PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Translation and Testing of the Risk Assessment Pressure Ulcer Score Scale Used Among Residents in Norwegian Nursing Homes
AUTHORS	Fossum, Mariann ; Söderhamn, Olle; Cliffordson, Christina; Söderhamn, Ulrika

VERSION 1 - REVIEW

REVIEWER	Dr. Jan Kottner Scientific Director Clinical Research Charité – Universitätsmedizin Berlin Department of Dermatology and Allergy Clinical Research Center for Hair and Skin Science Berlin GERMANY
REVIEW RETURNED	28-Jun-2012

THE STUDY	Pressure ulcer (PU) risk assessment is important and therefore this work is important too. However, my major concern is that a cross sectional design was used to validate a prognostic score. Estimating the predictive validity (=probability of future events) is only possible in longitudinal (cohort) studies. A further design weakness is the convenience sample. (1) Reliability, page 5: Because PU risk factors are not necessarily related with each other, the sum score can be considered as causal indicator model (see for instance Kottner J, Streiner DL. Int J Nurs Stud. 2010 Jul;47(7):926-8). Calculating alpha and item correlations is not appropriate here because PU risk scores are not unidimensional (= homogeneous). (2) Reliability, page 6: Do not use Pearson's r for calculating associations between rater scores. For measuring interrater reliability you must use ICCCs. (3) Validity, page 6: Why do you used the BMI 23 as splitting variable for the known groups comparisons? Is there any evidence for that? (4) Validity, page 6: Do not use a principal component analysis when you want to confirm a scale structure. PCA is for selecting and deleting items to create scores. Here you must use confirmatory analysis. Above all, you cannot identify different factors and calculate an overall alpha. (5) Sensitivity, page 7: Please see above. You cannot validate a predictive risk score with the concurrent presence or absence of Pus.
RESULTS & CONCLUSIONS	The presented results and conclusions do not provide good evidence about the scale performance. Table 4, page 9: Irrespective of the questionable method you

	present 4 factors here. Please label the factors in a consistent manner: “Nutritional status” and “physical condition” are conceptually functional characteristics and thus comparable. The factor “major risk factors” is conceptually completely different. By the way, that physical activity, mobility and moisture load on one factor clearly indicates that this scale is not homogenous.
REPORTING & ETHICS	STROBE was used, but maybe STARD or GRRAS provide better frameworks for reporting.

REVIEWER	Pedro L Pancorbo-Hidalgo Professor of Clinical Nursing and Head of departament of Nursing. Faculty of Health Sciences. University of Jaén. Spain I have no conflict of interests.
REVIEW RETURNED	08-Jul-2012

THE STUDY	Overall it is an interesting manuscript, but there are some points that need further clarification or modification, in order to improve the article. The title refers to "Evaluation of the Risk Assessment .... used among residents Nevertheless the article is not about using the scale, but about the translation and validation process. I think that the actual title could confound the readers. Perhaps the authors should rewrite the title in a more descriptive way, according the aim an the results. Aim: The purpose stated was " to translate and test the psychometric properties", this include to establish reliability and validity, but not the clinical validation as a diagnostic tool (sensibility, specificity,..). Perhaps the authors should explain better the aim, matching with methods and results. Methods: The process of translation-back-translation should be explained better. Which kind of linguistic changes were done? Perhaps, the main problem is the design. With a cross-sectional design you can't test the predictive power or properties of the Scale. The Risk Assessment Pressure Ulcers Scale is intended to detect or predict patients at risk for developing pressure ulcers, so it has no sense to score people with an actual Pressure ulcer (because they have risk by definition). So if you assess the risk (using the RAPS) and the presence of PU at the same time, you are trying to use the RAPS as a diagnostic tool for PU presence (instead of direct skin inspection). Due to this, your data on sensibility, specificity and area under the ROC curve have poor reliability. Clinical validation of a risk assessment scale usually need a prospective design , assessing the patients at the beginning, following them during a period in order to check if new pressure ulcers appear. My suggestion for the authors is to remove data about sensibility, specificity and so on, in the manuscript; focusing just in the psychometrics properties of the scale. Please explain the reason for excluding people with lower extremity amputation or receiving enteral nutrition in the validation study, because these are high-risk patients so your sample could be

	biased. Who did the assessment with the scale and the skin inspection? The same person? Reliability. Homogeneity or internal consistency. It seems that authors propose a 3-factor structure for the RAPS, so you should provide not only the global Cronbach's alpha, but the figures for each factor (the authors do in table 4, but should refers them in the previous paragraph about homogeneity).
RESULTS & CONCLUSIONS	Construct validity. Your data about validity in “known groups” (table 3) need interpretation. There are differences in the mean score between groups, but if you take into account the cut-off point there is no clinical differences between groups (in BMI and CC groups), because both of them score more than cut-off point (31 p.), so both groups are “no risk groups”. Authors should explain better these facts and consider as limitation. Limitations. The authors should consider as a limitation that their research do not estimate the concurrent validity using a well validated scale, like Braden scale, in order to check if this new scale classify the patients in a similar way. Is there a Norwegian version of Braden scale?
GENERAL COMMENTS	Thank you for this opportunity to review this manuscript and collaborate in this open access journal. This paper deals with an important issue in nursing care, such as pressure ulcers prevention. Using a risk assessment tool properly validated could be a way to improve preventive care, although there isn’t consensus between experts. So it is important to develop research like this piece, with validation data.

VERSION 1 – AUTHOR RESPONSE

Reviewer Comments

Answers to reviewers’ questions and changes in the manuscript
Reviewer 1.
However, my major concern is that a cross sectional design was used to validate a prognostic score. Estimating the predictive validity (=probability of future events) is only possible in longitudinal (cohort) studies. A further design weakness is the convenience sample.

Text is added page 12.

(1) Reliability, page 5: Because PU risk factors are not necessarily related with each other, the sum score can be considered as causal indicator model (see for instance Kottner J, Streiner DL. Int J Nurs Stud. 2010 Jul;47(7):926-8). Calculating alpha and item correlations is not appropriate here because PU risk scores are not unidimensional (= homogeneous).

The item-to-total correlations are removed; however we have chosen to present the Cronbach’s alpha because the lack of calculation of the reliability for the total study group. We have discussed p. 10 that

the obtained value must be interpreted with caution.

(2) Reliability, page 6: Do not use Pearson's r for calculating associations between rater scores. For measuring interrater reliability you must use ICCCs.

Text is changed p. 6 and ICCs are calculated and presented in Table 1.

(3) Validity, page 6: Why do you used the BMI 23 as splitting variable for the known groups comparisons? Is there any evidence for that?

This has been explained in the text p. 7.

(4) Validity, page 6: Do not use a principal component analysis when you want to confirm a scale structure. PCA is for selecting and deleting items to create scores. Here you must use confirmatory analysis. Above all, you cannot identify different factors and calculate an overall alpha.
(5) Sensitivity, page 7: Please see above. You cannot validate a predictive risk score with the concurrent presence or absence of PUs.

We have chosen to use the factor analysis and we have discussed p. 11 that it will be valuable to have results from a factor analysis to be able to compare the results from the Swedish study conducted by Lindgren et al. [10, 11]. Further testing should include confirmatory analysis to confirm the obtained results in the present study and in the study by Lindgren et al. [11].

Please see my remarks above. The presented results and conclusions do not provide good evidence about the scale performance.

Because the manuscript is revised, some calculations are removed and new analyses have been performed the conclusion is consider to be appropriate for the manuscript.

Table 4, page 9: Irrespective of the questionable method you present 4 factors here. Please label the factors in a consistent manner: "Nutritional status" and "physical condition" are conceptually functional characteristics and thus comparable. The factor "major risk factors" is conceptually completely different. By the way, that physical activity, mobility and moisture load on one factor clearly indicates that this scale is not homogenous.

This factor has been renamed (Table 3) p. 9

STROBE was used, but maybe STARD or GRRAS provide better frameworks for reporting.

We have changes and used STARD.

Reviewer 2.
The title refers to "Evaluation of the Risk Assessment .... used among residents Nevertheless the article is not about using the scale, but about the translation and validation process. I think that the actual title could confound the readers. Perhaps the authors should rewrite the title in a more descriptive way, according the aim and the results.

The title is changed.

Aim: The purpose stated was "to translate and test the psychometric properties", this include to establish reliability and validity, but not the clinical validation as a diagnostic tool (sensibility, specificity,..). Perhaps the authors should explain better the aim, matching with methods and results.

Since calculation of sensitivity and specificity is removed, the aim is assessed as appropriate for the revised manuscript.

Methods: The process of translation-back-translation should be explained better. Which kind of linguistic changes were done?

Text has been added p. 4.

My suggestion for the authors is to remove data about sensibility, specificity and so on, in the manuscript; focusing just in the psychometrics properties of the scale.

The suggestion is followed.

Please explain the reason for excluding people with lower extremity amputation or receiving enteral nutrition in the validation study, because these are high-risk patients so your sample could be biased.

The patients with lower extremity amputation or receiving enteral nutrition were not able to participate in the study based on the requirement of measure height and weight for the BMI calculation p. 5.

Who did the assessment with the scale and the skin inspection? The same person?

New text is added p.5.

Homogeneity or internal consistency. It seems that authors propose a 3-factor structure for the RAPS, so you should provide not only the global Cronbach's alpha, but the figures for each factor (the authors do in table 4, but should refers them in the previous paragraph about homogeneity).

We have decided to not refer them in the previous paragraph about homogeneity.

Construct validity. Your data about validity in “known groups” (table 3) need interpretation. There are differences in the mean score between groups, but if you take into account the cut-off point there is no clinical differences between groups (in BMI and CC groups), because both of them score more than cut-off point (31 p.), so both groups are “no risk groups”. Authors should explain better these facts and consider as limitation.

Based on the fact that the calculations of sensitivity and specificity values are removed, cut-off is no longer presented.

Limitations. The authors should consider as a limitation that their research do not estimate the concurrent validity using a well validated scale, like Braden scale, in order to check if this new scale classify the patients in a similar way. Is there a Norwegian version of Braden scale?

Limitations are discussed and text is added p. 12.

VERSION 2 – REVIEW

REVIEWER	Jan Kottner Scientific Director Clinical Research, Department of Dermatology, Charité-Universitätsmedizin Berlin, Germany I have no conflict of interest regarding this work.

REVIEW RETURNED	12-Aug-2012

THE STUDY	(1) Data analysis, reliability, page 6: You cannot use alpha for scores in which items reflect different aspects on a construct because item scores do not have to be correlated. PU risk scales sum up different non related factors for building a predictive score. This is the basic logic behind them. Please see for instance the COsensus-based Standards for selection of health Measurement Instruments (COSMIN) and many more publications dealing with health measurement properties. In the response you say that alpha "must be interpreted with caution". Alpha cannot be interpreted because it is meaningless here. (2) Data analysis, reliability, page 6: Please add which ICC case you calculated and why. (3) Data analysis, validity, page 7: Please do not use principal component analysis for establishing factors of a fixed scale. You must use confirmatory factor analysis.
RESULTS & CONCLUSIONS	(4) Results, reliability, page 7: ICCCs are reliability coefficients. They not indicate agreement. Please revise in the text and the heading of the table 1. (5) Results, validity, page 9 table 3: Again, when presenting factors, then the headings must represent comparable conceptual entities. Nutritional status and physical condition are groups of risk factors. Consequently the factor 1 cannot be labeled "risk factors". (6) Study limitations, page 12: Given the wide range of available PU risk scales today it appears that there are no scales that might be used as "gold standard". Every tool is related to setting, users, risk profiles and many more factors.
REPORTING & ETHICS	(7) You used STARD for structuring your manuscript. However, STARD is useful when comparing index and reference test results. This is not done here and therefore most items do not apply. Please consider another framework for reporting.

REVIEWER	Pedro L. Pancorbo-Hidalgo Professor and Head of the department of Nursing Faculty of Health Sciences University of Jaen. Spain No conflict of interests.
REVIEW RETURNED	15-Aug-2012

THE STUDY	The sample used in this study is not fully representative of the residents population, due to some exclusion criteria. So, this fact should be considered as a limitation.
RESULTS & CONCLUSIONS	In Table 3. Factor analysis, the name proposed for factor 1 "Risk factors" isn't a good option, because, all the factors are risk factors and this name doesn't provide additional information. Perhaps your could name it as Mobility, which happens to have the highest eingevalue, or a name like this. The 3-factor solution proposed by the Varimax rotation, have 2 factors difficult to explain (factor 1 and factor 3). Authors should include their explanation about this facts and implications.

GENERAL COMMENTS	The authors have adressed most of the reviewers recommendation and, now, the article is better focused.

VERSION 2 – AUTHOR RESPONSE

Response to reviewers' comments
Thank you for the reviewers' comments, which we found very helpful. The following describes how each comment has been addressed and its location in the manuscript. Text that has been added to the manuscript is highlighted in red in the text.

Reviewer Comments	Answers to reviewers' questions and changes in the manuscript
Reviewer 1. However, my major concern is that a cross sectional design was used to validate a prognostic score. Estimating the predictive validity (=probability of future events) is only possible in longitudinal (cohort) studies. A further design weakness is the convenience sample.	Text is added page 12.
(1) Reliability, page 5: Because PU risk factors are not necessarily related with each other, the sum score can be considered as causal indicator model (see for instance Kottner J, Streiner DL. Int J Nurs Stud. 2010 Jul;47(7):926-8). Calculating alpha and item correlations is not appropriate here because PU risk scores are not unidimensional (= homogeneous).	The item-to-total correlations are removed; however we have chosen to present the Cronbach's alpha because the lack of calculation of the reliability for the total study group. We have discussed p. 10 that the obtained value must be interpreted with caution.
(2) Reliability, page 6: Do not use Pearson's r for calculating associations between rater scores. For measuring interrater reliability you must use ICCCs.	Text is changed p. 6 and ICCs are calculated and presented in Table 1.
(3) Validity, page 6: Why do you used the BMI 23 as splitting variable for the known groups comparisons? Is there any evidence for that?	This has been explained in the text p. 7.
(4) Validity, page 6: Do not use a principal component analysis when you want to confirm a scale structure. PCA is for selecting and deleting items to create scores. Here you must use confirmatory analysis. Above all, you cannot identify different factors and calculate an overall alpha.	We have chosen to use the factor analysis and we have discussed p. 11 that it will be valuable to have results from a factor analysis to be able to compare the results from the Swedish study conducted by Lindgren et al. [10, 11]. Further testing should include confirmatory analysis to confirm the obtained results in the present study and in the study by Lindgren et al. [11].
(5) Sensitivity, page 7: Please see above. You cannot validate a predictive risk score with the	Because the manuscript is revised, some calculations are removed and new analyses have been performed the conclusion is consider to be appropriate for the manuscript. This factor has been renamed (Table 3) p. 9.

concurrent presence or absence of PUs.	
Please see my remarks above. The presented results and conclusions do not provide good evidence about the scale performance.	
	We have changes and used STARD.
Table 4, page 9: Irrespective of the questionable method you present 4 factors here. Please label the factors in a consistent manner: “Nutritional status” and “physical condition” are conceptually functional characteristics and thus comparable. The factor “major risk factors” is conceptually completely different. By the way, that physical activity, mobility and moisture load on one factor clearly indicates that this scale is not homogenous.	The title is changed. Since calculation of sensitivity and specificity is removed, the aim is assessed as appropriate for the revised manuscript.
STROBE was used, but maybe STARD or GRRAS provide better frameworks for reporting.	
	Text has been added p. 4.
Reviewer 2.	
The title refers to "Evaluation of the Risk Assessment .... used among residents Nevertheless the article is not about using the scale, but about the translation and validation process. I think that the actual title could confound the readers. Perhaps the authors should rewrite the title in a more descriptive way, according the aim and the results.	The suggestion is followed. The patients with lower extremity amputation or receiving enteral nutrition were not able to participate in the study based on the requirement of measure height and weight for the BMI calculation p. 5.
Aim: The purpose stated was "to translate and test the psychometric properties", this include to establish reliability and validity, but not the clinical validation as a diagnostic tool (sensibility, specificity,..). Perhaps the authors should explain better the aim, matching with methods and results.	New text is added p.5. We have decided to not refer them in the previous paragraph about homogeneity.
Methods: The process of translation-back-	Based on the fact that the calculations of

translation should be explained better. Which kind of linguistic changes were done?	sensitivity and specificity values are removed, cut-off is no longer presented.
My suggestion for the authors is to remove data about sensibility, specificity and so on, in the manuscript; focusing just in the psychometrics properties of the scale.	Limitations are discussed and text is added p. 12.
Please explain the reason for excluding people with lower extremity amputation or receiving enteral nutrition in the validation study, because these are high-risk patients so your sample could be biased.	
Who did the assessment with the scale and the skin inspection? The same person?	
Homogeneity or internal consistency. It seems that authors propose a 3-factor structure for the RAPS, so you should provide not only the global Cronbach's alpha, but the figures for each factor (the authors do in table 4, but should refers them in the previous paragraph about homogeneity).	
Construct validity. Your data about validity in “known groups” (table 3) need interpretation. There are differences in the mean score between groups, but if you take into account the cut-off point there is no clinical differences between groups (in BMI and CC groups), because both of them score more than cut-off point (31 p.), so both groups are “no risk groups”. Authors should explain better these facts and consider as limitation.	
Limitations. The authors should consider as a limitation that their research do not estimate the concurrent validity using a well validated scale, like Braden scale, in order to check if this new scale classify the patients in a similar way. Is there a Norwegian version of Braden scale?	