PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The diagnostic and predictive accuracy of the Clinical Frailty Scale among hospitalised older medical patients: a systematic review and meta-analysis protocol
AUTHORS	Leahy, Aoife; O'Connor, Margaret; Condon, Jennifer; Heywood, Sarah; Shanahan, Elaine; Peters, Catherine; Galvin, Rose

VERSION 1 – REVIEW

REVIEWER	Michaël Laurent Imelda Hospital Bonheiden, Belgium
REVIEW RETURNED	23-Jul-2020

GENERAL COMMENTS	This protocol for a systematic review and meta-analysis aims to investigate the diagnostic properties of the clinical frailty scale (CFS) for frailty and its ability to predict subsequent adverse outcomes in hospitalized older adults. The protocol is generally well written and the methods are appropriate. However the research question is challening and I see the following specific points: MAJOR COMMENTS: 1/ The authors plan to compare the CFS both against the reference standard of the Frailty Index, which however cannot be considered "golden standards" necessarily. They also plan to assess the predictive ability for adverse outcomes in hospitalized older adults. Why not compare the CFS against other instruments in predicting adverse outcomes, wouldn't that give a more accurate idea of the "best" frailty instrument? 2/ Why would you exclude the intervention arms of any RCTs? The diagnostic accuracy can still be judged from the baseline of an intervention arm of an RCT? Also, would an intervention in an RCT necessarily dilute the predictive ability of the CFS? Maybe these data would need to be analyzed separately but since RCTs are arguably the highest level of evidence, I find it strange to simply exclude intervention arms. 3/ In the abstract you state that "A cut-off score of >4 will be used to identify frailty." (I would say "A CFS score of >4 will be used to identify frailty.", because you speak about the QUADAS-2 just before that so this is somewhat confusing). But more generally: a lot of studies investigate associations using the CFS as a continuous orthogonal scale and e.g. Cox proportional hazard models, whereas other studies analyze it in a categorical

	manner (>4 or not). If I read the abstract you plan mainly a categorical analysis, but potentially this will make it very difficult to analyze the majority of studies likely using continuous analyses? How will you deal with this problem? 4/ The outcomes of adverse events are described in insufficient detail. Falls for example, may be analyzed as time to first fall, rate of falls or number of falls. Can you be more specific a priori, or will you just see what is available in studies and pool what is possible? 5/ What about other important adverse outcomes like fall-related injuries, fractures, cognitive decline, length of hospital stay, delirium, pressure ulcers etc.? Wouldn't it be possible to collate any additional secondary outcomes that you find, apart from the main ones you have defined, and summarize those as secondary outcomes? 6/ The authors have not fully followed the PRISMA checklist. Will you attempt to contact corresponding authors in case insufficient details are available in any studies? Who is the guarantor of this review? Will you use GRADE to evaluate the overall confidence in the evidence that arises from your meta-analysis? Will you assess publication bias? Which data items will be extracted from the full-text reports? Only fully published studies or also conference abstracts etc.? Heterogeneity by I^2? (these points were not specified in the protocol although they are part of the PRISMA checklist, please revise and submit a new version of the checklist) MINOR COMMENTS: 7/ In the abstract, "This systematic review and meta-analysis examines" should be "aims to examine", because you are not reporting the results of the meta-analysis yet. 8/ I also suggest to change this sentence, both in the abstract and on page 4 (end of introduction) to "This systematic review and meta-analysis aims to examine, in hospitalized adults ≥65 years, 1) …" because the target population of HOSPITALIZED older adults is not so clear in the aims. 9/ In the abstract and on page 4, you write "to predict adverse outcomes in hospitalised older adults". However, "unplanned ED visit" is something which happens after discharge, not during hospitalization. So I would suggest something like "the predictive validity of the CFS in hospitalised older adults to predict subsequent adverse outcomes." P.S.: note that the abbreviation "ED" needs to be defined upon first use, best to avoid abbreviations in the abstract. 10/ I think it is inappropriate that you predefine your ultimate target journal as a peer-reviewed "ageing" journal. You cannot know for certain where this work will ultimately be published, so I suggest to just state "peer-reviewed journal".

REVIEWER	Jai Darvall University of Melbourne, Australia
REVIEW RETURNED	23-Aug-2020

GENERAL COMMENTS	Leahy and colleagues describe an interesting proposal for a systematic review and meta-analysis examining the predictive ability of the Clinical Frailty Scale when compared to "reference" tools- the

	frailty index and phenotypic model. This is a worthy investigation, however I have some concerns: I am not sure why restricting the inclusion populations to "hospitalised medical patients" is necessary? What about surgical/trauma/intensive care unit populations? This is unnecessarily restrictive. The discussion paragraph could perhaps be reworked. It gives some information regarding the Comprehensive Geriatric Assessment, but this is not well synthesised with the rest of the manuscript. There is a fundamental issue which is not well addressed in the protocol. How do the authors propose to report/resolve the discrepancy in comparing individual studies' CFS scores with different "reference standard" tools (namely the FI and phenotypic score) when these belong to different paradigms of frailty? As the deficit model and phenotypic model measure different constructs of frailty, it is not meaningful to report "sensitivity and specificity" for the CFS as a screening test against these different scores. It would be more meaningful to choose one, and study the correlation against that "gold standard" measure (I would recommend the FI- deficit model). Otherwise how do the study authors propose pooling these quite different things in a meta-analysis. There is also insufficient detail on the proposed methods of meta-analysing results. On a related issue, more detail is required in the inclusion criteria about how studies will be selected. For example, will the 5-item Fried phenotype score required to be measured for inclusion in this arm of the study? It is also insufficient to state studies measuring the "reference standard Frailty Index" will be included- there are myriad different frailty indices in the literature. Perhaps the authors could reference the work of Searle et al in pre-specifying some preconditions which must be met in the construction of a valid frailty index to warrant inclusion in this systematic review. The comment: "We will include patient and public involvement in the dissemination of our systematic review" is not detailed enough. How will individual patients and/or the public be involved in this dissemination. This statement requires clarification, it is not simply a tickbox. Table 1 should be corrected, it inaccurately dichotomises CFS scores 1-4 as "robust" and 5-7 (or 5-9) as "frail". A better characterisation would be as "frail" and "non-frail" (consistent with the original definitions of Rockwood et al, eg a CFS = 4 is "vulnerable", not "robust").

VERSION 1 – AUTHOR RESPONSE

MAJOR COMMENTS:

1/ The authors plan to compare the CFS both against the reference standard of the Frailty Index, which however cannot be considered "golden standards" necessarily.

They also plan to assess the predictive ability for adverse outcomes in hospitalized older adults. Why not compare the CFS against other instruments in predicting adverse outcomes, wouldn't that give a more accurate idea of the "best" frailty instrument?

There are a multitude of screening tools for frailty. Operationalising frailty remains controversial however it is accepted in the literature that both the phenotypical and deficit models are appropriate methods to diagnose this condition, while one could argue that these are not gold standard tests. In acknowledgement of this, we have deliberately used the term 'reference standard' for the Frailty phenotype (phenotypic model) and Frailty Index (deficit model) as opposed to gold standard in the manuscript.

The reviewer makes an excellent point that one could assess the predictive accuracy of any of the many frailty screening tools outlined in the literature. We have chosen to review the CFS as it is a widely used instrument in clinical practice and in the literature. Broadening the review to incorporate multiple screening tools would make the review very unwieldy.

2/ Why would you exclude the intervention arms of any RCTs? The diagnostic accuracy can still be judged from the baseline of an intervention arm of an RCT?
Also, would an intervention in an RCT necessarily dilute the predictive ability of the CFS? Maybe these data would need to be analyzed separately but since RCTs are arguably the highest level of evidence, I find it strange to simply exclude intervention arms.

Apologies for the oversight. The baseline data from RCTs will be included for DA studies. We will exclude the intervention arm of trials for the predictive accuracy component of the review as the intervention/exposure may impact the incidence of adverse outcomes.

3/ In the abstract you state that "A cut-off score of >4 will be used to identify frailty." (I would say "A CFS score of >4 will be used to identify frailty.", because you speak about the QUADAS-2 just before that so this is somewhat confusing).

This has been changed in the manuscript as per the reviewer suggestion. (Page 3 Line 22)

But more generally: a lot of studies investigate associations using the CFS as a continuous orthogonal scale and e.g. Cox proportional hazard models, whereas other studies analyze it in a categorical manner (>4 or not). If I read the abstract you plan mainly a categorical analysis, but potentially this will make it very difficult to analyze the majority of studies likely using continuous analyses? How will you deal with this problem?

This is a very useful point. The focus of this review is to examine the diagnostic and predictive accuracy of the CFS. By its nature, the accuracy of a tool is a dichotomous outcome and the CSF score of >4 has been identified in the literature to 'diagnose' people with frailty. In studies where the data are presented as a continuous measure, we will email the authors for additional data if the information for 2x2 tables at a cut-off of >4 cannot be readily extracted. (Page 7, Line 28-29)

4/ The outcomes of adverse events are described in insufficient detail. Falls for example, may be analyzed as time to first fall, rate of falls or number of falls. Can you be more specific a priori, or will you just see what is available in studies and pool what is possible?

The outcomes of interest have been clarified in the revised manuscript. (Page 6, Line 32-34)

5/ What about other important adverse outcomes like fall-related injuries, fractures, cognitive decline, length of hospital stay, delirium, pressure ulcers etc.?
Wouldn't it be possible to collate any additional secondary outcomes that you find, apart from the main ones you have defined, and summarize those as secondary outcomes?

We have accepted the reviewer's comments and updated the manuscript. (Page 6, Line 30-32)

6/ The authors have not fully followed the PRISMA checklist.

The PRISMA guidelines were originally designed to enhance the transparency of the conduct and reporting of systematic reviews of RCTs. However, we have followed the PRISMA guidelines in so far as possible for the protocol of this review of DA and predictive accuracy studies.

Will you attempt to contact corresponding authors in case insufficient details are available in any studies?

We will contact corresponding authors if insufficient data. This is documented in the methods section. (Page 7, Line 28-29)

Who is the guarantor of this review?
Rose Galvin (Page 1, Line 26-29)

Will you use GRADE to evaluate the overall confidence in the evidence that arises from your meta-analysis?

We will use the GRADE framework to assess the quality of the evidence for predictive and diagnostic accuracy from the meta-analysis as suggested by the reviewers. A section describing our approach to same has been included in the updated manuscript. (Page 8, Line 11-16)

Will you assess publication bias?

Yes, we will assess publication bias. This has been clarified in the updated manuscript. (Page 8, Line 8-9)

Which data items will be extracted from the full-text reports?

We will include data extraction table to show data items to be extracted.

Only fully published studies or also conference abstracts etc.?

Abstracts will be used to check for subsequent peer-reviewed publications but we will exclude such abstracts if there is no subsequent peer reviewed publication. (Page 7, Line 3-4)

Heterogeneity by I^2? (these points were not specified in the protocol although they are part of the PRISMA checklist, please revise and submit a new version of the checklist)

Statistical heterogeneity will be explored using the variance of logit-transformed sensitivity and specificity, where smaller values indicate less heterogeneity between studies. (Page 8, Line 1-2)

MINOR COMMENTS:
7/ In the abstract, "This systematic review and meta-analysis examines" should be "aims to examine",

because you are not reporting the results of the meta-analysis yet.

Updated as per the reviewers suggestions. Page 3, Line 6

8/ I also suggest to change this sentence, both in the abstract and on page 4 (end of introduction) to "This systematic review and meta-analysis aims to examine, in hospitalized adults ≥65 years, 1) …" because the target population of HOSPITALIZED older adults is not so clear in the aims.

Updated as per the reviewers suggestions. Page 3, Line 7, Page 5, Line 17-19

9/ In the abstract and on page 4, you write "to predict adverse outcomes in hospitalised older adults". However, "unplanned ED visit" is something which happens after discharge, not during hospitalization. So I would suggest something like "the predictive validity of the CFS in hospitalised older adults to predict subsequent adverse outcomes."
P.S.: note that the abbreviation "ED" needs to be defined upon first use, best to avoid abbreviations in the abstract.

Updated as per the reviewers suggestions. Page 3 Line 9

10/ I think it is inappropriate that you predefine your ultimate target journal as a peer-reviewed "ageing" journal. You cannot know for certain where this work will ultimately be published, so I suggest to just state "peer-reviewed journal".

Updated as per the reviewers suggestions. Page 3, Line 31

Reviewer: 2

Reviewer Name: Jai Darvall
Institution and Country: University of Melbourne, Australia
Competing interests: None declared

Please leave your comments for the authors below
Leahy and colleagues describe an interesting proposal for a systematic review and meta-analysis examining the predictive ability of the Clinical Frailty Scale when compared to "reference" tools- the frailty index and phenotypic model. This is a worthy investigation, however I have some concerns:

I am not sure why restricting the inclusion populations to "hospitalised medical patients" is necessary? What about surgical/trauma/intensive care unit populations? This is unnecessarily restrictive.

Our preliminary searches indicated that the CFS has been used in over 70 peer-reviewed publications to examine diagnostic and predictive accuracy across a number of populations. We are particularly interested in frailty in hospitalised older adults and have identified over 30 papers that meet our inclusion criteria. The researchers took the view that including all these populations would make the studies included too heterogenous and affect the applicability of results to our population of interest.

The discussion paragraph could perhaps be reworked. It gives some information regarding the Comprehensive Geriatric Assessment, but this is not well synthesised with the rest of the manuscript.

The discussion section has been revised to reflect the three main points that have arisen in the review:
1. The choice of reference standard
2. The choice of hospitalised older adults

3. The rationale for the use of a frailty screening tool for older adults

There is a fundamental issue which is not well addressed in the protocol. How do the authors propose to report/resolve the discrepancy in comparing individual studies' CFS scores with different "reference standard" tools (namely the FI and phenotypic score) when these belong to different paradigms of frailty?

As the deficit model and phenotypic model measure different constructs of frailty, it is not meaningful to report "sensitivity and specificity" for the CFS as a screening test against these different scores. It would be more meaningful to choose one, and study the correlation against that "gold standard" measure (I would recommend the FI- deficit model). Otherwise how do the study authors propose pooling these quite different things in a meta-analysis. There is also insufficient detail on the proposed methods of meta-analysing results.

We welcome the opportunity to clarify this point. In the first instance we will pool the totality of evidence to explore the DA of the CFS against both reference standards combined. We will then complete a sub-group analysis exploring the DA of studies using either measure. This has been clarified in the updated manuscript. (Page 8, Line 7-9)

On a related issue, more detail is required in the inclusion criteria about how studies will be selected. For example, will the 5-item Fried phenotype score required to be measured for inclusion in this arm of the study? It is also insufficient to state studies measuring the "reference standard Frailty Index" will be included- there are myriad different frailty indices in the literature. Perhaps the authors could reference the work of Searle et al in pre-specifying some preconditions which must be met in the construction of a valid frailty index to warrant inclusion in this systematic review.

Addressing the reviewers comments regarding the frailty index, we agree that each frailty index which is used will need to be assessed to determine its validity as a reference standard. Searle et al provide a useful template for assessing this. We will include all studies in our analysis, but we will acknowledge those studies where the frailty index does not adhere to the above criteria.

The comment: "We will include patient and public involvement in the dissemination of our systematic review" is not detailed enough. How will individual patients and/or the public be involved in this dissemination. This statement requires clarification, it is not simply a tickbox.

We will involve a subgroup of older adults in our local Patient and Public Involvement Group who will review the paper and provide input. (Page 8, Line 18 - 24)

Table 1 should be corrected, it inaccurately dichotomises CFS scores 1-4 as "robust" and 5-7 (or 5-9) as "frail". A better characterisation would be as "frail" and "non-frail" (consistent with the original definitions of Rockwood et al, eg a CFS = 4 is "vulnerable", not "robust").

Updated as per the reviewer's suggestion. Page 5 Line 21-22

VERSION 2 – REVIEW

REVIEWER	Michael Laurent Imelda Hospital, Belgium
REVIEW RETURNED	02-Nov-2020

GENERAL COMMENTS	The authors have sufficiently addressed all of my comments.

REVIEWER	Jai Darvall University of Melbourne, Australia
REVIEW RETURNED	05-Nov-2020

GENERAL COMMENTS	Leahy et al have prepared a considered response to the points raised in review. Whilst I still have fundamental concerns about the ability of a systematic review and mooted meta-analysis to pool the results of two very different constructs of frailty (the phenotypic and deficit models), the revised approach to analysis by examining frailty index diagnostic accuracy, and phenotypic model accuracy, separately is welcomed. I also have residual concerns about exclusion of populations other than "hospitalised medical patients". Although the authors are concerned about heterogeneity in included studies, this is exactly the point when assessing the diagnostic accuracy of a screening tool (which the CFS is). If the results are not shown to apply to populations like intensive care unit patients, trauma patients, surgical patients, then the generalisability of any findings from this review will be limited. If the authors insist on limiting their included population, the title of the manuscript must be changed. I would suggest: The diagnostic and predictive accuracy of the Clinical Frailty Scale among hospitalised older medical patients: a systematic review and meta-analysis protocol." Similarly, all reference to the population of interest in the manuscript must be changed from "hospitalised older adults over 65 years of age" to "hospitalised older adults with a medical diagnosis over 65 years of age". Specific exclusion criteria must also be present in the manuscript, such as those populations listed above. Overall, if these changes are made, I think this manuscript is acceptable.

VERSION 2 – AUTHOR RESPONSE

Reviewer 2

The title of the manuscript must be changed. I would suggest: The diagnostic and predictive accuracy of the Clinical Frailty Scale among hospitalised older medical patients: a systematic review and meta-analysis protocol."

The title has been amended as per the reviewer's suggestion (Page 1, Line 4)

Similarly, all reference to the population of interest in the manuscript must be changed from "hospitalised older adults over 65 years of age" to "hospitalised older adults with a medical diagnosis over 65 years of age".

We acknowledge the reviewer's comment and have updated the relevant sections to highlight that this review will focus on older adults with medical diagnoses. See changes as follows, (Page 3, Line 7), (Page 3, Line 16), (Page 5, Line 21-22), (Page 6, Line 4), (Page 8, Line 6), (Page 8, Line 14), (Page 8, Line 32)

Specific exclusion criteria must also be present in the manuscript, such as those populations listed above.

We have amended the manuscript to reflect that the following populations are excluded: intensive care patients, surgical patients, cardiac patients, renal patients and orthopaedic patients (Page 7, Line 10-11)

VERSION 3 – REVIEW

REVIEWER	Jai Darvall University of Melbourne, Australia
REVIEW RETURNED	09-Dec-2020

GENERAL COMMENTS	Leahy and colleagues have now addressed the remaining concerns raised in review, with the exception of semantics around "medical complaints". I would prefer to see the word medical "diagnosis" used, as mentioned in the previous review, as it is less ambiguous. This I believe can be done in the editorial process, other than this, the manuscript has now addressed my remaining concerns.