PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Development and appropriateness of a scoring method for International Classification of Functioning, Disabilities, and Health assessment in older patients with heart failure: a Delphi survey of expert panel in Japan
AUTHORS	Shiota, Shigehito; Kitagawa, Toshiro; Goto, Naoya; Fujisita, Hironori; Tamekuni, Yurika; Nakayama, Susumu; Mio, Naoki; Kanai, Kana; Naka, Makiko; Yamaguchi, Mizuho; Mochizuki, Mariko; Ochikubo, Hiroyuki; Hidaka, Takayuki; Yasunobu, Yuji; Nakano, Yukiko; Kihara, Yasuki; Kimura, Hiroaki

VERSION 1 – REVIEW

REVIEWER	Roberti, Javier Universidad Nacional de Entre Rios
REVIEW RETURNED	05-Mar-2022

GENERAL COMMENTS	Thank you for the opportunity to review this paper. I have some comments that could improve clarity for readers. In the abstract, rewrite second sentence. The sentence that starts with Tertile... does not add much information to the reader who has not read the article. Bear in mind that what was evaluated is not fully introduced in the abstract. In the abstract, the conclusion is too similar to the objective. Introduction When you mention the amount in yens, it would be useful to add the amount in international dollars or USD. Page 7, line 57, when you say that it is not widely used in clinical practice, why is that? Is it possible to add information on that point? It would be interesting to establish the problem. PAge 8, line 10, this is not very clear. This is an important point. The ICF could be introduced and explained either here in the introduction or in methods. For those not familiar with it, the objective seems a bit obscure. You say: to utilise an assessment based on ICF (a tool to assess HF patients functionality?) it's necessary to develop guidelines for the assessment... "Assessment" is mentioned too often and it is not clear what the expert group had to validate. check the paper title, there, the idea is clear I think. Method How do you define "expert", please add some information to have a better idea on how you selected participants. Please remember that this is critical as the acceptability of your results may be associated to the expertise level of those who participated in the process.

	Page 9, line 30. The three questions are not clear. They also appear in line 46 with different wording. This is a very important point. The questions or statements or dimensions assessed should be very clear for the reader. Also, it would be useful if you could add part of the questionnaire as supplemental material. Perhaps the problem is that these were not questions but items. You have to find the way to describe very clearly what the group had to evaluate. In this explanation, you should include the categories, of course. For the second round, did you provide those items or categories with no agreement in the first round? what was different in this round? This is not explained. When you describe the medians I guess to refer to the total medians of each category, but in the table, you show three medians for each category. Perhaps you can add a column with this medians. When you mention tertile 7-9 you add %, why? is it a typo? I understand you refer to the likert scale of appropriateness not to any percentage. Discussion The first paragraph is not very informative, it should present the main findings very clearly. You use b455 and walking and this is too specific for this section and also please remember that not all readers are so familiar with the instruments. Please correct several punctuation errors throughout the text. Please check that there are several stops in the wrong position before reference numbers, or stops followed by small letters. Also, don't start a sentence with a number (line43). Page 25. Check grammar of figure that describes Round 1 "an gratuities" Hope this helps in improving your manuscript. Good work!

REVIEWER	Li, Shanqun
REVIEW RETURNED	16-Mar-2022

GENERAL COMMENTS	1.The background introduction only explained the importance of comprehensive assessment of the elderly with heart failure, and did not explain clearly why the ICF scoring rules were developed and validated for validity. 2.The article does not clearly describe how to find the corresponding ICF assessment rules from the final included literature.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comments

1. In the abstract, rewrite second sentence.

The sentence that starts with Tertile... does not add much information to the reader who has not read the article. Bear in mind that what was evaluated is not fully introduced in the abstract.

Response:
Based on reviewers' comments, we have revised our description of methods and results to make the research methodology easier for readers to understand.

Measures: We conducted a literature review of ICF linking rules and developed a questionnaire on scoring methods linked to ICF categories in older people with HF. In the Delphi rounds, we sent the expert panel a questionnaire consisting of three questions for each of the 43 ICF categories (1) explanation of the scoring, (2) existing assessment batteries linked to the ICF categories, and (3) how to score the existing assessment batteries. The expert panel responded to the questionnaire items on a 1 (very inappropriate) – 9 (very appropriate) Likert scale and repeated rounds until a consensus of 'Appropriate' and 'Agreement' was reached on all items.
Results: A total of 21 panel members responded to all the Delphi rounds. In the first Delphi round, six question items in four ICF categories did not reach a consensus of 'Agreement', but the result of our modifications based on panel members' suggestions reached to a consensus of 'Appropriate' and 'Agreement' on all questions in the second Delphi round. (Page 3, 46-56)

2. In the abstract, the conclusion is too similar to the objective.

Response:
In order to clearly explain the objectives and conclusions of the study we have added the significance of the study to the objectives and future work to the conclusions.

"Objective: The number of older patients with heart failure (HF) is increasing in Japan and has become a social problem. There is an urgent need to develop a comprehensive assessment methodology based on the common language of health care; the International Classification of Functioning, Disability and Health (ICF). The purpose of this study was to develop and confirm the appropriateness of a scoring methodology for 43 ICF categories in older people with HF.
Conclusion: The ICF-based scoring method for older people with HF developed in this study was found to be appropriate. Future work is needed to clarify whether comprehensive assessment and information sharing based on ICF contributes to preventing readmissions." (Page 3, 35-40, 57-60)

3. Introduction: When you mention the amount in yens, it would be useful to add the amount in international dollars or USD.

Response: Thank you for Reviewer 1's valuable comments. Based on the reviewer's comments, we have added USD to Yens. "In addition, cardiovascular disease accounts for 20.6% of all the cases requiring nursing care, and the annual medical costs exceed 6 trillion yen (USD 46 billion)." (Page 5, line 80)

4. Introduction: Page 7, line 57, when you say that it is not widely used in clinical practice, why is that? Is it possible to add information on that point? It would be interesting to establish the problem.

Response: We have added the following explanations to clearly explain why the ICF is not widely used in clinical practice.
"However, the ICF has not been widely used in clinical practice because of the complexity of the coding and the unreliability of the scores. [24-28]" (Page 5, 110-112)

5. Introduction: Page 8, line 10, this is not very clear. This is an important point. The ICF could be introduced and explained either here in the introduction or in methods. For those not familiar with it, the objective seems a bit obscure. You say: to utilise an assessment based on ICF (a tool to assess

HF patients' functionality?) it's necessary to develop guidelines for the assessment... "Assessment" is mentioned too often and it is not clear what the expert group had to validate. check the paper title, there, the idea is clear I think.

Response: Based on the reviewer's comments we have added a commentary on the ICF to the introduction and modified it to clearly explain the objectives of the study.
"To promote the use of the ICF in clinical practice, the World Health Organisation provides the ICF Core Set and the ICF Linking Rules. The ICF Core Set is a set of identified ICF categories for assessing a patient's special health condition or special medical background. [29] The ICF Linking Rules are a method of linking ICF categories with existing assessment methods. [30,31] The ICF core set for chronic ischaemic heart disease and the Geriatric ICF core set have already been developed, but these ICF categories are not appropriate for adaptation to older patients with heart failure. [32,33] Therefore, 43 ICF categories were selected for the comprehensive assessment of older patients with heart failure through the questionnaire survey of a multidisciplinary group of medical professionals and care professionals. [34,35] The 43 ICF categories specific to older patients with heart failure consisted of 17 body functions and one body structure, 19 activities and participation, and 6 environmental factors. However, in order to efficiently utilize ICF-based assessments in clinical practice, it is necessary to develop scoring methods linked to existing assessments." (Page 5, line 112-Page 6, line 125)

6. Method: How do you define "expert", please add some information to have a better idea on how you selected participants. Please remember that this is critical as the acceptability of your results may be associated to the expertise level of those who participated in the process.

Response:
In this study, we included professionals who mainly engaged in community care for heart failure in Hiroshima Prefecture, Japan. Five family physicians and ten care managers were recommended by the Hiroshima Care Manager Association. All five family physicians are specialists in internal medicine who engage in home visits while all ten care managers are board members of the Hiroshima Care Manager Association and leaders in their respective communities. In addition, we included in our panel 11 medical multidisciplinary professionals involved in heart failure care at specialised medical institutions recommended by the Hiroshima Heart Health Promotion Project. The 11 medical multidisciplinary members were: two cardiologists, three nurses certified in chronic heart failure nursing, two physiotherapists with registered instructors of cardiac rehabilitation, one occupational therapist with registered instructors of cardiac rehabilitation, one certified pharmacist, one nutritionist, and one social worker.
We have amended the description to clearly explain the composition of the expert panel and the rationale for its selection as follows:
"We established an expert multidisciplinary panel consisting of 26 medical and care professionals in Hiroshima Prefecture, Japan. The members of the expert committee were professionals with leadership roles in community care, all of whom have expertise in the assessment, treatment, and care of older patients with heart failure. Five home physicians and ten care managers were recommended by the Hiroshima Care Manager Association. All five home physicians are specialists in internal medicine who engage in home visits while all ten care managers are board members of the Hiroshima Care Manager Association and leaders in their respective communities. In addition, we included 11 medical multidisciplinary professionals involved in heart failure care at specialised medical institutions recommended by the Hiroshima Heart Health Promotion Project in our panel.[37] The 11 medical multidisciplinary members were: two cardiovascular physicians, three nurses certified in chronic heart failure nursing, two physiotherapists with registered instructors of cardiac rehabilitation, one occupational therapist with registered instructors of cardiac rehabilitation, one certified pharmacist, one nutritionist, and one social worker." (Page 6, line 144-Page 7, line 158)

7. Method: Page 9, line 30. The three questions are not clear. They also appear in line 46 with different wording. This is a very important point. The questions or statements or dimensions assessed should be very clear for the reader. Also, it would be useful if you could add part of the questionnaire as supplemental material. Perhaps the problem is that these were not questions but items. You have to find the way to describe very clearly what the group had to evaluate. In this explanation, you should include the categories, of course.

Response:
We agree with the reviewer's comments.
We have prepared supplementary material 1 to explain the questionnaire clearly. We have also revised the description of the 'Development of the questionnaire' and the description of the Figures to clearly explain the three questions for the 43 ICF categories.

Development of the questionnaire
“We set three questions for each of the 43 ICF categories and prepared 1 (very inappropriate) - 9 (very appropriate) Likert scale responses to assess appropriateness. Appropriateness was evaluated on a median response scale with the following three levels: 1-3 as “inappropriate”, 4-6 as “uncertain”, and 7-9 as “appropriate”. The three questionnaire items were as follows: 1) Appropriateness of the 43 ICF category scoring descriptions, 2) appropriateness of existing assessment batteries linked to each ICF categories, and 3) appropriateness of the scoring methods for each ICF categories linked to existing assessment batteries. All questionnaires were developed using a Google Form, with a description of each ICF category and the rationale for scoring. (Supplemental materials 1)”(Page 7, line 179-Page 8, line 188)

8. Method: For the second round, did you provide those items or categories with no agreement in the first round? what was different in this round? This is not explained.

Response:
In the second round we sent the same three questions in the 43 ICF categories as in the first round. However, we revised the questionnaire descriptions and scoring methods for the ICF categories that the panel members in the first round answered "Inappropriate" (1-3), "Uncertain" (4-6) or ‘Disagreement’, based on the panel's suggestions.

We have modified the description as follows to ensure that the research methodology is clearly communicated to the reader.

In addition, the results of Delphi Round 1 were added to Supplementary Material 3.
“In the first round, the HFC mailed a sheet with instructions on how to conduct the ICF category adequacy assessment, as well as the URL and QR codes for the questionnaire. The panel members responded to three questions in 43 ICF categories on a scale of 1-9. In addition, panel members provided open-ended suggestions for improvements to the questions they scored 1-6. The HFC collated the panel members’ responses. We revised the scoring descriptions and existing assessment batteries linked to the ICF categories responded to as ‘Inappropriate’, ‘Uncertain’ or ‘Disagreement’ based on the panel’s suggestions. The definition of ‘Disagreement’ in this article is given in Analysis. In the second round, the HFC emailed the revised questionnaire and feedback based on the panel members' responses. As in the first round, the panel members rated the appropriateness of three question items in the 43 ICF categories. In addition, the panel members provided suggestions for improvements to the scoring methods on those onesscored 1-6. The HFC compiled the panel members' responses and assessed their appropriateness. We also revised the descriptions of the questionnaire or scoring methods based on the panel's suggestions. The revised questionnaire was

emailed to the panel members, and a final consensus was reached after confirming that there were no comments for revision." (Page 8, lines 195-211)

9. Method: When you describe the medians I guess to refer to the total medians of each category, but in the table, you show three medians for each category. Perhaps you can add a column with this medians.

Response:
Thank you for the beneficial review comments. In this study, we asked our expert panel to respond to three questions for each of the 43 ICF categories on a 1-9 scale of appropriateness.

In order to clearly communicate the findings of the study to the reader, we have added 'three questions' to the notation in Table 2 and amended the headings within the table.

10. Method: When you mention tertile 7-9 you add %, why? is it a typo? I understand you refer to the likert scale of appropriateness not to any percentage.

Response:
In this study, following the RAND/UCLA Appropriateness User's Manual, the definition of 'agreement' was defined as 80% or more of the respondents' medians being within the same Tertile (1-3, 4-6, 7-9) as the median respondent. The Tertiles in Table 2 showed the percentage of respondents within the same Tertile as the median respondent.

However, as the reviewer commented, our description was not clear. We provide a clear explanation of the definitions of 'agreement' and 'disagreement' and the order to which we have amended the description as follows. In addition, we have revised the description of 'Tertile (%)' to 'Number of outside median tertile (/21)' in Table 2.

" In accordance with the RAND/UCLA guidelines, we defined 'Agreement' or 'Disagreement' according to the number of panellists who rated outside the range of tertiles (1-3; 4-6; 7-9) including the median. 'Agreement' was defined as fewer than one-third of panellists who rated outside the range of the tertile values. 'Disagreement' was defined as when more than one-third of panellists rated the extremes (1-3 range and 7-9 range) not including the median. "(Page 8, line 218-Page 9, line 223)

11. Discussion: The first paragraph is not very informative, it should present the main findings very clearly. You use b455 and walking and this is too specific for this section and also please remember that not all readers are so familiar with the instruments.

Response:
As the reviewer commented, our Discussion write-up did not provide enough information for the reader. In order to clearly explain the findings in this study, we have modified our description as follows:

"We have developed a comprehensive assessment for older people with heart failure based on ICF for widespread use in clinical practice and verified the appropriateness of the scoring method using the RAND Delphi method. In this study, we drew on our literature review and the ICF Reference Guide to link existing assessment batteries for 28 of the 43 ICF categories. In the first Delphi round, 'agreement' was not reached on six questions in the four ICF categories, and the explanation and scoring methods were modified. In the second round of Delphi, all question items of the 43 ICF category were reached to a consensus of 'Appropriate' and 'Agreement'" (Page 14, line 318-Page 15, line 325)

12. Discussion: Please correct several punctuation errors throughout the text. Please check that there are several stops in the wrong position before reference numbers, or stops followed by small letters. Also, don't start a sentence with a number (line43).
Response:
Based on the reviewer's comments, we have checked the text of the discussion and corrected the comma after the reference number. In addition, we have revised the text beginning with the number.

13. Figure: Page 25. Check grammar of figure that describes Round 1 "an gratuities"

Response: Based on the comments, we have checked the grammar of "an gratuities" in the diagram of Figure and revised the description.

“We sent the expert panel a questionnaire and a gratuity.”

Reviewer 2 comments
1.The background introduction only explained the importance of comprehensive assessment of the elderly with heart failure, and did not explain clearly why the ICF scoring rules were developed and validated for validity.

Response:
We have added the following explanation to clearly explain the background of our study - the need for an ICF-based scoring method for older people with heart failure.
“However, the ICF has not been widely used in clinical practice because of the complexity of the coding and the unreliability of the scores. [24-28] To promote the use of the ICF in clinical practice, the World Health Organisation provides the ICF Core Set and the ICF Linking Rules. The ICF Core Set is a set of identified ICF categories for assessing a patient's special health condition or special medical background. [29] The ICF Linking Rules are a method of linking ICF categories with existing assessment methods. [30,31] The ICF core set for chronic ischaemic heart disease and the Geriatric ICF core set have already been developed, but these ICF categories are not appropriate for adaptation to older patients with heart failure. [32,33] Therefore, 43 ICF categories were selected for the comprehensive assessment of older patients with heart failure through the questionnaire survey of a multidisciplinary group of medical professionals and care professionals. [34,35] The 43 ICF categories specific to older patients with heart failure consisted of 17 body functions and one body structure, 19 activities and participation, and 6 environmental factors. However, in order to efficiently utilize ICF-based assessments in clinical practice, it is necessary to develop scoring methods linked to existing assessments.” (Page 5, line 110-Page 6, line 125)

2.The article does not clearly describe how to find the corresponding ICF assessment rules from the final included literature.

Response:
Based on the reviewers' comments, we have revised the description to clarify the process of developing the questionnaire through a literature review. The revised description is shown below. In addition, we have presented the results of our literature review in Supplementary Material 2.

“In the qualitative analysis, we excluded 19 references dealing with disease-specific assessment batteries that could not be adapted to older patients with heart failure (e.g., stroke, musculoskeletal disease, hand surgery, low back pain). Eight articles on ICF linking rules were included. Finally, we

employed 11 existing assessment batteries on eight articles links to the 43 ICF categories (Supplemental material 2). [40-47] Eleven existing assessment batteries were included: assessment of ADL (such as Functional Independence Measure (FIM) and Barthel Index), assessment of general health-related quality of life (such as Short Form 36 and the European Quality of Life instrument (EQ-5D), The World Health Organization Quality of Life (WHOQOL)), assessment of general health status (such as the Nottingham Health Profile(NHP), the World Health Organization Disability Assessment Schedule (WHODAS 2.0)), and assessment of falls (such as Falls Efficacy Scale-International (FES-I), the Swedish version of the Falls Efficacy Scale (FES[S]), the Activities-specific Balance Confidence Scale (ABC), and the modified Survey of Activities and Fear of Falling in the Elderly (SAFFE)). We identified these existing assessment batteries as linked to 20 of the 43 categories. However, only the FIM and BI were employed in the questionnaire, as they did not match the objectives of this study for the assessment of general health-related quality of life, general health status and falls. Therefore, we developed a scoring methodology for ICF categories other than ADL, based on the Italian ICF Guidelines and the ICF Reference Guide. [38,39, 48] Finally, we decided to provide 30 existing assessment batteries linking to ICF categories, and to score the remaining 13 categories using only the scoring descriptions (Table 2). “ (Page 10, line 252-273)

VERSION 2 – REVIEW

REVIEWER	Roberti, Javier Universidad Nacional de Entre Rios
REVIEW RETURNED	06-Jun-2022

GENERAL COMMENTS	The authors have improved the manuscript. Good work! There remain some very minor issues to correct before acceptance that Im sure the authors will be able to address. Page 55, line 201. please check that you are describing the categories correctly, aren't they appropriate, inappropriate and uncertain based on the score itself 1 to 9 and agreement and disagreement based on the dispersion of scores? This line is confusing. Page 55, line 206, Please clarify what panel members had to score in round 2, did they score everything again or just those items that had not reached consensus in the previous round? Page 57, line 248, some of the details of the review process are not necessary here as they can be seen in the PRISMA graph, page 57, line 267, sentence starting with "however....". this is not clear, check grammar page 58- please check first paragraph there are repeated sentences. page 61. line 324. please check grammar in "were reached to...". Also, please correct UCLS to UCLA. Hope these comments help to improve this manuscript.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Comments to the Author:
1. Page 55, line 201. please check that you are describing the categories correctly, aren't they appropriate, inappropriate and uncertain based on the score itself 1 to 9 and agreement and disagreement based on the dispersion of scores? This line is confusing.

Response: Based on reviewer's comment, we have moved the definitions of the three levels of appropriateness ratings ('Appropriate', 'Uncertain', and 'Inappropriate') and the agreement ratings ('Agreement' and 'Disagreement') under the Delphi process and funding consensus section to improve logical organization.

2. Page 55, line 206, Please clarify what panel members had to score in round 2, did they score everything again or just those items that had not reached consensus in the previous round?

Response: Based on the reviewer's comment, we have revised the manuscript to reflect our intended meaning. For round 2, the panel members scored all items again.

Revision:
"As in the first round, the panel members scored again the appropriateness of three of the question items in all 43 ICF categories." (Page 8, line 214-215)

3. Page 57, line 248, some of the details of the review process are not necessary here as they can be seen in the PRISMA graph,

Response: Based on the reviewers' comment, we have deleted the details on the inclusion and exclusion processes and retained only the information on the number of references included.

Revision:
"Following a two-stage screening process, we conducted a qualitative analysis of 26 references." (Page 10, line 249-250)

4. page 57, line 267, sentence starting with "however....". this is not clear, check grammar

Response: Based on the reviewer's comment, we have rewritten Line 257 for clarity and language.

Revision:
However, we employed only the FIM and the BI. We did not include assessment batteries for general health-related quality of life, general health status, and falls in the questionnaire because these were not consistent with the aims of this study. (Page 10, line 265-267)

5. page 58- please check first paragraph there are repeated sentences.

Response: We have reviewed the text and accordingly removed any redundant sentences.

6. page 61. line 324. please check grammar in "were reached to...".
Also, please correct UCLS to UCLA.

Response: Based on the reviewer's comment, we have corrected the grammar of the relevant line

and corrected UCLS to UCLA.

Revision:
“This study was based on the RAND/UCLA Delphi method, but face-to-face meetings could not be conducted because of the current coronavirus pandemic.” (Page 15, line355)