



Published in final edited form as:

*J Clin Epidemiol.* 2018 October ; 102: 99–106. doi:10.1016/j.jclinepi.2018.06.007.

## A Survey of Delphi Panelists after Core Outcome Set Development Revealed Positive Feedback and Methods to Facilitate Panel Member Participation

Alison E. Turnbull, DVM, MPH, PhD<sup>a,b,c</sup>, Victor D. Dinglas, MPH<sup>a,b</sup>, Lisa Aronson Friedman, Sc.M<sup>a,b</sup>, Caroline M. Chessare, MS<sup>a,b</sup>, Kristin A. Sepúlveda, BA<sup>a,b</sup>, Clifton O. Bingham III, MD<sup>a,b,d</sup>, Dale M. Needham, FCPA, MD, PhD<sup>a,b,e</sup>

<sup>a</sup>Outcomes After Critical Illness and Surgery (OACIS) Group, Johns Hopkins University, Baltimore, MD, USA

<sup>b</sup>Division of Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

<sup>c</sup>Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

<sup>d</sup>Divisions of Rheumatology and Allergy and Clinical Immunology, Johns Hopkins University, Baltimore, MD, USA

<sup>e</sup>Department of Physical Medicine and Rehabilitation, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

### Abstract

**Objective(s):** To elicit feedback on consensus methodology used for core outcome set development.

**Study Design:** Online survey of international Delphi panelists participating in a recent Core Outcome Set for clinical research studies evaluating acute respiratory failure (ARF) survivors.

**Setting:** Panelists represented 14 countries (56% outside USA).

**Results:** Seventy (92%) panelists completed the survey, including 32 researchers, 19 professional association representatives, 4 research funding representatives, and 15 ARF survivors/caregiver members. Among respondents, 91% reported that the time required to participate was appropriate and 96% were not bothered by reminders for timely response. Attributes of measurement instruments and voting results from previous rounds were evaluated differently across stakeholder groups. When measurement properties were explained in the stem of the survey question, 59 (84%) panelists (including 73% of survivors/families) correctly interpreted information about an instrument's reliability. Without a reminder in the stem, only 20 (29%) panelists (including 38% of researchers), correctly identified properties of a core outcome set.

---

**Corresponding Author:** Alison E. Turnbull, DVM, MPH, PhD, Pulmonary and Critical Care Medicine, Johns Hopkins University, 1830 E. Monument St, 5<sup>th</sup> Floor, Baltimore, MD 21205. Phone: (410) 955-2190; turnbull@jhmi.edu.

**Declarations of Interest:** None

**Conclusions:** This international Delphi panel, including >20% patients/caregivers, favorably reported on feasibility of the methodology. Providing all panelists pertinent information/reminders about the project's objective at each voting round is important to informed decision-making across all stakeholder groups.

### Keywords

consensus methods; core outcome set development; Delphi study; stakeholders; feedback strategies

---

## Introduction

A Core Outcome Set (COS) is a minimum collection of outcomes reported in all studies within a specific field.[1,2] Similarly, a Core Outcome Measurement Set (COMS) contains the measurement instruments used to assess outcomes within a COS. Core set adoption improves trial efficiency, facilitates comparisons and meta-analyses within a field, and helps to prevent bias from selective outcome reporting, while still permitting researchers to evaluate additional outcomes of relevance to their study.[3,4] Incorporating input from a panel of diverse stakeholders helps to ensure core sets contain the outcomes and measures which are most valued by patients, families, clinicians, clinical researchers, and research funding organizations.

The modified Delphi consensus methodology is a common way to reach consensus on COS/COMS projects.[5][6] However, a Delphi process, which involves multiple rounds of voting by a large panel of stakeholders, can also be challenging because all panelists must understand fundamental properties of outcomes and measurement instruments to serve as informed voters. Because patients and family caregivers are essential stakeholders, but often have no clinical research experience, integrating their input into the Delphi process can be challenging.[7,8] Substantial effort also may be required to ensure a high participation rate among panelists during each round of voting. Delphi moderators must decide how best to prepare panel members for voting, what background information about outcomes and measurement instruments to provide, and how to ensure timely voting.

To help future Delphi moderators navigate these design decisions, we elicited Delphi panel member feedback. We recently conducted a Delphi process to develop COS/COMS that included stakeholders from >16 countries, including ARF survivors and their caregivers. These stakeholders participated in 5 rounds of voting, reviewed information on 36 outcomes and 75 measurement instruments, and more than 90% of panelists voted in each of the 5 Delphi rounds. Therefore, we asked stakeholders to report on the burden of participation and reminders to vote in each round, and on how they weighed provided background information and feedback from other stakeholder groups when voting. We also asked two questions assessing stakeholder understanding of key information needed to inform voting.

## Materials and Methods

We conducted a cross-sectional, online survey (Qualtrics, Provo, UT) of 76 stakeholders who recently participated in an international, 2-stage Delphi consensus process to develop

both a COS and COMS for post-discharge clinical research studies evaluating acute respiratory failure (ARF) survivors,[9,10] approximately 3 months after completion of the Delphi process. To develop the survey, we generated questions based on the expertise and experience of the researchers administering the Delphi, and reviewed questions asked of panelists in previous evaluations of Delphi processes.[11–13] Survey questions were tested for clarity and readability, with iterative refinement, using input from 4 ARF survivors/ caregivers and 6 clinical researchers. The final result was a 30-question survey with both multiple choice and open-ended/free text questions assessing: 1) the burden of Delphi participation, 2) how panelists used background information provided by the research team to prepare themselves for voting, 3) how panelists considered and weighed feedback and voting from earlier Delphi rounds, 4) panelist understanding of information provided about measurement instrument properties, and 5) panelist understanding of how core outcome sets are used in research. The complete text of the survey instrument is available at [www.improveLTO.com/delphi-methods](http://www.improveLTO.com/delphi-methods).

The 5-stage Delphi process occurred from January 5, 2016 to October 10, 2016. The Delphi panel included representatives of each of the 21 members of the International Forum for Acute Care Trialists (InFACT) organization, as well as clinical researchers identified through random sampling of a pre-existing database of corresponding authors on studies of ICU survivors, and representatives of clinicians, ICU patients and caregivers identified by professional associations and patient family advisory councils.[9] The 5 rounds of voting were completed in 157 days, with the median number of weeks for response to each round of 1 (IQR: 0, 2). Each panelist received an e-mail invitation containing a link to the follow-up survey regardless of their participation rate during the Delphi process. All initial e-mail invitations included the names and affiliations of study investigators, and requested survey completion within 5 days. Reminder e-mails were sent to panelists who had not completed the survey on days 7, 15, 28, 35, and 48 following the initial invitation, after which telephone calls and text messages were used to contact non-respondents.

Survey response rate for this study was defined as the proportion of Delphi panel members sent an invitation who subsequently completed the follow-up survey. Responses to multiple choice survey questions were summarized using counts and percentages for categorical variables, and medians and interquartile ranges (IQR) for continuous variables. Response options to questions about the burden of survey participation, and about considering the voting results from previous rounds and stakeholder groups, used a 5-point Likert scale with the following options: strongly agree, agree, neutral, disagree, strongly disagree, and in some cases, not applicable. Likert-scale responses to questions about the importance panel members placed on educational information when they were voting were: extremely important, very important, moderately important, slightly important, and not at all important. Differences in response to multiple choice questions across stakeholder groups were compared using Fisher's exact test. A P-value <.05 was considered statistically significant. Responses to open-ended questions were not evaluated as part of this analysis of survey findings. All descriptive statistics and tests were performed using SAS® version 9.4 (2013, Cary, NC). The Institutional Review Board of Johns Hopkins University approved this study.

## Results

Of the 76 invited Delphi panelists, 70 (92%) completed the follow-up survey, including 32 clinical researchers, 19 clinical professional associations representatives, 4 US research funding representatives, and 15 ARF survivors and caregivers. Among all responding panelists, 35 (50%) were male, 39 (56%) resided outside of the U.S., and 23 (33%) were physicians with specialty training in critical care medicine. Participating clinicians reported a median of 15 (IQR 9 – 21) years of professional experience (Table 1).

The counts and percentages of panelists agreeing or strongly agreeing with statements about the time and burden associated with being a panel member are presented in Table 2, stratified by stakeholder group. Among all stakeholders completing the survey, 67 (96%) agreed that participating in the Delphi process was important, 64 (91%) agreed that the time required to participate in the process was appropriate, and 62 (89%) agreed that they would participate in a research studying using a Delphi consensus process again. Compared to other stakeholders, a similar proportion of former ARF patients and family caregivers felt the time required to participate was appropriate (87% vs 93%,  $P=.60$ ) but relatively fewer patients and caregivers reported that they would participate in a future Delphi process (73% vs 93%,  $P=.06$ ). Among responding clinical researchers, 28 (97%) said they planned to use the Core Outcome Set resulting from the Delphi consensus process.

Only 3 (4%) panelists reported being bothered by the study team's attempts to contact panelists who had not voted following the initial survey release in each round of the Delphi. During the 5 rounds of voting in the Delphi consensus process, the Delphi administrators issued a total of 680 reminders (median of 1 reminder[interquartile 0-2] per round), including e-mails, phone calls, and text messages, with >90% of panelists voting in each of the 5 Delphi rounds. During the 5 rounds of voting, 73 (94%) of panelists ever received an e-mail reminder(s); 37 (47%) received >2 e-mail reminders during a single round of voting, and 32 (41%) ever received a telephone call(s) or text message(s) during the entire 5-round Delphi process.

Among survey respondents, 63 (90%) reported considering the written comments of other panelists and 58 (83%) reported considering voting results from outside of their own stakeholder group (Table 2). Consideration of prior voting by other stakeholder groups was highest among representatives of funding organizations (100%) and lowest among patients and caregivers (67%). Clinical researchers prioritized the voting results of other clinical researchers first, then patients and caregivers, and then clinicians (47% vs 28% vs 25%). Clinicians were similarly prioritized voting by clinical researchers and patients and caregivers (37% vs 37%). In contrast, patients and caregivers placed the most weight on the voting of clinicians, followed by other patients and caregivers, and then clinical researchers (60% vs 33% vs 7%).

While voting on measurement instruments[10] for assessing each core outcome[9], panelists were provided information about each measurement instrument under consideration. This information was summarized via a 1 or 2 page "Measure Card" which displayed standardized information about 14 attributes of the instrument. These attributes included

practical considerations, such as the estimated time required for completion and licensing fees. When psychometric evaluations of the instrument had been performed in the target population, the measure card included the results of that evaluation, as well as an assessment of any such evaluation using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative checklist[14,15]. Among responding panelists, 14% reported reviewing between 0 - 25% of the Measure Cards during the Delphi process, 16% reported reviewing 26% - 50%, 14% reported reviewing 51% - 75%, 29% reported reviewing 76% - 99%, and 27% reported reviewing 100% of Measure Cards. The importance of each section of the Measure Card, according to the responding panelists, is presented in Table 3, stratified by stakeholder group. The measure attribute most frequently rated as extremely important or very important (83%) was the estimated time required for completion.

Panelist's understanding of information provided about measurement instrument properties was assessed by a single multiple-choice question in which participants were asked to interpret a measure with both poor reliability and an excellent COSMIN rating (Table 4). A brief reminder of the COSMIN rating system was included in the stem of the question, as was the case in the actual Delphi process. There were 59 (84%) panel members who answered this question correctly, with correct answers ranging from 100% among representatives of funding organizations to 73% among patients and caregivers, and no significant difference in the proportion of patients and caregivers who answered correctly when compared to other panel members (73% vs 87%,  $P=0.23$ ).

Understanding of how core outcome sets are used in research was assessed by a single multiple-choice question without any guidance provided in the question stem or instructions, unlike in the actual Delphi process (Table 4). Overall, 20 (29%) panel members answered the question correctly, 25 (36%) said they were unsure, and 25 (36%) answered incorrectly. Patients and caregivers were more likely to answer that they were unsure of the correct answer (87% vs 22%,  $P<0.01$ ), but less likely to provide an incorrect answer (7% vs 44%,  $P=0.01$ ) or a correct answer (7% vs 35%,  $P=0.05$ ).

## Discussion

This cross-sectional survey of 70 recent Delphi panelists from 14 countries found that both clinicians and lay participants considered the time required to participate in COS/COMS development appropriate, and the vast majority of clinical researchers intend to use the resulting core sets in their research. While almost half of panelist received a phone or text reminder at some point during the Delphi process, only 3 panelists reported being bothered by the study team's contacts. While most panelists considered the feedback and results from previous rounds of voting when casting votes, the importance placed on the prior voting of different stakeholder groups varied substantially. Standardized information provided about measurement instruments under consideration was reviewed by most panelists. Finally, all stakeholder groups could interpret information about a measurement instrument's reliability when guidance was provided in the stem of the question. However, without such guidance, respondents struggled to correctly identify the properties of a core set approximately 3 months after completion of the Delphi project.

To our knowledge, studies evaluating Delphi panelists' feedback have occurred with mental health guidelines, but not following COS/COMS development.[11–13] Despite this difference, the feedback obtained in our study was similar, with high levels of agreement that the time and effort required to reach consensus over multiple Delphi rounds was worthwhile. The enthusiasm from participating clinical researchers for utilizing the resulting COMS in their own research is encouraging, and suggests that including active clinical researchers in the consensus process may help drive COS/COMS adoption.

Substantial effort, including multiple e-mail reminders, telephone calls, and text messages, helped achieve a >90% response rate during each of the 5 Delphi rounds. Despite such repeated contact from the study team, only 4% of responding panelists reported being bothered by these reminders. Although social desirability bias may have influenced panelists to provide positive feedback, it appears that reminders and personalized contact through telephone calls and text messages were generally an acceptable and effective way to optimize participation among a Delphi panel that is otherwise very busy with other activities. Researchers looking to foster high voting participation rates should collect, at the time of panel recruitment, robust contact information (e.g., e-mail addresses, multiple phone numbers, optimal time(s) to telephone panelist – see example contact information form used in this Delphi project at [www.improvelto.com/participant-contact-information-sheet](http://www.improvelto.com/participant-contact-information-sheet)), and budget sufficient time for study personnel to repeatedly follow-up with non-responders.

In our survey, the panelist responses indicated that they placed differential weight on feedback from different stakeholder groups. This finding contrasts with a recent randomized trial that reported no evidence that Delphi panelist voting was influenced by whether feedback was combined across stakeholder groups, stratified by stakeholder group, or provided only for the panelist's own stakeholder group.[16] However this lack of difference may have been explained by a very high degree of agreement on outcomes overall in the randomized trial. If panelists truly make decisions based largely on the opinions of a particular stakeholder group, our finding suggests that presenting feedback from multiple stakeholders separately may be optimal. Conversely, if Delphi administrators prefer that all panel member input be weighed equally, feedback should not be stratified by stakeholder group. Future research using interviews and focus groups may also help clarify how panelists use feedback and background information during the Delphi process.

Former ARF patients and caregivers were included as voting panelists in this Delphi process to help ensure core sets included outcomes and measurement instruments valued by these stakeholders. Despite the many background materials provided and the numerous items under consideration, the majority of patients and caregivers reported reviewing many parts of these materials and being willing to participate again in another Delphi process. Although they generally reported reviewing fewer parts of the measurement cards, they performed nearly as well as other stakeholders when provided with guidance on how to interpret information about an instrument's reliability and COSMIN rating. Therefore, we recommend that investigators leading COS/COMS development efforts include reminders about the purpose and properties of core sets at the start of each round of voting and guidance on how to interpret information about psychometric evaluations whenever this information is presented. These additional steps both help patients and caregivers to



participate as voting panelists, and assist other stakeholder groups to be more informed voters.

This survey was conducted 3 months after the completion of the last round of Delphi voting which may have limited panel member recall. To minimize participant burden in our survey, we asked panelists only two questions to assess their understanding of core set properties and their ability to interpret information provided about measurement instruments. This limited our ability to infer how well stakeholders grasped additional fundamental principles that are central to making informed decisions about core set composition. However, by varying the amount of guidance provided in the stem of the two questions, a more nuanced picture emerged in which panelists appeared not to retain essential information approximately 3 months after voting, but were generally capable of manipulating complex information when a reminder of how to interpret the information was included with the question. These results also underscore that many clinical researchers are unfamiliar with core outcome sets, and continued education on their properties and benefits are necessary to facilitate widespread adoption. We were not able to address many of the questions that vex investigators designing Delphi processes for COMS/COS development, including the number of participants to include in each stakeholder group, the number items each panel member can effectively review, and whether all stakeholders should participate in all voting rounds.

## Conclusions

A diverse, international group of stakeholders, including former patients and their caregivers, provided positive feedback on their experiences as panelists in a 5-round, modified Delphi consensus process to develop both a Core Outcome Set and Core Outcome Measurement Set. To achieve a high response rate during each round of voting, study team members received repeated reminders which the vast majority of panelists did not find bothersome. Repeating essential information in the stem of survey questions may help stakeholders remember vital principles about Core Outcome Sets and facilitate informed voting during multi-round Delphi processes.

## Acknowledgements:

We thank Rakesh Allamneni, MBBS for his assistance with the initial drafting of the survey questions.

## References

- [1]. Clarke M Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39. doi:10.1186/1745-6215-8-39. [PubMed: 18039365]
- [2]. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;13:132. doi:10.1186/1745-6215-13-132. [PubMed: 22867278]
- [3]. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. *PLOS ONE* 2008;3:e3081. doi:10.1371/journal.pone.0003081. [PubMed: 18769481]
- [4]. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365. doi:10.1136/bmj.c365. [PubMed: 20156912]

- [5]. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med* 2011;8:e1000393. doi:10.1371/journal.pmed.1000393. [PubMed: 21283604]
- [6]. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. *PLoS One* 2016;11:e0146444. doi:10.1371/journal.pone.0146444. [PubMed: 26785121]
- [7]. Harman NL, Bruce IA, Kirkham JJ, Tierney S, Callery P, O'Brien K, et al. The Importance of Integration of Stakeholder Views in Core Outcome Set Development: Otitis Media with Effusion in Children with Cleft Palate. *PLoS ONE* 2015;10:e0129514. doi:10.1371/journal.pone.0129514. [PubMed: 26115172]
- [8]. Turnbull AE, Sahetya SK, Needham DM. Aligning critical care interventions with patient goals: A modified Delphi study. *Heart Lung J Crit Care* 2016;45:517–24. doi:10.1016/j.hrtlng.2016.07.011.
- [9]. Turnbull AE, Sepulveda KA, Dinglas VD, Chessare CM, Bingham CO, Needham DM. Core Domains for Clinical Research in Acute Respiratory Failure Survivors: An International Modified Delphi Consensus Study. *Crit Care Med* 2017;45:1001–10. doi:10.1097/CCM.0000000000002435. [PubMed: 28375853]
- [10]. Needham DM, Sepulveda KA, Dinglas VD, Chessare CM, Aronson Friedman L, Bingham CO III, et al. Core Outcome Measures for Clinical Research in Acute Respiratory Failure Survivors: An International Modified Delphi Consensus Study. *Am J Respir Crit Care Med* 2017. doi:10.1164/rccm.201702-0372OC.
- [11]. Hart LM, Jorm AF, Kanowski LG, Kelly CM, Langlands RL. Mental health first aid for Indigenous Australians: using Delphi consensus studies to develop guidelines for culturally appropriate responses to mental health problems. *BMC Psychiatry* 2009;9:47. doi:10.1186/1471-244X-9-47. [PubMed: 19646284]
- [12]. Hart LM, Bouchier SJ, Jorm AF, Kanowski LG, Kingston AH, Stanley D, et al. Development of mental health first aid guidelines for Aboriginal and Torres Strait Islander people experiencing problems with substance use: a Delphi study. *BMC Psychiatry* 2010;10:78. doi:10.1186/1471-244X-10-78. [PubMed: 20932326]
- [13]. Chalmers KJ, Bond KS, Jorm AF, Kelly CM, Kitchener BA, Williams-Tchen A. Providing culturally appropriate mental health first aid to an Aboriginal or Torres Strait Islander adolescent: development of expert consensus guidelines. *Int J Ment Health Syst* 2014;8:6. doi:10.1186/1752-4458-8-6. [PubMed: 24467923]
- [14]. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil* 2012;21:651–7. doi:10.1007/s11136-011-9960-1.
- [15]. Robinson KA, Davis WE, Dinglas VD, Mendez-Tellez PA, Rabiee A, Sukrithan V, et al. 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. doi:10.1016/j.jclinepi.2016.08.014. [PubMed: 27865899]
- [16]. MacLennan S, Kirkham J, Lam TBL, Williamson PR. A randomised trial comparing three Delphi feedback strategies found no evidence of a difference in a setting with high initial agreement. *J Clin Epidemiol* 2017. doi:10.1016/j.jclinepi.2017.09.024.



### What is new?

#### Key Findings:

- After 5 rounds of voting, >90% of Delphi panelists reported that the time required to participate was appropriate and they were not bothered by repeated reminders to encourage timely voting.
- While most panelists considered the feedback and results from previous rounds of voting when casting votes, the importance placed on the prior voting of different stakeholder groups varied substantially.
- Without guidance, respondents struggled to correctly identify the properties of a core set approximately 3 months after completion of the Delphi project.

#### What this adds to what is known:

- All stakeholder groups benefit from repeated guidance on principles for core outcome set development during voting.

#### What is the implication? What should change now?

- A high participation rate from a diverse international panel of stakeholders including patients and caregivers can be achieved during core outcome set development with real-time provision of pertinent information and timely voting reminders.

Table 1.

## Characteristics of survey respondents

Characteristic	All Panel Members (n=70)	Clinical Researchers (n=32)	Clinicians/Professional Assoc. (n=19)	US Federal Research Funding Organizations (n=4)	Patients and Caregivers (n=15)
Male, n (%)	35 (50)	21 (66)	7 (37)	1 (25)	6 (40)
Age, n (%)					
25 - 44	29 (41)	13 (41)	7 (37)	1 (25)	8 (53)
45 - 64	38 (54)	19 (59)	12 (63)	3 (75)	4 (27)
65	3 (4)	0 (0)	0 (0)	0 (0)	3 (20)
Country of residence, n (%)					
United States	31 (44)	8 (25)	10 (53)	4 (100)	9 (60)
Canada, United Kingdom, and Australia	28 (40)	13 (41)	9 (47)	0 (0)	6 (40)
Other*	11 (16)	11 (34)	0 (0)	0 (0)	0 (0)
Years of education, median (IQR)	20 (18-22)	21 (19-22)	20 (19-21)	22 (20-22)	18 (16-18)
Clinical work: Type of training, n (%) <sup>†</sup>					
Physician - Critical Care	23 (55)	17 (81)	5 (28)	1 (100)	0 (0)
Physical, Occupational, or Respiratory Therapist and/or Speech Language Pathologist	10 (24)	4 (19)	6 (33)	0 (0)	0 (0)
Nurse or Nurse Practitioner	6 (14)	0 (0)	4 (22)	0 (0)	2 (100)
Physician - Physical Medicine & Rehabilitation	2 (5)	0 (0)	2 (11)	0 (0)	0 (0)
Other clinical training	1 (2)	0 (0)	1 (6)	0 (0)	0 (0)
Years of professional experience, median (IQR) <sup>‡</sup>	15 (9-21)	14 (8-20)	17 (9-21)	19 (14-23)	22 (5-38)
Area(s) of professional expertise, n (%) <sup>§</sup>					
Physical health and functioning	40 (57)	25 (78)	13 (68)	1 (25)	N/A
Mental health	16 (23)	11 (34)	4 (21)	0 (0)	N/A
Cognitive function	17 (24)	11 (34)	6 (32)	0 (0)	N/A
Other	8 (11)	4 (13)	3 (16)	1 (25)	N/A
None	7 (10)	2 (6)	3 (16)	1 (25)	N/A
Professional interest in critical illness, n (%) <sup>§</sup>					
Research - Clinical	48 (69)	32 (100)	15 (79)	1 (25)	0 (0)

Characteristic	All Panel Members (n=70)	Clinical Researchers (n=32)	Clinicians/Professional Assoc. (n=19)	US Federal Research Funding Organizations (n=4)	Patients and Caregivers (n=15)
Research - Basic or translational	15 (21)	10 (31)	4 (21)	1 (25)	0 (0)
Clinical work	42 (60)	21 (66)	18 (95)	1 (25)	2 (13)
None of the above	15 (21)	0 (0)	0 (0)	2 (50)	13 (87)

**Abbreviations:** IQR, Inter-quartile Range; Assoc, Association

\* Other countries: 1 each from Belgium, France, Germany, Greece, Ireland, Italy, Netherlands, Norway and Singapore, and 2 from Brazil

<sup>†</sup> 42(60%) panel members selected clinical training. (Clinical Researchers = 21; Clinicians/Professional Associations = 18; US Federal Research Funding Organizations=1; and Patients/Caregivers=2). Other clinical training includes anesthesiology (n=2).

<sup>‡</sup> Two funding body representatives responded and reported 23 and 14 years of professional experience.

<sup>§</sup> Panel members could select > 1 response

Table 2.

Respondent level of agreement with statements about Delphi method and process

Statement, N (%) selecting response options agree/strongly agree*	All Panel Members (n=70)	Clinical Researchers (n=32)	Clinicians/ Professional Assoc. (n=19)	US Federal Research Funding Organizations (n=4)	Patients and Caregivers (n=15)	P value for comparison of caregivers vs. other panel members <sup>#</sup>
Participating in this Delphi Consensus process was important	67 (96)	31 (97)	18 (95)	4 (100)	14 (93)	0.52
In my own future research, I plan to follow the Core Outcome Set resulting from this project <sup>†</sup>	36 <sup>§</sup> (88)	28 (97)	8 (67)	N/A	N/A	N/A
The time required of me to participate in the Delphi Consensus process was appropriate*	64 (91)	28 (88)	19 (100)	4 (100)	13 (87)	0.60
I was bothered by the attempts made to contact me to complete each survey (e.g., e-mails, phone calls or text messages) <sup>†,‡</sup>	3 (4)	1 (3)	0 (0)	0 (0)	2 (13)	0.11
I would participate again in a research study using a Delphi Consensus process*	62 (89)	29 (91)	18 (95)	4 (100)	11 (73)	0.06
When I was voting in the Delphi consensus process I considered the prior voting results from the other stakeholder groups in my new voting*	58 (83)	26 (81)	18 (95)	4 (100)	10 (67)	0.11
When I was voting in the Delphi consensus process I considered the prior written comments from the other panel members in my new voting*	63 (90)	28 (88)	19 (100)	4 (100)	12 (80)	0.16
When I was voting in the Delphi consensus process, in reviewing prior voting results from the 4 stakeholder groups, I placed the most weight on results from...						
Clinical Researchers	25 (36)	15 (47)	7 (37)	2 (50)	1 (7)	0.06
Clinicians / Professional Association	24 (34)	8 (25)	5 (26)	2 (50)	9 (60)	
US Federal Research Funding Organizations	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	
Patients and Caregivers	21 (30)	9 (28)	7 (37)	0 (0)	5 (33)	

\* Response options were Strongly agree, agree, neutral, disagree and strongly disagree

† Response options were Strongly agree, agree, neutral, disagree, strongly disagree and not applicable.

‡ I responded 'not applicable'

§ 29 responded 'not applicable', including 3 clinical researchers, and 7 clinicians/Professional Association representatives.

# Fisher's exact test

**Table 3.** Proportion of panel members rating each section of the Measure Card as ‘Very Important’ or ‘Extremely Important’.\*

	All Panel Members (n=70)	Clinical Researchers (n=32)	Clinicians/Professional Assoc. (n=19)	US Federal Research Funding Organizations (n=4)	Patients and Caregivers (n=15)
Section 1: Number of questions	42 (60)	20 (63)	13 (68)	3 (75)	6 (40)
Section 2: Description of instrument	48 (69)	22 (69)	13 (68)	4 (100)	9 (60)
Section 3: Instrument versions	14 (20)	8 (25)	4 (21)	1 (25)	1 (7)
Section 4: Recall period	23 (33)	12 (38)	6 (32)	2 (50)	3 (20)
Section 5: Scoring information	32 (46)	14 (44)	11 (58)	2 (50)	5 (33)
Section 6: Estimated time to complete	58 (83)	28 (88)	17 (89)	4 (100)	9 (60)
Section 7: Administer to (e.g., patient, proxy)	47 (67)	22 (69)	16 (84)	3 (75)	6 (40)
Section 8: Requires trained administrators	49 (70)	24 (75)	17 (89)	3 (75)	5 (33)
Section 9: Mode of administration (e.g., in-person, phone)	51 (73)	28 (88)	15 (79)	3 (75)	5 (33)
Section 10: Licensing fee information	42 (60)	21 (66)	16 (84)	3 (75)	2 (13)
Section 11: Required equipment	52 (74)	27 (84)	16 (84)	4 (100)	5 (33)
Section 12: Number of published critical care publications using instrument	38 (54)	19 (59)	12 (63)	4 (100)	3 (20)
Section 13: Measurement properties of instrument and highest COSMIN rating	44 (63)	24 (75)	14 (74)	3 (75)	3 (20)
Section 14: Online example	26 (37)	13 (41)	6 (32)	2 (50)	5 (33)

\* Response options were extremely important, very important, moderately important, slightly important, and not at all important

Multiple choice questions assessing understanding of instrument measurement properties and core outcome sets for research

Table 4.

Questions with correct responses bolded	All Panel Members (n=70)	Clinical Researchers (n=32)	Clinicians/Professional Assoc. (n=19)	US Federal Research Funding Organizations (n=4)	Patients and Caregivers (n=15)
"A study reported "Poor" reliability (reliability is the degree to which a measure produces comparable and consistent results) of a muscle strength measure, but the study had an "Excellent" COSMIN rating (COSMIN is used to rate a study's evaluation of the measurement properties of the instrument. COSMIN does NOT rate the instrument itself, but helps readers understand if they can have confidence in the results of studies evaluating measurement properties of surveys and tests.) Please select the statement that is TRUE."	5 (7)	2 (6)	1 (5)	0 (0)	2 (13)
The measure has poor reliability and the study had poor methods in evaluating and reporting the reliability of the measure					
<b>The measure has poor reliability, but the study had excellent methods in evaluating and reporting the reliability of the measure</b>	59 (84)	27 (84)	17 (89)	4 (100)	11 (73)
The measure has excellent reliability and the study had excellent methods in evaluating and reporting the reliability of the measure	3 (4)	1 (3)	0 (0)	0 (0)	2 (13)
The measure has excellent reliability, but the study had poor methods in evaluating and reporting the reliability of the measure	3 (4)	2 (6)	1 (5)	0 (0)	0 (0)
"A hypothetical Core Outcome Set has 7 Core Domains, with 1 measurement tool recommended for each Domain. For researchers designing new studies in this field, what is the minimum number of measurements tools that should be used in their study?"					
At least 1 of the 7 measurements tools from the core outcome set	12 (17)	5 (16)	6 (32)	1 (25)	0 (0)
At least 4 of the 7 measurements tools from the core outcome set	13 (19)	7 (22)	4 (21)	1 (25)	1 (7)
Only the 7 measurements tools from the core outcome set (i.e., no other measures should be used)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
<b>At least all 7 of the measurements tools from the core outcome set (i.e., additional measures may also be used)</b>	20 (29)	12 (38)	5 (26)	2 (50)	1 (7)
I am not sure	25 (36)	8 (25)	4 (21)	0 (0)	13 (87)