rounds. Among 75 measurement instruments evaluated, the following met *a priori* consensus criteria: EQ-5D and

36-item Short Form Health Survey version 2 (optional) for the "satisfaction with life and personal enjoyment" and "pain" outcomes, and both the Hospital Anxiety and Depression Scale and the Impact of Events Scale-Revised for the "mental health" outcome. No measures reached consensus for the following outcomes: cognition, muscle and/or nerve function, physical function, and pulmonary function. All measures considered for pulmonary function met consensus criteria for exclusion. The following measures did not reach the threshold for consensus but achieved the highest scores for their respective outcomes: the Montreal Cognitive Assessment (cognition), manual muscle testing and handgrip dynamometry (muscle and/or nerve function), and 6-minute-walk test (physical function).

Conclusions: This Core Outcome Measurement Set is recommended for use in all clinical research evaluating ARF survivors after hospital discharge. In the future, researchers should evaluate measures for outcomes not reaching consensus.

Keywords: patient outcome assessment; follow-up studies; Core Outcome Measurement Set; clinical trials; intensive care

(Received in original form February 16, 2017; accepted in final form May 15, 2017)

Supported by NHLBI grant R24 HL111895, Patient-Centered Outcomes Research Institute grant SC14-11402-10918 (C.O.B.), and National Institute of Arthritis and Musculoskeletal and Skin Diseases grant P30-AR053503 (C.O.B.).

Author Contributions: D.M.N.: had full access to all data in the study, takes full responsibility for the integrity of the data and the accuracy of the data analysis, and supervised the study; K.A.S., C.M.C., V.D.D., and D.M.N.: contributed to the acquisition of data; L.A.F.: conducted the statistical analysis; D.M.N., K.A.S., and A.E.T.: drafted the article; all authors: developed the study concept and design, interpreted the data, provided critical revisions for important intellectual content, and read and approved the final manuscript.

Correspondence and requests for reprints should be addressed to Dale M. Needham, F.C.P.A., M.D., Ph.D., Division of Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, 1830 East Monument Street, 5th Floor, Baltimore, MD 21205. E-mail: dale.needham@jhmi.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 196, Iss 9, pp 1122–1130, Nov 1, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201702-0372OC on May 24, 2017

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: There is substantial heterogeneity in the outcome measures used within the rapidly expanding research literature on acute respiratory failure survivors' outcomes after hospital discharge. This heterogeneity limits the field's ability to synthesize or interpret study findings and contributes to potential bias resulting from selective reporting of study results.

What This Study Adds to the

Field: In this study, we present a consensus-based set of core outcome measures that are recommended for use in all clinical research studies in which researchers choose to evaluate acute respiratory failure survivors after hospital discharge. The set of core outcome measures was created using a rigorous, modified Delphi consensus process incorporating the perspectives of an international panel of clinical researchers, clinicians, patients/caregivers, and U.S. federal funding organization members for clinical research in the field.

The NHLBI, along with the American Thoracic Society, the Society of Critical Care Medicine, and the Multisociety Task Force for Critical Care Research, recommends giving priority to research evaluating the outcomes of intensive care unit (ICU) survivors after hospital discharge (1–7). Consistent with such recommendations, researchers in a growing number of studies are evaluating ICU survivors' outcomes after discharge, with more than 300 original research articles published since 2000 (8). This rapid growth in research publications has made comparing, synthesizing, and interpreting these results increasingly challenging (9, 10), with a recent scoping review demonstrating the use of 250 unique measurement instruments across 425 publications in the field (8).

One relatively new methodological approach to addressing this issue is the creation of a Core Outcome Set along with an accompanying set of core measurement instruments. A Core Outcome Set is a minimum collection of outcomes reported

in all studies within a specific field (11, 12). Importantly, this approach does not prevent researchers from evaluating additional outcomes; however, it serves as the minimum standard to ensure that essential outcomes within a given field are consistently assessed using the same measurement instruments in all studies. This consistency facilitates comparisons and meta-analyses and may prevent bias resulting from selective outcome reporting (13, 14). After identifying essential outcomes for evaluation in all studies in a Core Outcome Set project, relevant stakeholders participate in a systematic and deliberate process to identify measurement instruments that (1) evaluate these essential outcomes, (2) have suitable measurement properties, and (3) are feasible for use. In this article, we refer to the resulting list of measurement instruments as a *Core Outcome Measurement Set*.

There are currently no other Core Outcome Measurement Set projects for acute respiratory failure (ARF) survivorship research registered with the Core Outcome Measures in Effectiveness Trials (COMET) Initiative (www.cometinitiative.org). Hence, our objective was to develop a Core Outcome Measurement Set for clinical research aimed at evaluating patient outcomes after hospital discharge among survivors of ARF, including acute respiratory distress syndrome, using a rigorous consensus methodology and an international panel of relevant stakeholders.

Methods

We conducted a three-round modified Delphi consensus process to identify a minimum set of measurement instruments for assessing a recommended Core Outcome Set (15). *Core outcomes* are defined as patient outcomes, health-related conditions, or aspects of health that are essential to evaluate in all studies within a specific clinical field (12). The following eight core outcomes for ARF survivors are recommended (15): survival, physical function, mental health, pulmonary function, pain, muscle and/or nerve function, cognition, and satisfaction with life or “personal enjoyment” (the term used to represent the concept of “health-related quality of life” when originally establishing the Core Outcome Set for ARF survivors [15]). The first seven outcomes in the

preceding list met *a priori* consensus criteria for inclusion in the Core Outcome Set (15), whereas “satisfaction with life or personal enjoyment” represented an eighth recommended core outcome because it was very close to reaching the *a priori* consensus threshold for inclusion in the Core Outcome Set (i.e., the threshold for inclusion was 70%, and this outcome achieved 69% [15]) and because there was strong and consistent support for its inclusion based on our prior pilot work conducted in preparation for this Delphi process (16) and in prior meetings of three different expert groups (2–4). The survival outcome was deemed not to need a consensus-based process for selecting a relevant measurement instrument and was not included in the present consensus process. Hence, panel members voted on measurement instruments for a total of seven outcomes.

The modified Delphi consensus methodology uses expert opinion to address questions for which empirical data are unavailable or inadequate (15). This method, used extensively in Core Outcome Set-related projects (17), involves participants' completing serial Internet-based surveys, referred to as *rounds*, related to the study question. Essential aspects of the Delphi approach include (1) recruiting a panel of informed, expert participants; (2) maintaining the anonymity of participants to ensure that voting occurs without intimidation or powerful participants disproportionately influencing results; and (3) providing a summary of voting results after each round so that each participant can compare his or her responses with those of the other participants before voting in the subsequent round. We report this research in accordance with current recommendations for creating core outcome sets via the Delphi process (12, 18). The complete study protocol for the Improving Long-Term Outcomes Research for Acute Respiratory Failure project is available from www.improvelto.com. The project was registered with the COMET Initiative (www.comet-initiative.org/studies/details/360) and funded by the NHLBI (grant R24 HL111895; see www.improvelto.com).

Recruitment of the Delphi Panel

A single Delphi panel was created to establish the Core Outcome Set (15) and perform the current task of establishing the associated measurement instruments. In

establishing this panel, we aimed to recruit a diverse group of participants encompassing four stakeholder groups relevant to this area: (1) critical care clinical researchers, (2) clinicians caring for critical care patients/survivors, (3) ICU survivors or caregivers of ICU survivors, and (4) U.S. federal research funding organizations that fund clinical research in this area. To avoid limiting representation of key stakeholder groups, the pool of clinical researchers, patients, and caregivers was not restricted to ARF; however, it was made clear to all panel members throughout the consensus process that the Core Outcome Measurement Set is intended for research studies specifically evaluating ARF survivors. Given that the end users of the Core Outcome Measurement Set are clinical researchers, an international recruitment strategy was used that included 1 representative from all 21 member groups of the International Forum for Acute Care Trialists organization, which represent more than 16 different countries across six continents. Additional individual clinical researchers were selected by random sampling from an existing database of corresponding authors (8, 15) to obtain two researchers with self-reported clinical research expertise in each of the following areas: physical, cognitive, and mental health outcomes. Finally, we purposefully invited nine clinical researchers who have published internationally recognized research on outcomes of ARF survivors.

To enroll representatives of clinicians as well as patients and caregivers, we recruited from the top four English-speaking countries in a scoping review of ICU survivorship research: the United States, the United Kingdom, Australia, and Canada (8). Details regarding this recruitment process have been reported previously (15), with invitation e-mails being sent that explained that survey completion would serve as informed consent. The Qualtrics (Provo, UT) online survey platform was used to collect demographic information about panel members, and DelphiManager software (COMET Initiative, Liverpool, UK) was customized for use in this project. The institutional review board of Johns Hopkins University approved this study.

Generating a Preliminary List of Outcome Measures

To prepare a preliminary list of outcome measures for the first Delphi round, we

selected up to five of the most commonly used instruments by referring to a scoping review conducted for this Delphi process (8) and for six of the recommended core outcomes (15) (no measures for the “pain” core outcome were reported in the scoping review [8]). Standardized “measure cards” written in nontechnical language (available from www.improvelto.com/instruments) were created for each outcome measure included in this preliminary list. These cards were carefully developed, pilot tested, and iteratively revised using input from clinicians, patients, and caregivers who were not members of the Delphi panel. Measure cards included information such as number of survey items, estimated time needed to complete, administration mode (e.g., patient or proxy or both), scoring information, need for specialized training for administration or scoring, licensing or purchasing information, cost, required equipment, number of times used in prior ICU survivorship research (8), and published measurement properties (e.g., validity, reliability) in ICU survivors. The Consensus-based Standards for the selection of health Measurement Instruments checklist was used to rate a study’s evaluation of measurement properties (19). Panel members were also provided with easy-to-understand descriptions of the measurement properties described on the measure cards. When available, the measure card included hyperlinks to online examples of instruments or videos demonstrating performance of tests.

Modified Delphi Methodology

Before starting each round of the Delphi process, participants were reminded of the goal of the consensus project (i.e., use for clinical research after hospital discharge of ARF survivors), the definition of a Core Outcome Measurement Set, and the *a priori* criteria for consensus (see below). Panel member support for each outcome measure was rated using the Grading of Recommendations Assessment, Development and Evaluation (or “GRADE”) scale (20), which is a 9-point scale commonly divided into three categories for Core Outcome Set–related projects: Not Important (score, 1–3), Important but Not Critical (score, 4–6), and Critical (score, 7–9). In addition, panel members were provided with an “Unable to Score” response option and were instructed

to use this response if they did not feel comfortable rating specific measures. Consensus for an instrument to be a part of the Core Outcome Measurement Set, was defined *a priori* as at least 70% of all respondents rating the measure as “Critical” (i.e., score of ≥ 7) and less than or equal to 15% of respondents rating the measure as “Not Important” (i.e., score of ≤ 3). This consensus definition, which has been used in other Delphi studies (21–23), ensured that a measure could not achieve consensus if a minority stakeholder group (i.e., patients/caregivers or clinicians) commonly rated it as “Not Important.” Anonymity of panel members was maintained throughout the three-round Delphi process, which commenced on May 2, 2016, and finished on October 10, 2016.

Round 1. Prior to voting, participants were asked to review: (1) measure cards (as described above) for each measurement instrument, (2) a scoping review on ICU survivorship research and related measurement instruments (8), and (3) descriptions of psychometric measurement properties. Panel members rated the importance of each of 38 preliminary outcome measures for 6 recommended core outcomes, and they suggested measures for the “Pain” core outcome. Panel members were explicitly asked to consider the appropriateness (i.e., measurement properties) and feasibility (i.e., ease of use, cost, and other requirements) of the measurement instrument in their voting. The survey solicited suggestions for other potential measures missing from the preliminary list provided.

Round 2. The same documents provided in round 1 were provided in round 2 along with relevant panel member comments from round 1 and measure cards for new measures suggested during round 1. Using feedback from round 1, we revised the name of the cognitive subdomain of “Intelligence” to “Intelligence/Cognitive Screening” in response to suggestions regarding more explicit inclusion of general cognitive screening instruments. Panel members were provided tables that displayed the percentage of panel members who rated each measure as either “Not Critical” or “Critical” for inclusion, aggregated across all round 1 participants as well as stratified by stakeholder group. Participants were shown their own round 1 score, the percentage distribution of votes

for each score across the 9-point scale, and the number of panel members who scored the outcome.

Round 3. The documents (excluding the scoping review [8]) and the scoring results and comments, as previously described for round 2, were provided in round 3. The survey asked participants to rerate outcome measures that had not met consensus criteria for inclusion or exclusion. Pain measures suggested during

round 1 had been rated only once; therefore, all were included for a second round of rating in round 3.

Statistical Reporting

Response rates were defined as the proportion of recruited panel members who completed each survey. Survey responses were summarized with descriptive statistics using SAS version 9.4 software (2013; SAS Institute, Cary, NC).

Results

The expert panel included a total of 77 representatives from the four stakeholder groups, comprised of 35 (45%) clinical researchers, 19 (25%) clinicians and representatives of clinician professional associations, 19 (25%) patients/caregivers, and 4 (5%) representatives of U.S. federal research funding organizations (Table 1; see also Table E1 in the online supplement).

Table 1. Characteristics of Panel Members

Characteristic	All Panel Members* (n = 77)	Clinical Researchers (n = 35)	Clinicians/Professional Associations (n = 19)	U.S. Federal Research Funding Organizations (n = 4)	Patients and Caregivers† (n = 19)
Male sex, n (%)	41 (53%)	24 (69%)	7 (37%)	1 (25%)	9 (47%)
Age, yr, n (%)					
25–44	31 (40%)	14 (40%)	7 (37%)	1 (25%)	9 (47%)
45–64	43 (56%)	21 (60%)	12 (63%)	3 (75%)	7 (37%)
≥65	3 (4%)	0 (0%)	0 (0%)	0 (0%)	3 (16%)
Country of residence, n (%)					
United States	35 (45%)	8 (23%)	10 (53%)	4 (100%)	13 (68%)
Canada, United Kingdom, or Australia	28 (37%)	13 (37%)	9 (47%)	0 (0%)	6 (32%)
Other‡	14 (18%)	14 (40%)	0 (0%)	0 (0%)	0 (0%)
Years of education, median (IQR)	20 (18–22)	21 (19–22.5)	20 (19–21)	22 (21–22)	16 (15–18)
Clinical work: type of training§, n (%)					
Physician: critical care	25 (57%)	19 (83%)	5 (28%)	1 (100%)	0 (0%)
Physical, occupational, or respiratory therapist and/or speech-language pathologist	12 (27%)	5 (22%)	7 (39%)	0 (0%)	0 (0%)
Nurse or nurse practitioner	6 (14%)	0 (0%)	4 (22%)	0 (0%)	2 (11%)
Physician: physical medicine and rehabilitation	2 (5%)	0 (0%)	2 (11%)	0 (0%)	0 (0%)
Other clinical training	4 (9%)	3 (13%)	1 (5%)	0 (0%)	0 (0%)
Years of professional experience¶, median (IQR)	14.5 (9–21)	13 (9.5–19.5)	17 (9.5–21)	18.5¶¶	N/A
Area of professional expertise**, n (%)					
Physical health and functioning	42 (55%)	27 (77%)	13 (68%)	1 (25%)	N/A
Mental health	17 (22%)	12 (34%)	4 (21%)	0 (0%)	N/A
Cognitive function	17 (22%)	11 (31%)	6 (32%)	0 (0%)	N/A
Other	8 (10%)	4 (11%)	3 (16%)	1 (25%)	N/A
None	8 (10%)	3 (9%)	3 (16%)	1 (25%)	N/A
Professional interest in critical illness**, n (%)					
Research: clinical	51 (66%)	35 (100%)	15 (79%)	1 (25%)	0 (0%)
Research: basic or translational	16 (21%)	11 (31%)	4 (21%)	1 (25%)	0 (0%)
Clinical work	44 (57%)	23 (66%)	18 (95%)	1 (25%)	2 (11%)
None of the above	19 (25%)	0 (0%)	0 (0%)	2 (50%)	17 (89%)

Definition of abbreviations: IQR = interquartile range; N/A = not applicable.

*One panel member represented both the clinical researcher and clinician/professional association groups; the total number of respondents was 76.

†A patient/caregiver was replaced by another patient/caregiver member after round 2. Data from both panel members are presented (patients, n = 10; caregivers, n = 9).

‡Representation from other countries: Singapore = 1, China = 1, France = 2, Germany = 1, Belgium = 1, Greece = 1, the Netherlands = 1, Norway = 1, Italy = 1, Ireland = 1, Brazil = 1, Panama = 1.

§A total of 44 (57%) panel members selected clinical training (23 clinical researchers, 18 clinicians/professional associations, 1 U.S. federal research funding organization, and 2 patients/caregivers), 5 of whom selected two types of clinical work. Other clinical training includes anesthesiology (n = 2), internal medicine (n = 1), and pharmacy (n = 1).

¶All panel members from the clinical researcher, clinician/professional association, and U.S. federal research funding organization groups (n = 58) provided data.

¶¶Two funding body representatives responded and reported 14 and 23 years of professional experience, respectively.

**Panel members could select more than one response.

There were 42 (55%) panel members from outside the United States, and 36 (47%) participants were female. The median professional experience level (excluding the patient and caregiver group) was 14.5 years (interquartile range, 9–21).

Across the Delphi rounds, there were 75 unique panel members available for voting (1 unique panel member was a representative for two stakeholder groups, and 1 patient/caregiver could not continue and was replaced, with both of these members included in the description of the 77 representatives in Table 1). In round 1, of the 75 unique panel members, 73 (97%) responded (Figure 1). Among the 38 outcome measures provided, none met consensus criteria for inclusion in the Core Outcome Measurement Set (Table E2). All measures were retained for voting during round 2. Panel members suggested 37 additional measures for consideration. In round 2, 68 (91%) of the 75 unique panel members responded (Figure 1). Of 75 outcome measures evaluated, a total of 50 measures met exclusion criteria, with no measure remaining for the “Pulmonary Function” core outcome (Table E3). In round 3, 72 (96%) of the 75 unique panel members responded (Figure 1). Among the remaining 22 outcome measures rated in round 3, a total of 13 measures met exclusion criteria, and 6 measures did not meet inclusion or exclusion criteria in the “Physical Function,” “Muscle and/or Nerve Function,” and “Cognition” core outcomes (Table E4). A list of all measurement instruments considered in this Delphi process, including the final inclusion/exclusion status for the Core Outcome Measurement Set, is provided in Table E5.

Results of the consensus process and related recommendations are provided in Table 2. For the “Mental Health” core outcome, two measures reached consensus: Hospital Anxiety and Depression Scale (with separate subscales for anxiety and depression symptoms) and Impact of Events Scale–Revised (evaluating post-traumatic stress disorder symptoms). These instruments have commonly been used in critical care clinical research, with some evaluation of their measurement properties (8, 19, 24–28) and recent recommendations for their use in clinical practice (29). For the “Pain” core outcome, panel members suggested eight existing instruments. By a large margin, consensus was reached for

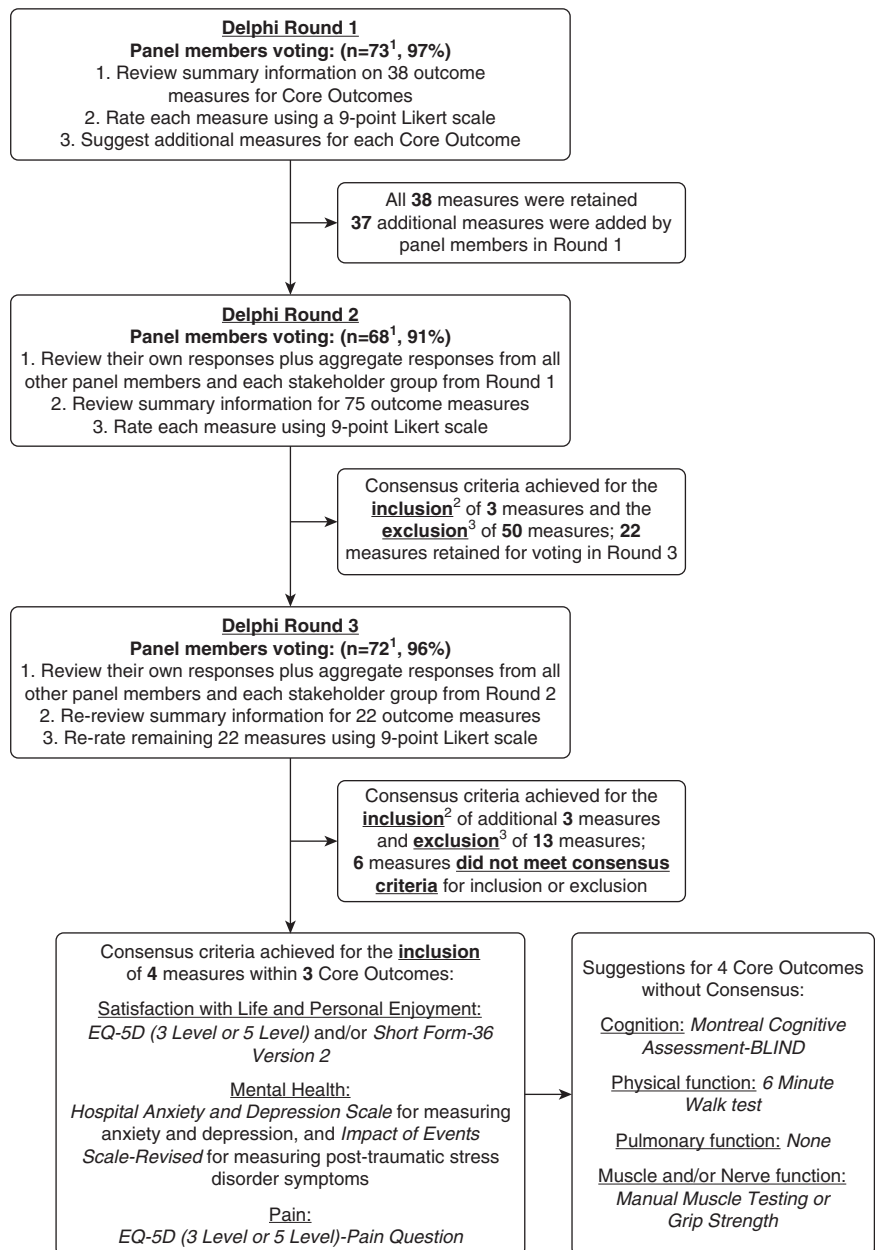


Figure 1. Flow diagram for modified Delphi consensus process. ¹Of 76 experts invited, 2 patients/caregivers declined to participate, and one was replaced. Of 75 participants, there was a nonresponse from 1 clinical researcher and 1 patient/caregiver in round 1; 2 clinical researchers and 5 patients/caregivers in round 2, and 3 patients/caregivers in round 3. ²Consensus criteria for including a measure in the Core Outcome Measurement Set: $\geq 70\%$ of responses rate the measure “Critical” (i.e., score ≥ 7) and $\leq 15\%$ of responses rate the measure “Not Important” (i.e., score ≤ 3). ³Consensus criteria for excluding from the Core Outcome Measurement Set: $>15\%$ of responses rated the measure “Not Important” (i.e., score ≤ 3).

use of the single “Pain” question within the EQ-5D for the Core Outcome Measurement Set; however, further evaluation of the measurement properties of this item (vs. other pain-specific instruments, including a visual analogue scale, as recommended by an expert

group [29]) is recommended as part of a future research agenda (Figure 2). For the recommended core outcome of “Satisfaction with Life and Personal Enjoyment,” the EQ-5D generic health-related quality of life instruments reached consensus early, followed by the

Table 2. Measurement Instruments for Core Outcomes of Acute Respiratory Failure Survivors

Core Outcome	Measurement Instrument*	Number of Questions	Estimated Time Needed to Complete (min)	Cost (US\$)	Core Outcome Measurement Set Configurations			Optional Expanded Configurations	
					Minimum Acceptable	Minimum + SF-36	Minimum + Cognitive Screen†	Minimum + SF-36 and Cognitive Screen†	Minimum + Cognitive Screen†
Survival	N/A‡	N/A	N/A	N/A	✓	✓	✓	✓	✓
Satisfaction with life and personal enjoyment (HQOL)	EQ-5D [§] (3L or 5L version) SF-36 version 2 (optional)	6 36	2 9	Usually free Variable	✓	✓	✓	✓	✓
Mental Health	Hospital Anxiety and Depression Scale Impact of Event Scale-Revised [¶]	14 22	4 6	Approximately \$150 per 100 No cost	✓	✓	✓	✓	✓
Pain	EQ-5D pain question	1 (accounted for in EQ-5D)	Accounted for in EQ-5D	Accounted for in EQ-5D	✓	✓	✓	✓	✓
Cognition	Montreal Cognitive Assessment-BLIND [†]	13	5	No cost	X	X	✓	✓	✓
Physical function**	None	—	—	—	—	—	—	—	—
Muscle and/or nerve function**	None	—	—	—	—	—	—	—	—
Pulmonary function**	None	—	—	—	—	—	—	—	—
Total number of questions in this Core Outcome Measurement Set configuration					42	78	55	91	
Estimated time needed to complete this Core Outcome Measurement Set configuration, min					12	21	17	26	

Definition of abbreviations: HQOL = health-related quality of life; N/A = not applicable; SF-36 = 36-item Short Form Health Survey.

*For more information about each instrument, including data on costs and licensing fees, please visit www.improvelto.com/instruments.

†Although there was no consensus on a cognition measurement instrument, the Montreal Cognitive Assessment (MoCA) received the highest rating (55% agreed MoCA was “Critical” for inclusion) and has a version (i.e., MoCA-BLIND) that can be used via phone, like the rest of the Core Outcome Measurement Set; therefore, we encourage inclusion of this instrument in the Core Outcome Measurement Set. However, measurement properties of the MoCA have not yet been assessed in acute respiratory failure survivors. Use of MoCA must be registered at its website.

‡We strongly suggest, at a minimum, collecting date and location (e.g., home, hospital, or hospice) of death.

§We recommend using the newer EQ-5D-5L instrument, which has five levels of responses for each question, rather than the EQ-5D-3L, which has three levels of responses. Licensing fees, as determined by the EuroQol Executive Office, are usually free of charge for academic research.

¶If research funding does not permit use of the SF-36 version 2, researchers may consider use of the freely available RAND 36-item Health Survey 1.0 questionnaire (RAND SF-36 v1). The RAND SF-36 v1 is similar to the SF-36 version 2 and is available free of charge. Its differences from the SF-36 version 2 are described elsewhere (37).

**There was no consensus on measurement instruments to include for these outcomes. However, if in-person assessments are part of a research protocol, the following tests for evaluating these core outcomes received the highest scores from the panel: Physical Function (6-min-walk test [50% agreed this was “Critical” for inclusion]), Muscle and/or Nerve Function (manual muscle test [49% agreed this was “Critical” for inclusion]), and Pulmonary Function (panel voted to exclude all tests that were considered [e.g., spirometry, St. George’s Respiratory Questionnaire]).

Overall

- How do acute respiratory failure (ARF) survivors and research staff participating in clinical research studies perceive the Core Outcome Measurement Set? **High priority**
- Can instruments developed with item response theory and computer adaptive testing (IRT/CAT) be used to decrease the number and redundancy of questions within the Core Outcome Measurement Set?
 - Evaluate instruments from Patient-Reported Outcome Measurement Information System (PROMIS®) together with Core Outcome Measurement Set in populations of ARF survivors to establish a “cross-walk” between commonly used existing instruments and newer IRT/CAT instruments **High priority**

Cognition

- Evaluate measurement properties of the Montreal Cognitive Assessment-BLIND (MoCA-BLIND) instrument in ARF survivors **High priority**

Pain

- Evaluate measurement properties of the pain item within the EQ-5D-5L in ARF survivors **High priority**
- Evaluate measurement properties of the pain visual analogue scale and other pain instruments in ARF survivors

Survival

- Develop standards for ascertaining and reporting survival status and related data (e.g., date and location of death)

Muscle and/or Nerve Function

- Further evaluate strength measures obtained by Manual Muscle Testing versus Handgrip Dynamometry in ARF survivors
- Compare the feasibility of Manual Muscle Testing versus Handgrip Dynamometry in ARF survivors

Physical Function and Symptoms

- Evaluate patient-reported outcome measures of physical function in ARF survivors recognizing that such measures may not closely correlate with performance-based measures **High priority**
- Continue to evaluate and compare brief performance-based measures of physical function in ARF survivors (e.g., 4-meter gait speed test)

Pulmonary Function and Symptom

- Evaluate and/or develop a patient-reported outcome measure of pulmonary function and symptoms in ARF survivors **High priority**

Figure 2. Future research agenda for improving outcome measurement in acute respiratory failure survivors. EQ-5D-5L = EQ-5D five-level version.

36-item Short Form Health Survey (SF-36) in the last Delphi round. Both of these instruments are commonly used in critical care clinical research and have been recommended previously (2, 29, 30). Although little comparative evaluation of different versions of these instruments has been performed within critical care (8, 19), the most recent versions (i.e., EQ-5D five-level version and SF-36 version 2) are recommended. To ensure comparability between future studies in the field, the EQ-5D five-level version is specifically recommended for all studies; and for investigators wanting more comprehensive assessment, the SF-36 version 2 can be added, as was done in prior research (31).

Discussion

This three-round modified Delphi consensus process with an international panel of stakeholders representing clinical researchers from more than 16 countries across 6 continents, clinicians,

patient/caregivers, and U.S. federal research funding organizations results in recommendation of a Core Outcome Measurement Set for clinical research evaluating postdischarge outcomes of ARF survivors. By referring to a preestablished set of recommended core outcomes, the panel considered a preliminary list of 38 outcome measures, suggested 37 additional measures, and ultimately reached the *a priori* consensus criteria for 4 measures of 3 core outcomes, with additional measures and issues identified for consideration and for future research in the field. Participation and retention of panel members were excellent throughout the Delphi process.

For the “Cognition” core outcome, no instrument reached the *a priori* threshold for consensus. However, the instrument with the highest rating (with 55% rating it as “Critical” for inclusion) was the free-of-charge Montreal Cognitive Assessment (MoCA), which has not previously been used in critical care studies (8). Importantly, the commonly used Mini Mental State Examination instrument

requires payment of a licensing fee and recently has been empirically demonstrated to be a poor cognitive screening instrument in ARF survivors (32). The MoCA is already being used in at least one large, multicentered, randomized trial (33). Evaluation of the measurement properties of the MoCA is a high priority, given that cognition is a core outcome, that this instrument is already being used in clinical research in ARF survivors, and that no other instrument achieved sufficient support for inclusion (Figure 2).

Acceptable combinations of instruments within the Core Outcome Measurement Set are summarized in Table 2. The minimum acceptable Core Outcome Measurement Set (Table 2) includes three unique instruments comprising 42 questions with an estimated completion time of only 12 minutes. Including the optional MoCA instrument without SF-36, as well as the optional MoCA instrument with SF-36, raises the total number of questions to 55 and 91 and requires a total of 17 and 26 minutes,

respectively. Instrument-specific information, including detailed licensing information, is freely available from the project's website (www.improvelto.com/instruments).

Notably, the three core outcomes that are commonly evaluated using performance-based tests (physical function, muscle and/or nerve function, and pulmonary function) did not achieve consensus for inclusion of a measurement instrument as part of the Core Outcome Measurement Set. In their qualitative feedback, panel members raised important concerns about the feasibility of performance-based tests as part of a mandatory Core Outcome Measurement Set, given the need for in-person assessment and the added administrator skill, cost, and time required. Existing data on ARF survivors clearly demonstrate that performance-based tests and patient-reported outcome measures (i.e., surveys) measure distinct aspects of patient outcomes (9, 34, 35). Hence, investigators should not presume that survey-based measures will yield findings similar to performance-based tests for these three core outcomes. Within the "Physical Functioning" core outcome, the 6-minute-walk test clearly had the highest rating in round 3, with 50% of respondents rating it "Critical" for inclusion in the Core Outcome Measurement Set (Table E4). Notably, this test was recently demonstrated as valid and responsive in ARF survivors (36), is the most commonly used method for assessing physical activity limitations in ARF research (8), and has been recommended by an expert group for clinical evaluation of ICU survivors after hospital discharge (29). For the "Muscle and/or Nerve Function" core outcome, two performance-based tests of muscle strength (handgrip dynamometry and manual muscle testing) reached round 3 with similar voting results (46% and 49%, respectively, rated as "Critical"). Empirical research comparing the measurement properties and the feasibility of these two tests is important (Figure 2). Finally, in terms of the "Pulmonary Function" core outcome, all tests and surveys evaluated in

the Delphi process met exclusion criteria for the Core Outcome Measurement Set. Evaluation of new or existing instruments for this core outcome is an important area for future research (Figure 2).

This study has potential limitations. Given the nature of this expert consensus process, its results may be anchored in the use of instruments that traditionally have been used in prior research. Moreover, rigorous evaluation of the measurement properties of many instruments is lacking (19), which makes recommendations less certain. Furthermore, there may be overlap in the content of the Core Outcome Measurement Set instruments, leading to potential redundancy or inefficiency. As such, all Core Outcome Measurement Sets require revision and/or updating over time for several reasons, including (1) expanded use of item response theory and computer adaptive testing instruments (e.g., National Institutes of Health Patient-Reported Outcomes Measurement Information System initiative [9]); (2) new evaluations of the measurement properties of existing and newly developed instruments for ARF survivors; and (3) evaluation of the appropriateness of the Core Outcome Measurement Set for non-English-speaking patient populations, including consideration of language and cultural validations of the instruments. This study could not address issues that are relevant to the field but outside the scope of this project, such as recommendations for standardized timing of administration of these instruments after hospital discharge or how these research data are analyzed (see relevant resources for data analysis available from: www.improvelto.com/stats-tools). Finally, alternative compositions of panel members (e.g., inclusion of nonclinician methodologists or exclusion or reduction of patient/caregiver representation in light of the psychometric and clinically oriented information considered in selecting outcome measures) may have yielded different results when selecting the Core Outcome Measurement Set. Despite these limitations,

this project takes a vital step toward standardizing the use of measurement instruments to achieve greater comparability and reduced reporting bias in clinical research evaluating the outcomes of ARF survivors after hospital discharge.

In conclusion, research evaluating the outcomes of ARF survivors after hospital discharge has grown rapidly, with few deliberate efforts made to standardize the outcomes assessed or the measurement instruments used. Using a rigorous, modified Delphi process, an international panel of clinical researchers, clinicians, patients/caregivers, and U.S. federal research funding organization members reached consensus on the following measures for clinical research evaluating postdischarge outcomes of ARF survivors: (1) the EQ-5D (optional addition of SF-36 version 2), (2) the Hospital Anxiety and Depression Scale, and (3) the Impact of Events Scale-Revised. Investigators may also consider using the MoCA for evaluating the "Cognition" core outcome, even though it did not reach the threshold for consensus. These instruments are recommended for use in all future studies in which researchers elect to evaluate postdischarge outcomes of ARF survivors. This Core Outcome Measurement Set will need to be reevaluated as additional data on new and existing measurement instruments emerge, especially for the core outcomes (i.e., Cognition, Muscle and Nerve Function, Physical Function, and Pulmonary Function) for which there was no consensus regarding appropriate measures. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Wesley Davis for assistance with survey development, Paula Williamson for methodological advice, and Bronagh Blackwood and John Marshall for assistance with recruitment of stakeholder representatives.

References

1. Angus DC, Mira JP, Vincent JL. Improving clinical trials in the critically ill. *Crit Care Med* 2010;38:527-532.
2. Angus DC, Carlet J; 2002 Brussels Roundtable Participants. Surviving intensive care: a report from the 2002 Brussels Roundtable. *Intensive Care Med* 2003;29:368-377.
3. Spragg RG, Bernard GR, Checkley W, Curtis JR, Gajic O, Guyatt G, Hall J, Israel E, Jain M, Needham DM, et al. Beyond mortality: future clinical research in acute lung injury. *Am J Respir Crit Care Med* 2010; 181:1121-1127.
4. Lieu TA, Au D, Krishnan JA, Moss M, Selker H, Harabin A, Taggart V, Connors A; Comparative Effectiveness Research in Lung Diseases Workshop Panel. Comparative effectiveness research in lung

- diseases and sleep disorders: recommendations from the National Heart, Lung, and Blood Institute workshop. *Am J Respir Crit Care Med* 2011;184:848–856.
5. Needham DM, Davidson J, Cohen H, Hopkins RO, Weinert C, Wunsch H, Zawistowski C, Bemis-Dougherty A, Berney SC, Bienvenu OJ, *et al*. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. *Crit Care Med* 2012;40:502–509.
 6. Deutschman CS, Ahrens T, Cairns CB, Sessler CN, Parsons PE; Critical Care Societies Collaborative USCIITG Task Force on Critical Care Research. Multisociety Task Force for Critical Care Research: key issues and recommendations. *Crit Care Med* 2012;40:254–260.
 7. Carson SS, Goss CH, Patel SR, Anzueto A, Au DH, Elborn S, Gerald JK, Gerald LB, Kahn JM, Malhotra A, *et al*.; American Thoracic Society Comparative Effectiveness Research Working Group. An official American Thoracic Society research statement: comparative effectiveness research in pulmonary, critical care, and sleep medicine. *Am J Respir Crit Care Med* 2013;188:1253–1261.
 8. Turnbull AE, Rabiee A, Davis WE, Nasser MF, Venna VR, Lolitha R, Hopkins RO, Bienvenu OJ, Robinson KA, Needham DM. Outcome measurement in ICU survivorship research from 1970 to 2013: a scoping review of 425 publications. *Crit Care Med* 2016;44:1267–1277.
 9. Needham DM. Understanding and improving clinical trial outcome measures in acute respiratory failure. *Am J Respir Crit Care Med* 2014;189:875–877.
 10. Blackwood B, Marshall J, Rose L. Progress on core outcome sets for critical care research. *Curr Opin Crit Care* 2015;21:439–444.
 11. Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39.
 12. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;13:132.
 13. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, *et al*. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;3:e3081.
 14. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
 15. Turnbull AE, Sepulveda KA, Dinglas VD, Chessare CM, Bingham CO III, Needham DM. Core domains for clinical research in acute respiratory failure survivors: an international modified Delphi consensus study. *Crit Care Med* 2017;45:1001–1010.
 16. Hodgson CL, Turnbull AE, Iwashyna TJ, Parker A, Davis W, Bingham CO, Watts NR, Finfer S, Needham DM. Core domains in evaluating patient outcomes after acute respiratory failure: international multidisciplinary clinician consultation. *Phys Ther* 2017;97:168–174.
 17. Gorst SL, Gargon E, Clarke M, Smith V, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and identification of gaps. *PLoS One* 2016;11:e0168403.
 18. Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Moher D, Schmitt J, Tugwell P, *et al*. Core Outcome Set-STAndards for Reporting: the COS-STAR Statement. *PLoS Med* 2016;13:e1002148.
 19. Robinson KA, Davis WE, Dinglas VD, Mendez-Tellez PA, Rabiee A, Sukrithan V, Yalamanchilli R, Turnbull AE, Needham DM. A systematic review finds limited data on measurement properties of instruments measuring outcomes in adult intensive care unit survivors. *J Clin Epidemiol* 2017;82:37–46.
 20. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–926.
 21. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
 22. Bartlett SJ, Hewlett S, Bingham CO III, Woodworth TG, Alten R, Pohl C, Choy EH, Sanderson T, Boonen A, Bykerk V, *et al*.; OMERACT RA Flare Working Group. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Ann Rheum Dis* 2012;71:1855–1860.
 23. Prinsens CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” – a practical guideline. *Trials* 2016;17:449.
 24. Rabiee A, Nikayin S, Hashem MD, Huang M, Dinglas VD, Bienvenu OJ, Turnbull AE, Needham DM. Depressive symptoms after critical illness: a systematic review and meta-analysis. *Crit Care Med* 2016;44:1744–1753.
 25. Nikayin S, Rabiee A, Hashem MD, Huang M, Bienvenu OJ, Turnbull AE, Needham DM. Anxiety symptoms in survivors of critical illness: a systematic review and meta-analysis. *Gen Hosp Psychiatry* 2016;43:23–29.
 26. Parker AM, Sricharoenchai T, Rappala S, Schneck KW, Bienvenu OJ, Needham DM. Posttraumatic stress disorder in critical illness survivors: a metaanalysis. *Crit Care Med* 2015;43:1121–1129.
 27. Stevenson JE, Colantuoni E, Bienvenu OJ, Sricharoenchai T, Wozniak A, Shanholtz C, Mendez-Tellez PA, Needham DM. General anxiety symptoms after acute lung injury: predictors and correlates. *J Psychosom Res* 2013;75:287–293.
 28. Chan KS, Aronson Friedman L, Bienvenu OJ, Dinglas VD, Cuthbertson BH, Porter R, Jones C, Hopkins RO, Needham DM. Distribution-based estimates of minimal important difference for Hospital Anxiety and Depression Scale and Impact of Event Scale-Revised in survivors of acute respiratory failure. *Gen Hosp Psychiatry* 2016;42:32–35.
 29. Major ME, Kwakman R, Kho ME, Connolly B, McWilliams D, Denehy L, Hanekom S, Patman S, Gosselink R, Jones C, *et al*. Surviving critical illness: what is next? An expert consensus statement on physical rehabilitation after hospital discharge. *Crit Care* 2016;20:354.
 30. Heyland DK, Stapleton RD, Mourtzakis M, Hough CL, Morris P, Deutz NE, Colantuoni E, Day A, Prado CM, Needham DM. Combining nutrition and exercise to optimize survival and recovery from critical illness: conceptual and methodological issues. *Clin Nutr* 2016;35:1196–1206.
 31. Needham DM, Dinglas VD, Bienvenu OJ, Colantuoni E, Wozniak AW, Rice TW, Hopkins RO; NIH NHLBI ARDS Network. One year outcomes in patients with acute lung injury randomised to initial trophic or full enteral feeding: prospective follow-up of EDEN randomised trial. *BMJ* 2013;346:f1532.
 32. Pfoh ER, Chan KS, Dinglas VD, Girard TD, Jackson JC, Morris PE, Hough CL, Mendez-Tellez PA, Ely EW, Huang M, *et al*.; NIH NHLBI ARDS Network. Cognitive screening among acute respiratory failure survivors: a cross-sectional evaluation of the Mini-Mental State Examination. *Crit Care* 2015;19:220.
 33. Huang DT, Angus DC, Moss M, Thompson BT, Ferguson ND, Ginde A, Gong MN, Gundel S, Hayden DL, Hite RD, *et al*.; Reevaluation of Systemic Early Neuromuscular Blockade Protocol Committee and the National Institutes of Health National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Network Investigators. Design and rationale of the reevaluation of systemic early neuromuscular blockade trial for acute respiratory distress syndrome. *Ann Am Thorac Soc* 2017;14:124–133.
 34. Needham DM, Wozniak AW, Hough CL, Morris PE, Dinglas VD, Jackson JC, Mendez-Tellez PA, Shanholtz C, Ely EW, Colantuoni E, *et al*.; National Institutes of Health NHLBI ARDS Network. Risk factors for physical impairment after acute lung injury in a national, multicenter study. *Am J Respir Crit Care Med* 2014;189:1214–1224.
 35. Chan KS, Aronson Friedman L, Dinglas VD, Hough CL, Shanholtz C, Ely EW, Morris PE, Mendez-Tellez PA, Jackson JC, Hopkins RO, *et al*. Are physical measures related to patient-centred outcomes in ARDS survivors? *Thorax* [online ahead of print] 20 Jan 2017; DOI: 10.1136/thoraxjnl-2016-209400.
 36. Chan KS, Pfoh ER, Denehy L, Elliott D, Holland AE, Dinglas VD, Needham DM. Construct validity and minimal important difference of 6-minute walk distance in survivors of acute respiratory failure. *Chest* 2015;147:1316–1326.
 37. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. *Ann Med* 2001;33:350–357.