PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Which outcomes are most important to measure in patients with COVID-19 and how and when should these be measured? Development of an international standard set of outcomes measures for clinical use in patients with COVID-19: a report of the International Consortium for Health Outcomes Measurement (ICHOM) COVID-19 working group
AUTHORS	Seligman, William ; Fialho, Luz; Sillett, Nick; Nielsen, Christina; Baloch, Farhala M; Collis, Philip; Demedts, Ingel KM; Fleck, Marcelo P; Floriani, Maiara A; Gabriel, Lucinda EK; Gagnier, Joel; Keetharuth, Anju; Londral, Ana; Ludwig, Ingvar IL; Lumbreras, Carlos; Moscoso Daza, Alejandro; Muhammad, Nasreen; Nader Bastos, Gisele A; Owen, Christine W; Powers, John; Russell, Anne-Marie; Smith, Michaela K; Wang, Tracy; Wong, Evan K; Woodhouse, Douglas C; Zimlichman, Eyal; Brinkman, Kees

VERSION 1 – REVIEW

REVIEWER	Horstman, Molly Center for Innovations in Quality,Effectiveness, and Safety (152)
REVIEW RETURNED	28-Apr-2021

GENERAL COMMENTS	This manuscript addresses a timely and important issue, namely the development of a standard set of outcome measures for COVID-19. This set of measures adds to the literature by including patient-reported outcomes, which is of particular interest in patients with COVID-19 given the prevalence of long-term symptoms. The methods closely follow the methods used for other measure sets developed by ICHOM. The manuscript is well written and easy to follow. After reading the manuscript, I have the following questions/comments: 1. Pg 8, line 54-55 – I could not find the search strategy included as a supplement to the article 2. Pg 10, line 8 – Pg 9 states that the inclusion of outcomes in the standard set required an 80% consensus, but on pg 10 it states that a 70% consensus was used for measures and case-mix variables. What was the rationale for changing the consensus threshold? Was this determined in advance? 3. Pg 10, line 38-39 – It would be helpful to know more about where the initial 51 measures came from, e.g. how many from the literature review, patient surveys, etc. At the very least, it would be interesting to know the source of the 13 final outcomes (e.g. the

	literature review, clinical trials, patient survey, existing ICHOM measure sets). 4. Pg 11, Table 1 – after reviewing the FLU-PRO measure citation, it is unclear how the list of symptoms would be used for the outcomes “meeting criteria for critical care admission” and “disease course severity”; based on the definitions provided, these seem like clinician-reported measures. 5. Pg 14, line 36-37 – it is unclear to me how these measures would help patients understand when to seek care. How do you see these measures actually guiding patient decision-making? 6. Pg 21, Figure 2 –from Table 1, cognitive status is a clinician measure, but on the figure, it is included in the PROM package. Will PROM be used to determine cognitive status? It is also unclear where FLU-PRO fits on the measurement timeline based on Figure 2.

REVIEWER	Hermelijn, Sergei Erasmus MC Sophia Children Hospital, Pediatric Surgery
REVIEW RETURNED	10-May-2021

GENERAL COMMENTS	The authors highlight an important issue in current studies describing a large variety of outcomes making comparison between studies difficult. They suggest utilisation of a standard set of outcome measures in order to follow-up patients recovery and measure treatment strategy outcomes. They have provided a succinct and clear list of outcomes relevant to the patient group as well as provided adequate definitions, suggestions for measurement instruments and measurement timelines. The manuscript is well written with a pleasant structure. I have some minor remarks and suggestions to make the manuscript more complete: Introduction - Line 8 add " and" - In the aim please specify that this study specifically pertains patient-centred clinical outcomes as is stated in your abstract. Methods - The composition of the working group is well described, however I suggest adding how the WG was established eg how were group members selected? Additionally, which type of clinicians were part of the group as this may have influenced which clinician measures have been added. - Please describe how outcome measures were identified and extracted from the literature review. - In figure 1 please number the four phases of the standard set development as it is easily confused with the calls. - It is unclear in what stage the Delphi process was performed, was a seperate three stage delphi done for identifying outcome domains and identifying measurement tools? Please state this more clearly in your text and figure.

	- Please specify the length of the rounds of the Delphi process. - Please address attrition rate and attrition bias in the rounds of the Delphi process. - Please state why a different cut-off for consensus (70%) was used for measures and case-mix variables. Results - Are the results of the literature review available? - Please supply the initial list of 51 outcome variables as well

REVIEWER	Salustri, Francesco UCL, Institute for Global Health
REVIEW RETURNED	13-May-2021

GENERAL COMMENTS	This paper is structured as a report and differ from standard academic articles. However, it addresses an interesting topic, i.e. providing a internationally standardised set of outcomes for clinical care of patients with SARS-CoV-2. The methodology has the acknowledged limitation regarding the selected sample to identify the set of outcomes. I am not against a publication of articles like this one in BMJ Open, even though I think the best outlet for report-like articles would be a commentary or other forms different from a research article. This is an editorial choice, though.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1	Response from Authors
Search strategy missing	The MEDLINE search strategy has now been included in the 'Methods' section
Threshold for inclusion of outcomes and case-mix variables	As for all ICHOM Standard Sets, the 70% threshold is used for all decisions except within the Delphi rounds for the selection of outcomes, where a more stringent method is required and where a Likert scale is used, and 80% of respondents must vote an outcome as 'essential to include' for it to be included in the final list of outcomes. The 70% consensus level is thought to be sufficient for the selection of outcome measures and case-mix variables. This explanation is now included in the manuscript.
Source of the initial 51 measures and of the final 13 outcomes	We have summarised the sources of different outcomes in Supplementary File 2. There was significant overlap between the sources, and the Working Group discussed merging and renaming outcomes even after voting, so it is not possible to define precisely the source of most of the outcomes on the final list. This, we would argue, is a positive sign that these are meaningful outcomes as they were prioritised by patients as well as evidence-based guidelines and other ICHOM Standard Set (i.e., experts and patients with chronic conditions and an interest in health-related quality of life).
Clarification of the use of FLU-PRO for critical care admission and disease course severity	While the FLU-PRO can be used as an adjunct in the measurement of disease course severity, the reviewer is correct that it does not assist with whether or not a patient met criteria for critical care admission, and as such

	this has been changed in the manuscript to 'Clinician Measure' which is more appropriate.
How measures would help patients understand when to seek care	While, of course, the clinical outcomes would not help patients in terms of their decision-making, it is anticipated that identification of issues which patients may not realise at the time are problems e.g., mental health symptoms, or productivity, may trigger patients to seek support that they may not otherwise access.
Clarification of determination of cognitive status outcome	The 'clinician measure' part of this is in determining whether the patient is cognitively able to understand and, with or without assistance, answer the PROMs. For the duration that the patient is unable to report their own outcomes, then only the clinician-reported and administrative measures are to be used when scheduled. Cognitive function is addressed briefly in PROMIS Global Health 1.2 and so a patient's own assessment of their 'ability to think' will be measured in this questionnaire. In Figure 2, particular mention is made of the cognitive assessment in order to highlight that the clinician must determine whether the patient is cognitively able to answer the PROMs before the administration of the PROM package at each time point on the timeline.
FLU-PRO on measurement timeline	The FLU-PRO is to be completed fortnightly for the first month and then monthly thereafter. It is designed to be completed by patients themselves and, as such, is counted as part of the 'PROM package' depicted in Figure 2

Reviewer 2	
Line 8 typo	This has been fixed – many thanks
Inclusion of patient-centred clinical outcomes in aim	The specific focus on patient centred outcomes has now been highlighted
Further detail on the establishment & make-up of the WG	Details on the recruitment for the WG have now been included in the 'Methods' section, and the make-up of the WG in the 'Results' section.
Further detail on how outcome measures were identified and extracted from the literature	This has now been included in the paragraph entitled 'Identification of potential outcomes, outcome measures and case-mix variables' where specific reference is made to the extraction of outcome measures
Numbering of phases in Figure 1	This is a very helpful suggestion and we have amended Figure 1 to ensure it is clearer.
Further detail around the Delphi process, including attrition rate and attrition bias	Further detail on the Delphi process has been included in the 'Methods' section and in the 'Results' section on the attrition rate. We have also included the possibility of attrition bias in the section on limitations in the 'Discussion' section, although we feel this is unlikely to have had a large impact given that we did not see much attrition at all.
Threshold for inclusion of outcomes and case-mix variables	This is now addressed in the manuscript. A different consensus-gathering process requiring 70% consensus from the WG for each item was used to agree on which measures and case-mix variables should be recommended in line with the methodology used in all ICHOM standard sets for this part of the process. The 70% consensus level is thought to be sufficient for the selection of outcome measures and case-mix variables whereas a more stringent threshold of 80% or more of the WG voting an outcomas 'essential to include' on the Likert scale is required in ICHOM methodology for the selection of the outcomes themselves.

Source of the initial 51 measures and of the final 13 outcomes	Please see above
Length of the rounds of the Delphi process	WG calls were held 2 weeks apart and members given 1 week to respond to each survey. These details are now included in the ‘Methods’ section