Author's response to reviews

Title: A cohort study investigating a simple, early assessment to predict upper extremity function after stroke - a part of the SALGOT study

Authors:

Hanna C Persson (hanna.c.persson@vgregion.se)
Margit Alt Murphy (margit.alt-murphy@neuro.gu.se)
Anna Danielsson (anna.danielsson@neuro.gu.se)
Åsa Lundgren-Nilsson (asa.lundgren-nilsson@neuro.gu.se)
Katharina S Sunnerhagen (ks.sunnerhagen@neuro.gu.se)

Version: 4 Date: 8 May 2015

Author's response to reviews: see over
Dr Toby Cumming  
Associate Editor  
BMC Neurology  

Dear Editor,  
Thank you for the additional comments to our manuscript “A cohort study investigating a simple, early assessment to predict upper extremity function after stroke - a part of the SALGOT study”, MS 1379509548160386.

The following comments were addressed: “Thank you for your detailed responses to the reviewers. The only issue that I think still needs to be addressed is the indirectness of using the ARAT to predict FMA, which is used to predict a drinking task (Reviewer 1 brings this issue up in points 5 and 6). The added text in the Discussion is vague re: ARAT (predictive validity for what?), and providing more detail on the drinking task in the Methods doesn't really answer the question. The fact that the ARAT wasn't used to directly predict the drinking task should be included as a limitation in the Discussion.”

We have tried to follow the suggestions given and we included our response as green bold text in the manuscript as well in the point to point revision note below.

- Line 136 (method): Validation of the cut-off score of FMA-UE $\geq 32$ to correctly classify patients’ motor ability to drink from a glass (the drinking task) was based on the entire cohort.
- Line 328 (discussion): Also the fact that the ARAT was not used to directly predict the drinking task is a limitation.
- Line 343 (discussion): Similarly, the predictive validity of ARAT-2 to directly predict a drinking task needs to be investigated in another independent cohort.

Looking forward to hear from you soon,

Yours sincerely  
Hanna C Persson  

Corresponding author:  
Hanna C Persson, PT PhD-student,  
Department of Clinical Neuroscience and Rehabilitation, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden,  
Telephone number: +4631342267, Fax: +46-31-3422497, E-mail: hanna.persson@neuro.gu.se
Reviewer 1

Dear Dr Stinear,

Thank you for your comments which have helped us to improve the manuscript. Below are point-to-point answers to the specific questions raised.

Minor essential revisions

1. FMA-UE is usually considered a measure of impairment, while ARAT is usually considered a measure of function. Please clarify throughout the manuscript.

   Thank you for pointing this out. We have used the terms included in the ICF classification of function, activity and participation. Functioning includes both body functions as well as activities.
   We have tried to clarify this throughout the manuscript.

Major compulsory revisions

1. This is not a prospective study, in that the process for selecting the ARAT items for the ARAT-2 was not complete until all the data had been collected and analysed. To be prospective, the study would need to test the predictive power of ARAT-2 from the outset. Please revise the title and manuscript accordingly.

   Thank you for making us aware of this. We have now changed the text throughout according to a cohort study.
   The title of the study has been changed: A cohort study investigating a simple, early assessment to predicts upper extremity function after stroke - a part of the SALGOT study
   In the methods, participants line 111: …a prospective cohort study has been omitted. In same section, in line 123 also includes: …a cohort of…

2. It’s unclear how predicting a binary outcome on the FM scale (which in turn relates to the ability to drink from a cup) solves the problem identified in the Introduction, line 96:
“However, in these [previous] studies … the outcome does not tell us if this predicted motor function can be useful for daily activities”. Drinking from a cup is only one of many important daily activities, and it doesn’t require much in the way of individuated finger movements, which are essential for other important daily activities.

Yes, we agree that this study only have investigated one of many activities possible to perform with the upper limb. We still think finger movements need to be involved to perform a drinking task, which we also have addressed in discussion, line 259. Grasping the glass is not possible without finger movements. We are aware of that other tasks can involve finger movement to a higher extent, but the drinking task is one, among other activities. We choose to use this task, in this study, to identify motor function required for a drinking task.

According to your suggestion, we have now defined the drinking task in the methods section, line 135; This drinking task requires the capacity to reach, grasp, lift, transport the glass as well as drink.

In the introduction line 97, we changed and to or.

The possibility of the cut-off of FMA-UE to correct identify patients’ motor ability to drink from a glass is given in the methods section, and has high sensitivity and specificity (line 135-141). Also see answer to question number 4 below.

3. The FMA-UE scores were binarised as below 32 points and 32 points or more, and this cut-off point was established in the authors’ previously published study. It seems that some of the 30 patients in the previous study might also have been part of the present study though, as both studies report on participants in the SALGOT study. Ideally, there should be no overlap, if the results from the first study of 30 patients are being used to rationalise the choice of cut-off in the second study.

It is correct that 30 of the patients were included both articles, however the cut-off was validated in the total cohort in the present study (line 135-141, also see question number 4). We have also confirmed this cut-off in other non-publish research material in persons after stroke.

In the present study at the end of the discussion, a suggestion of further research including validation of the FMA-UE cut-of is now added: Line 340: In future studies, the discriminative validity using a cut-off at FMA-UE score of ≥32 to identify persons that are able to perform a drinking task needs to be assessed in another stroke cohort.
4. The previous study only included patients who could perform the drinking task with their affected upper limb. These patients had FMA-UE scores ranging from 32 to 64, and ARAT scores ranging from 24 to 54. Patients who couldn’t perform the task were excluded, so it’s impossible to know whether their scores exceeded 32 on the FMA-UE or 24 on the ARAT. The choice of cut-off score would have more support if it was shown to clearly distinguish between those who can perform the task, and those who can’t. This is partly addressed in Line 135 onwards, where the accuracy of the FMA-UE cut-off in correctly classifying patients’ ability to “perform a drinking task” was evaluated at 3 time points during the study. However, these data and analyses don’t seem to be reported, and there are no details provided on the drinking task. What constitutes successful performance?

The ability of FMA-UE to correctly identify if the patients were able to perform the drinking task with their affected arm is provided in the methods section, line 135-141. This drinking task requires the capacity to reach, grasp, lift, transport the glass as well as drink. Validation of the cut-off score of FMA-UE ≥32 to correctly classify patients’ motor ability to drink from a glass (the drinking task) was based on the entire cohort. The results were: at 10 days sensitivity 98% (CI 95% 0.91-1.0) and specificity 89% (0.77-0.96), at 1 month sensitivity 100% (0.92-1.0) and specificity 93% (0.84-0.98) and at 12 months sensitivity 100% (0.85-1.0) and specificity 96% (0.87-1.0). In the manuscript we have also included additional reasons to the classification errors in the validation process, line 142. The majority of classification errors occurred in data gathered at 10 days post stroke, on the group of patients with moderate/mild UE impairment but with a poor hand function and inability to grip and perform the drinking task (n=6).

According to the protocol of the SALGOT-study, we analyzed the movement “drinking from a glass” using a kinematic camera system. If a patient could not participate in the kinematic movement analysis, this was systematically reported by the test leader (physiotherapist). Reasons for not able to perform the drinking task could be insufficient motor function, but also due to cognitive impairments, inability to come to the test room, fatigue or other. All patients were also (before assessment started, at day 10, month 1 and month 12) asked about their ability to perform “drinking from a glass with the paretic arm”, and responded with yes/no. This self-responded question was used in the second step to find out if a patient could perform the drinking task even if not kinematic movement analysis was used.
We have now in the manuscript also defined the drinking task in the methods section, line 135. *This drinking task requires the capacity to reach, grasp, lift, transport the glass as well as drink.*

5. More generally, it seems that this study has used a drinking task to represent meaningful upper limb function, used a FMA-UE score to predict who can perform the drinking task, then used two items from the ARAT scale to predict who will exceed the FMA-UE score cut-off. The relationship between ARAT-2 and the drinking task is therefore somewhat indirect, via the FMA-UE score. A more robust test of the ARAT-2 would be to see whether it directly predicts which patients can perform a (defined) drinking task, or preferably more than one functionally meaningful tasks.

We have now given a more detailed definition of the drinking task in the methods section, line 135.

In the present study in the end of the discussion, a suggestion of further research now is added, line 338. *In future studies, the discriminative validity using a cut-off at FMA-UE score of $\geq 32$ to identify persons that are able to perform a drinking task needs to be assessed in another stroke cohort. Similarly, the predictive validity of ARAT-2 to directly predict a drinking task needs to be investigated in another independent cohort.*

We have also in the discussion at line 328 included a sentence of limitation: *Also the fact that the ARAT was not used to directly predict the drinking task is a limitation.*

6. The results and discussion describe the ARAT-2 as able to correctly predict whether patients can perform a drinking task. It’s important to note that the ARAT-2 was used to predict whether patients would exceed the FMA-UE score cut-off. This isn’t quite the same thing as being able to perform a drinking task (which is not described).

We agree to that reaching the cut-off and be able to perform a drinking task is not the same thing, and we have tried to change accordantly throughout the manuscript.

In addition, the drinking task now also is defined in the methods section, line 135. We have also in the discussion at line 328 included a sentence of limitation: *Also the fact that the ARAT was not used to directly predict the drinking task is a limitation.*

7. It’s perhaps not surprising that the predictions made at 3 days were most accurate for FMA-UE score at 10 days. At 3 days it is probably more clinically relevant to be able to make accurate predictions for time points later in recovery.
The reasons to the chosen times for testing and follow-up have been described in the discussion, line 310, and have now been changed to: In this study we tried to cover both, the need for early evaluation (day 3 and day 10) since the length of stay in hospital is getting shorter, as well as the need for a later assessment, at 1 month, which can be useful for more impaired patients who will require a longer period of rehabilitation (Winters C et al 2014). We also have in clinical practice experienced a need of the short prediction (7 days) in the acute care to give correct information to the patient and in short time goal setting/rehabilitation planning.

In order to make this clearer, we have now also changed a sentence in the methods section Line 163-166: These time-points were considered as being of possible clinical importance for both early and long term rehabilitation planning.

8. Was a standardized approach used for the clinical scales, such as that outlined by Yozbatiran et al. for the ARAT (NNR, 2008)? How many raters administered the clinical scales and were they trained for consistency?

Thank you for making us aware of the lack of the information of the administration procedure of assessments. The standardized approach outlined by Yozbatrian has been followed. The reference to Yozbatrian and also the Swedish reference Nordin et al have now been included in the method section line 162: The ARAT was performed in a standardized manner (Yozbatrian et al and Nordin et al) at 3 days and 1 month post-stroke.

We have also included the following information in the methods section, line 146-149: Three physiotherapists, after joint training, performed the clinical assessments according to a standardized procedure (REF). The majority of the assessments were performed at the hospital, and if the patient was unable to travel, assessment was conducted in the patient’s home, nursing home or rehabilitations unit. Global neurological deficits were detected by physicians…

9. How does predicting whether a patient will achieve one of two outcomes on the FM scale help therapy teams and patients? How might this information be used by therapy teams?

In this cohort, a high concordance between the dichotomized FMA-UE and the patient’s possibility to perform the drinking task was shown. In the discussion section, we have pointed out the clinical importance of improved knowledge of the patient probable UE function at later stages. We have now also added in line 304: This information can be
used for planning of the content of rehabilitation; i.e. training vs compensatory strategies as well need for care and assistance.

10. What was the therapy dose completed by the patients in this study? How might variations in therapy dose affect their outcomes, and the accuracy of the predictions? Or does ARAT-2 make predictions that are accurate for any therapy dose? What might this mean for the role of therapy?

This study includes patients from one stroke unit in Sweden with stroke care that is organized with governmentally funded rehabilitation, and with a rehabilitation that starts at the first days after onset. It is likely to believe this has an impact on the results. Table 1 briefly includes information of amount of received rehabilitation. In line 155-158 this is have now been changed to: Physiotherapist and occupational therapist were available in the primary care system as well as in the community and in nursing homes. The level of rehabilitation received at the different test occasions during the first year is described in table 1, and follows the Swedish national guidelines (reference).

If the results could be generalized in a larger population is a question in coming research, now addressed in the end of the discussion, line 340-343.

11. The discussion notes (line 235) that accurate predictions are more difficult for patients with higher levels of initial impairment. This is true of prediction tools that rely solely on clinical assessment. In contrast, accuracy for these patients can be higher when using measures of corticomotor pathway integrity. Simple clinical assessment is unable to detect intact, but not yet functional, corticomotor connections that can support subsequent recovery. This point could be usefully added to the possible reasons outlined in line 236 onwards.

Thank you for pointing this out. We have now inserted this sentence, line 256-258: Another explanation could be that the clinical assessment scores at stroke onset does not give enough information to distinguish between patients with similar initial impairment but with different recovery potential and thereby predicted outcomes (ref Stinear, The PREP algorithm..2012).

12. It would be helpful for the reader if the description of the third approach (line 260) was more informative, and identified as the algorithm from reference 13, in place of
“objective evaluation”. This will help the reader make the link back to the previous points in the discussion.

Thank you for notice this. We have now added in line 280 …clinical assessment and TMS or MRI evaluation, using an algorithm (ref Stinear, The PREP algorithm..2012) yielded…

We agree that this will help the reader to refer back to previous discussions within the manuscript.

13. Line 288 comments on the need for early evaluation – at days 3 and 10. The following sentence mentions assessments at 1 month. The next paragraph continues in this vein. However it’s not clear in the method that assessments at 1 month were used to make predictions at 12 months, as the focus is on predictions made from assessments at 3 days. In the aim it is defined ….that a sub-set of items form ARAT administrated at 3 days and 1 month post stroke…..

In the method section we now tried to make this clearer, line 176-179: The selected ARAT items from 3 days post stroke were used to predict the UE function (to detect the motor function required for a drinking task using the paretic arm) at 10 days, 1 month and 12 months post stroke; and the ARAT items from 1 month post stroke were used to predict the UE function at 12 months post stroke.

In discussion, line 310 we also made changes: In this study we tried to cover both, the need for early evaluation (day 3 and day 10) since the length of stay in hospital is getting shorter, as well as the need for a later assessment, at 1 month, which can be useful for more impaired patients who will require a longer period of rehabilitation (reference). Hopefully this will make it clearer for the reader.

14. It would be good to identify the need for validation of the ARAT-2 in a new cohort of patients, in a prospective study.

Yes, we agree to this, and two sentences of further research have now been added at the end of the discussion, line 340-343.
Reviewer 2

Dear Dr Kaffengerber.

Thank you for your comments which have helped us to improve the manuscript. Below are point-to-point answers to the specific questions raised.

Major Compulsory Revisions:

1. Reporting of results

The authors are using two items of the Action Research Arm Test (ARAT) to predict functional relevant arm function measured with the Fugl-Meyer Assessment (FMA). They conclude that testing the ARAT items “pour water from glass to glass” and “place hand on top of head”, predicts the ability of performing a drinking task. This conclusion seems circular as predicting the ability to perform a certain movement by testing a similar movement reveals naturally a good sensitivity and poor to moderate specificity. The authors may consider re-thinking the reporting of the results: Whereas the FMA is a complex time demanding procedure, 2 simple items of the ARAT may be performed within a few minutes even in severely impaired patients. Showing that ARAT-2 is able to stratify motor function similar to FMA might be of major importance. In this context the authors could consider analysing the reliability of ARAT-2 and FMA-UE when assessed at the same time point (3 days and 1 month after stroke).

Thank you for these important considerations. That ARAT-2 ability to predict motor function required for a drinking task could be a circular conclusion has been discussed in Line 260 in the discussion: The items in ARAT, “Pour water from glass to glass”, consists of a similar movement as when drinking, but to receive one point the patient only needs to initiate the task. Initial function in shoulder abduction and finger extension has been described as important predictors for outcomes in UE function at 6 months (references).

To re-organize the results to investigate the ARAT-2 possibility to report motor function in similarity to FMA-UE would be important and interesting in another study, also to analyze the reliability of ARAT-2 and FMA-UE. Unfortunately this is not covered by the purpose of this study. This is however partly discussed in line 283: The new ARAT-2 does not cover all aspects of functioning, but it could be considered useful as an early predictor of UE motor function that is required for a daily activity, such as drinking from a glass after a stroke.
In the methods section, in data handling and statistics, line 201 the following sentence has been added: *To provide strong evidence in most circumstances (in a stroke population) positive likelihood ratio values should be above 10 and negative likelihood ratios should be below 0.1 (reference).*

2. Study design

a. The study design isn’t precisely described. In the methods section, “Participants”, 1st paragraph the SALGOT study is mentioned for the first time, which comes, somehow, surprising for the reader. The authors should clarify if the reported study is a sub-study of the SALGOT study or if it is a retrospective sub-analysis of the SALGOT study data. In this context it might stress the importance of the presented study to focus on relevant data rather than to summarise different data in table 1 without further explanation (e.g. number of thrombolysis). If the reported study is a separate sub-study, it should have received ethical approval, which – according to the second to last sentence of this section – only the SALGOT study has.

Thank you for making us aware of the lack of early information of the SALGOT study. We have now highlight that the present study is a part of the SALGOT-study and changed the title to: *A cohort study investigating a simple, early assessment to predict upper extremity function after stroke - a part of the SALGOT study.*

In the methods, participants line 111: …*a prospective cohort study*…. has been omitted. In same section, in line 123 also includes: …*a cohort of*…

In table 1 we included information we expected to be of importance to the reader to improve the understanding of the study population. We chose to provide information of thrombolysis, due to this is one important factor that can affect the recovery process. To increase the readability in table 1, we deleted information on number of women (consisted of information that matched the men), the mean and SD in NIHSS, FMA-UE and ARAT, information of pain and sensory function. We have also changed from given both the absolute numbers and the percentage, to only give the percentage of the group. Finally, the information of received rehabilitation has been changed and hopefully improved.

b. The authors do not explain why they have chosen the described testing and follow-up time points. If it is a prospective study the motivation for the chosen time-points should be given. If it is a retrospective study this should be stated.
Thank you for this suggestion. The reasons to the chosen time points for testing and follow-up have been described in the discussion, on line 310-313: *In this study we tried to cover both, the need for early evaluation (day 3 and day 10) since the length of stay in hospital is getting shorter, as well as the need for a later assessment, at 1 month, which can be useful for more impaired patients who will require a longer period of rehabilitation (Winters C et al 2014).*

In order to make this clearer, we have now also changed two sentences in the methods section Line 162-164: *The ARAT was performed in a standardized manner [6, 20] at 3 days and 1 month post-stroke. These time-points were considered of possible clinical importance for both early and long term rehabilitation planning.*

### 3. Language and general writing

Unfortunately, partly due to poor language partly due to incoherence, the content of the article is not very comprehensive. Especially the background and the methods sections aren’t well structured, which makes it difficult to extract the main purpose and procedure of the study. I strongly advise a re-writing of the article with support of an expert or native English speaker.

E.g., methods section, “Clinical assessments and procedure”, 1st paragraph, next to last sentence: “The sensitivity ranged between…” might be more conclusive in the result section. Thank you for suggestions to improve the readability. The paper was checked by a native Australian (professional editor) before submission. We have tried to improve the language in the background and the discussion in general.

The manuscript has been checked again by a native English language reviewer before this re-submission.

The sentence, mention above is about the sensitivity and specificity of the FMA-UE ≥32 to correctly classify the patients’ motor ability to perform a drinking task. Due to this is not a part of the aim, we chose to have this information in the method section. We have clarified the sentence above, line 135-145: *This drinking task requires the capacity to reach, grasp, lift, transport the glass as well as drink. Validation of the cut-off score of FMA-UE ≥32 to correctly classify patients’ motor ability to drink from a glass (the drinking task) was based on the entire cohort. The results were: at 10 days sensitivity 98% (CI 95% 0.91-1.0) and specificity 89% (0.77-0.96), at 1 month sensitivity 100% (0.92-1.0) and specificity 93% (0.84-0.98) and at 12 months sensitivity 100% (0.85-1.0) and specificity 96% (0.87-1.0). These results confirmed the use of this cut-off in the subsequent analysis. The majority of classification errors occurred in data gathered at 10 days post stroke, on the group of*
patients with moderate/mild UE impairment but with a poor hand function and inability to grip and perform the drinking task (n=6).

E.g., methods section, “Data handling and statistics”, 2nd paragraph, 3rd sentence: “The minimal number of items…” is an almost exact repetition to the 3rd sentence in the paragraph “ARAT and reduction of items.”

The sentence in ARAT reduction of items, Line 167 this have now been changed to: 2) the minimum number of items needed to capture most of the variance in the ARAT at day 3 was explored:

The line 190 has been changed to: The fewest number of items needed….. population (n=117).

E.g., results, 2nd paragraph, third last sentence: “The predictive value …” should be part of the methods section and has already been mentioned there (“Data handling and statistics”, 2nd paragraph).

Thank you for pointing this out. This sentence in the result section, line 223-228, is now changed to: Figure 3 shows the ROC curves of the ARAT-2 to classify the motor function required for a drinking task (FMA-UE ≥ 32 points), and a cut-off level of 2 points…… and month 12

E.g., discussion: The first 2 sentences are stating the same fact.

Thank you for this suggestion, we have omitted the second sentences in the beginning of the discussion.

E.g, discussion, 3rd paragraph, 2nd sentence:
“In order to identify…was used” is part of the methods section and has already been mentioned there.

Thank you. This sentence in the discussion section, line 296, is now changed to: The items’ differences in difficulty were identified with a Mokken analysis (reference) in which similar differences in difficulty also were shown using Rasch analysis. This provided

Minor Essential Revisions:
1. abstract, result section, 2nd sentence Without further explanation it’s hard to understand, what the cut-off does separate (favourable versus unfavourable motor function?).

   Line 47: changed to: ARAT-2 is a sum score (0-6) with a cut-off at 2 points, where >2 is considered an improvement.
To have enough space in the abstract to insert this information, we choose to omit the 95% confidence intervals.

2. abstract, result section, 3rd sentence “The sensitivity varied….” Please, make clear what the variation of sensitivity reflects (different time points?).
   This has now been changed to: At the different time points, the sensitivity varied between…(line 48)

3. background, last paragraph This sentence gives the impression that the ARAT was administered on the stroke unit 3 days and 1 month post stroke. I assume that patients haven’t stayed on the stroke unit for 1 month. It might be worth mentioning in which setting the further testing has taken place (in the method section).
   This has now been changed to: …of items from ARAT, administered at 3 days and 1 month… in coherence to the aim in the abstract.

4. methods, “Clinical assessments and procedures”, last paragraph Similar to sensory function and pain assessment the Bamford classification isn’t used or referred to any further in the article. The authors might skip this information. The last sentence refers to table 1, in which “the details of the type of rehabilitation” should be described; but table 1 only refers to the number of patients obtaining or not obtaining therapy.
   The information in table 1 was aimed to provide important information of the study population. The Bamford classification is used to sub group and describe the ischemic stroke and was therefore also used in the present study. We have after you pointed this out changed the table as stated above.
   The sentence of the rehabilitation in method section, line 155-159: Physiotherapist and occupational therapist were available in the primary care system as well as in the community and in nursing homes. The level of rehabilitation received at the different test occasions during the first year is described in table 1, and follows the Swedish national guidelines (reference).

5. methods, “ARAT and reduction of items”, second last sentence “From the selected… was determined.” Again, it should be made clear, what the cut-off does separate.
   This is in line 174 changed to: …..was determined where higher score indicate better function.

6. results, 2nd sentence “Between participating patients…” should rather be “Between completing patients…” as also the drop-outs participated initially.
This is in line 210 changed to: *There were no significant differences between the sex of the participants, initial stroke severity or initial UE function when comparing patients who completed the study (n=73) and those who dropped out at 12 months (n=39)*; discussion, Overall the discussion and the conclusion are well balanced. Unfortunately, the authors have compiled pure description of other studies without associating it with their own findings.

We have tried to improve the discussion in general, aspects of further research has been added, line 340-343.

**Discretionary Revisions**

8. Figures 1 and 2

Figure 1 and figure 2 do not make the described procedure any clearer (a simple but precise narrative description in the text might have greater impact). The authors may consider using a table instead of figure 1 to mark the time points at which what score was assessed.

Thank you for the suggestion of changing figure 1. We think that figure 1 not only contributes with exact time point of different assessments, but also with knowledge of the drop-outs at 1 and 12 months. We believe this is difficult to illustrate in a table, and would like to keep figure 1.

Figure 2 was aimed to be a complement and visual illustration to the procedure to choose the ARAT items. The text in the method has been slightly changed, and we believe that figure 2 contributes to the readability in the manuscript, and our suggestion is therefore to keep the figure. A new up-dated figure 2 has been added.