PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Can the STarT Back Tool predict health related quality of life and work ability after an acute/subacute episode with back or neck pain? – a psychometric validation study in primary care
AUTHORS	Forsbrand, Malin; Grahn, Birgitta; Hill, Jonathan; Petersson, Ingemar; Post Sennehed, Charlotte; Stigmar, Kjerstin

VERSION 1 – REVIEW

REVIEWER	Victor Hoe Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, MALAYSIA
REVIEW RETURNED	09-Feb-2018

GENERAL COMMENTS	Thank you for the opportunity to review your paper on “Can the STarT Back Tool predict health related quality of life and work ability after an acute/subacute episode with back or neck pain? – a prospective cohort study in primary care”. Overall it is a well-written report which is easy to read and understand. There is only minor issue that needs to be addressed, 1. Has the SBT been validated for the Swedish population? 2. p-value is usually not presented as p=0.0000 but as p<0.0001 3. Table 1 – to include the type of intervention, the SBT total score is not necessary. 4. Present the regression analysis in a Table format. It is unclear how the variable age, sex, treatment or time to follow-up was assessed in the regression analysis. The data should be included. 5. There should be an explanation of the use and the meaning of the results of the test Nagelkerke R2 and Cox-Snell R2. 6. The discussion needs to clarify why to use SBT as a predictive tool, where the original design of the SBT was for targeted treatment pathways. If the therapist uses it for targeted treatment pathways, will the predictive value of SBT still be as robust?

REVIEWER	Martin Underwood Warwick CTU, UK
REVIEW RETURNED	27-Feb-2018

GENERAL COMMENTS	Thank you for asking me to review this paper that is adding to our knowledge. I do, however, have some substantial concerns about some aspects of the work presented. 1. Participants for this study are people invited to join a cluster randomised trial of a workplace intervention. Included a mixture of people who joined the trial and those who were assessed and not included in the trial. This mean a mixed group of participants with three distinct sets of characteristics that may affect outcome. It is

	thus not, strict reporting a cohort study 2. It should also be noted that these are not an inception cohort but a group who have self-referred for treatment of their back pain to a physiotherapy department 3. Some more detail is needed on screening ahead of consideration for study entry. That 35 primary care centres over one year only saw 329 people with acute or sub-acute low back pain who had self-referred for physiotherapy seems unlikely – this equates to less than ten per site per year 4. The actual number of people recruited to the overall programme of work is unclear. Whilst what is described in this paper is internally consistent it is not externally consistent. The Trial registration website for this study indicates that 364 people were recruited to the trial upon which this study is based; whilst in this paper 162 people were included who had been part of the trial. I have not been able to identify a published protocol or final report of the randomised controlled trial to explore this further 5. Whilst the Startback tool has not been used to inform participant management the intervention arm of the trial specifically targeted psychosocial yellow flags for poor prognosis of low back pain. This approach covers the same domains as the StartBack Tool and might reduce any apparent effect of its use as a screening tool by addressing psychosocial risk factors. Without the main paper being available it is difficult to know how these factors might interact 6. That there are only full data on those included in the RCT makes interpretation of the final results extremely challenging. Fundamentally, we are being asked to believe that the StartBack Tool predicts outcome in terms of health related quality of life when we have not, and cannot be, presented with the baseline data for health related quality of life, and back pain severity, for the whole cohort. Thus we do not know if what is being reported for differences are those engendered by future effects of StartBack status at baseline or they are simply reflecting the relationship that was already present at baseline. It would not be surprising in pain severity, health related quality of life, and StartBack Status were associated at baseline. The authors need to work out how best to address this problem that seriously impedes the interpretation of their findings. 7. The authors may wish to reflect rather more on the discriminant ability of the startback tool for predicting poor outcome when compared to other measures when compared , for example, to clinician assessment or baseline pain severity and relevance for clinical practice. It is suggested here that StartBack is marginally superior to other measures but AUCs of 0.68 and 0.73. Many may feel these data do not indicate a sufficiently high level of discrimination for this to be worthwhile A minor point is that in the background it I suggested that evidence based guidelines suggest use of brief questionnaires for risk stratification. To be specific it is UK NICE Guidance that recommends this and the second reference is this sentence does not refer to evidence based guidance

REVIEWER	Prawit Janwantanakul Department of Physical Therapy, Faculty of Allied Health Sciences, Chulalongkorn University, Bangkok, Thailand
REVIEW RETURNED	11-Mar-2018

GENERAL COMMENTS	General comments ========== This prospective cohort study aimed to evaluate the SBT's predictive

	validity for HRQoL and work ability outcomes at long-term follow-up in a population with acute/subacute back and/or neck pain. The topic is of interest in the field and the paper is very well-written. My main concern is about the inclusion of those with neck pain with and without low back pain into the study because the SBT is developed and validated to predict future disability due to low back pain. This may affect the internal validity of the findings and the authors did not state clearly on this issue. Specific comments ========== 1. INTRODUCTION •The authors mentioned that the SBT is developed and validated to predict future disability due to low back pain of any duration. What is the justification for studying in a population with neck pain? Are there any limitations for translating findings of previous studies in LBP to a population with NP? •Comments on using SBT, which is developed for LBP, in the population of NP would be useful. How appropriate is it? 2. METHODS •How was non-specific LBP and NP confirmed? •It is not clear to me how patients were identified as having acute or subacute LBP and NP. Did you ask about previous episodes? •Did you have exclusion criteria for the study? •A statement about the completion of follow up is required. •The authors used a nonparametric approach which was chosen based on the distribution of the data. Please specify the reason for using a nonparametric statistic. 3. RESULTS •No description of the Kruskal Wallis test and Mann Whitney U test for studying if there were any differences between the SBT risk groups on follow-up data on poor or good HRQoL and work ability. •It is unclear how the medians in Table 2 are derived. The authors did not clearly state how their outcomes were collected during the follow up. •I wonder whether the authors had information about duration from the onset of symptoms to the first encounter with a clinician. Would it affect the findings of the study? 4. DISCUSSION •Discussion regarding the effect of the broad variation in time to follow up on the findings would be useful. •What about the clinical implications of the findings? How can clinicians use the findings in their day-to-day work?

REVIEWER	Achim Elfering University of Bern, Switzerland
REVIEW RETURNED	15-Mar-2018

GENERAL COMMENTS	BMJOPEN-2018-021748 This interesting study reports the prognostic validity of the STarT Back Tool in prediction of health-related quality of life and work ability in a large sample of primary care patients across 35 health care centers. The writing of the study is clear and well-developped. I have only a few points author(s) should consider ina revision:

	1) To investigate a varying number of patients from 35 health care centers means to have a multilevel-data structure. Author(s) should report whether health-related quality of life and work ability differed systematically between 35 health care centers (Intra-class correlations). If so, a multilevel-regression approach is more adequate than the current approach. 2) Some data were reported separately for patients having back pain or « patients reporting neck pain plus back pain patients with neck or shoulder pain (NP) of without back pain » (cf. Table 1). This separation is unclear to me. Moreover, it would be more clear and interesting when the prognostic validity of the STarT Back Tool would have been tested not only for the total of patients but also separately for patients reporting solely LBP, solely neck and shoulder pain, and those patients who reported LBP and neck and shoulder pain. 3) Time of follow-up after primary care did vary much. Author(s) should discuss time in more depth (see as an examples for a systematic investigation of follow-up time after primary care : Melloh, M., Elfering, A., Käser, A., Rolli Salathé, C., Barz, T., Röder, C., & Theis, J.-C. (2015). What is the best time point to identify patients at risk of developing persistent low back pain? Journal of Back and Musculoskeletal Rehabilitation, 28(2), 267-276. doi: 10.3233/BMR-140514 4) Working ability was assessed, but to me it was unclear whether the sample indeed included only individuals who were employed or had an active working status at baseline. 5) Two thirds of participants were also included in a RCT (intervention or control group). One third of participants did not. Author(s) tested whether RCT participation predicted health-related quality of life and work ability. Author(s) should also test whether RCT participation interacted with STarT Back Tool in prediction of health-related quality of life and work ability.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1(Victor Hoe):
Reviewer reports:

Thank you for the opportunity to review your paper on “Can the STarT Back Tool predict health related quality of life and work ability after an acute/subacute episode with back or neck pain? – a prospective cohort study in primary care”. Overall it is a well-written report which is easy to read and understand. There is only minor issue that needs to be addressed,

Authors ‘response: Thank you for the encouraging and positive support.

Authors’ action: We have made changes in the manuscript and our response to each comment is described more in detail below. Please note that changes in the manuscript are marked in red.
…….

1. Has the SBT been validated for the Swedish population?

Authors ‘response: Yes, it has. We are sorry for not being clear enough about this. The SBT was first cross-culturally adapted and validated for Swedish conditions in a low back pain population (Betten et al, 2015) and recently the SBT was validated against the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire for patients with both back and neck pain in a Swedish primary care setting (Forsbrand et al 2017).

Betten C, Sandell C, Hill JC, Gutke A. Cross-cultural adaptation and validation of the Swedish STarT Back Screening Tool. European Journal of Physiotherapy 2015;17(1):29-36.

Forsbrand M, Grahn B, Hill JC, Petersson IF, Sennehed CP, Stigmar K. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. BMC musculoskeletal disorders 2017;18(1):89.

Authors' action: We have clarified this in the introduction: "The SBT is cross-culturally adapted and validated in Swedish19 and recently also for a population with both back and neck pain in primary care 20". Please see Introduction, page 4, line 87-88.

…….
2. P-value is usually not presented as p=0.0000 but as p<0.0001

Authors 'response: We agree. It is more common to present the p-value as you describe.
Authors' action: We have changed all "p=0.000" to "p<0.001" in the manuscript, table 2 included.

…….

3. Table 1 – to include the type of intervention, the SBT total score is not necessary.

Authors 'response: Thank you for the advice. We agree – the SBT total score is not necessary and it is better to include information about the type of intervention.
Authors' action: We have excluded the information about SBT total score and we have instead added information about the type of intervention in table 1, please see Table 1, page 11 in the Results section.

…….

4. Present the regression analysis in a Table format. It is unclear how the variable age, sex, treatment or time to follow-up was assessed in the regression analysis. The data should be included.

Authors 'response: Thank you for your comment on this and the suggestion of presenting the regression analysis in a table format. We agree that it will be clearer for the readers to understand the regression analysis if we present it in a table format. The reason why we, from the beginning, chose to present the results of the analysis in text instead of in a table were to save space in the article, and to not to exceed the recommended maximum number of tables and figures as BMJ Open recommend a maximum of five tables and figures. We hope that BMJ Open will accept another table.

Authors' action: We have added a table (table 3) in the manuscript where you also can see how the variables age, sex, treatment and time to follow-up were entered into the regression analysis, please see page 13 in the Results section.
We have removed "double information" from the text on the results of Hosmer and Lemeshow´s test, Cox-Snell R^2 and Nagelkerke R^2 that is now found in table 3, please see page 13.
We have rephrased: "The regression analysis showed that the SBT high risk group could significantly predict poor HRQoL (OR 6.16, CI 1.50-25.26, B=1.82, p=0.012) and poor work ability (OR 5.08, CI 1.75-14.71, B=1.62, p=0.003) at long-term follow-up also after adjusting for age, sex, treatment and time to follow-up (Table 3)", please see Results, page 12, line 253-256.
…….

5. There should be an explanation of the use and the meaning of the results of the test Nagelkerke R2 and Cox-Snell R2.

Authors 'response: Thank you for your comment on this. Cox-Snell R^2 and Nagelkerke R^2 are useful statistics that are described as approximate equivalent to R^2 in the linear regression model and are used to present how much of the variation in the dependent variable that is explained of the predictor (independent) variables. Larger R^2 values indicate that more of the variation is explained by the model, to a maximum of 1. In our analyses, the model for HRQoL accounts for between 12% and 21% of the variance (Cox-Snell R^2=0.12. Nagelkerke R^2=0.21) and the model for work ability accounts for between 11% and 16% of the variance (Cox-Snell R^2=0.11. Nagelkerke R^2=0.16).

References:
Brace N, Snelgar R, Kemp R. SPSS for Psychologists: Palgrave Macmillan 2012.
Cox, D. R., and E. J. Snell. 1989. The Analysis of Binary Data, 2nd ed. London: Chapman and Hall.
Nagelkerke, N. J. D. 1991. A note on the general definition of the coefficient of determination. Biometrika, 78:3, 691-692.

Authors' action: We think that interested readers can see the results of the Nagelkerke R2 and Cox-Snell R2 in the new added table and no other action is taken. Please see table 3, page 13 in the Results section.

……

6. The discussion needs to clarify why to use SBT as a predictive tool, where the original design of the SBT was for targeted treatment pathways. If the therapist uses it for targeted treatment pathways, will the predictive value of SBT still be as robust?

Authors 'response: This is an important question and we are happy to address how we have thought about it. In the clinical trial WorkUp, where we identified patients to this psychometric study, SBT was not used to target interventions at baseline. The physiotherapists in the study were not aware of the results from the SBT. In the current study we have not included information on how the patients were treated. From the clinical trial we know that all patients received evidence based physiotherapy and in the intervention group, a work place dialogue as an add-on. For patients not included in the trial, we have no information on how they were treated, more than that they applied for physiotherapy.

Authors' action: We have added text in the discussion. Please see page 19, line 389-395:
"The SBT is primarily designed as a "stratified care tool" which involves targeting treatment to subgroups of patients based on their key characteristics 63 but in this study, we wanted to study if the SBT could predict the important outcomes HRQoL and work ability when applied in an RCT of neck and back pain. In this study, the physiotherapists did not target treatment based on SBT. However, we accept that some of the constructs within the SBT may have been addressed by the intervention provided which may have affected SBTs ability to predict the above mentioned outcomes."

……
Reviewer #2 (Martin Underwood):
Reviewer reports:
Thank you for asking me to review this paper that is adding to our knowledge. I do, however, have some substantial concerns about some aspects of the work presented.

Authors 'response: Thank you for your thoroughly review of the manuscript. We have answered the different concerns that you have highlighted below.

Authors' action: We have made changes in the manuscript and our response to each comment is described more in detail below. Please note that changes in the manuscript are marked in red.

…….

1. Participants for this study are people invited to join a cluster randomised trial of a workplace intervention. Included a mixture of people who joined the trial and those who were assessed and not included in the trial. This mean a mixed group of participants with three distinct sets of characteristics that may affect outcome. It is thus not, strict reporting a cohort study

Authors 'response: Thank you for your well-asserted point. We agree that the study population is a mixed group of participants and we will do our best to be transparent about that. We agree with you that the study population consisted of more than one single cohort and actually represents three groups: a) Patients included in the trial that received the intervention; b) Patients included in the trial that were in the control group and c) Patients that were not included in the trial because they failed to meet the inclusion criteria of needing to be at risk of developing persistent disability, defined using the Short form of the Örebro Musculoskeletal Pain Screening Questionnaire (Linton et al 2011) cut-off of 40 points.

This means that the sample includes patients who are at risk and patients who are not at risk of developing persistent disability. We wanted to include the widest possible group with the study population as we believe this is better representative of the primary care population that might be referred for physiotherapy. This study population is also consistent with another recent Swedish study of the SBT (Forsbrand et al 2017).

References:
Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire. Spine 2011;36(22):1891-5. doi: http://dx.doi.org/10.1097/BRS.0b013e3181f8f775.

Forsbrand M, Grahn B, Hill JC, Petersson IF, Sennehed CP, Stigmar K. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. BMC musculoskeletal disorders 2017;18(1):89.

Authors' action: We have changed the title. We have removed "cohort study" and changed to "a psychometric validation study": "Can the STarT Back Tool predict health related quality of life and work ability after an acute/subacute episode with back or neck pain? – a psychometric validation study in primary care", please see page 1, line 1-3.
We have also changed: "We conducted a prospective psychometric validation study with long-term follow up", please see page 5, line 97 in the Methods section. Through the whole manuscript the word "cohort" is removed.

…….

2. It should also be noted that these are not an inception cohort but a group who have self-referred for treatment of their back pain to a physiotherapy department

Authors 'response: We agree that this can be written more clearly.
Authors' action: We have added some text on this. Please see Methods, page 5, line 102-104:

“Participants were consecutively recruited between January 2013 and January 2014 from 35 primary care centers in the southern parts of Sweden, as part of an RCT 32. Patients that all applied for physiotherapy treatment on self-referral due to….”

…….

3. Some more detail is needed on screening ahead of consideration for study entry. That 35 primary care centres over one year only saw 329 people with acute or sub-acute low back pain who had self-referred for physiotherapy seems unlikely – this equates to less than ten per site per year

Authors ‘response:
We agree that this seems a very low number of patients for an inclusion period of one year and we can accept that this number is not representative of everyone who consulted primary care for one of these problems. We are aware of this and will add it as a limitation in the discussion. However, whilst our study population might be a selected sample of people, we don´t think that this will have substantially affected the psychometric validation questions examined in this study.

The study population for this study was identified when including patients in connection to a clinical trial (Sennehed CP, Holmberg S, Axen I, Stigmar K, Forsbrand M, Petersson IF, Grahn B. Early Workplace Dialogue in Physiotherapy Practice Improved Work Ability at One-Year Follow-Up - Workup a Randomised Controlled Trial in Primary Care. Pain 2018 doi: 10.1097/j.pain.0000000000001216 [published Online First: 2018/03/20])
and in that article you can find more details about the inclusion and exclusion of patients to the clinical trial.

For this study, it has to be clarified that it was 329 patients that completed the SBT at the first visit during one year and as in the clinical trial, we do not know if there were patients that due to different reasons (lack of time, forgot to ask) were not asked to be a part of the study. Furthermore, we do not know how many patients refused to participate. There were no registers on patients that were potentially eligible to the study. We agree that being able to report on how many potential patients that were available and not included, would have increased the quality of the trial and also this study. It is a limitation that we do not have data on eligible but non-participating patients.

Authors’ action:
We have added this as a limitation in the discussion, please see page 16, line 311-314:
“We accept that our study population (n=329) is unlikely to be representative of all individuals consulting primary care for acute/subacute BP and/or NP. However, even if they are a selected group of participants, we don´t think that this will have substantially affected the psychometric validation questions examined in this study.”
……..

4. The actual number of people recruited to the overall programme of work is unclear. Whilst what is described in this paper is internally consistent it is not externally consistent. The Trial registration website for this study indicates that 364 people were recruited to the trial upon which this study is based; whilst in this paper 162 people were included who had been part of the trial. I have not been able to identify a published protocol or final report of the randomised controlled trial to explore this further

Authors ‘response: The trial WorkUp included 364 patients over a period of two years, of whom 352 were eligible for analyses. In this study we have used only patients that were included during the first year of inclusion for the trial. There is no study protocol published, but the first article of the main

outcome is accepted for publication in Pain (Sennehed et al 2018-03-12).
Reference:
Sennehed CP, Holmberg S, Axen I, Stigmar K, Forsbrand M, Petersson IF, Grahn B. Early Workplace Dialogue in Physiotherapy Practice Improved Work Ability at One-Year Follow-Up - Workup a Randomised Controlled Trial in Primary Care. Pain 2018 doi: 10.1097/j.pain.0000000000001216 [published Online First: 2018/03/20])

Authors' action: We have added the accepted article of the clinical trial as a reference in the manuscript, please see the Methods, page 5, line 102-103: "Participants were consecutively recruited between January 2013 and January 2014 from 35 primary care centers in the southern parts of Sweden, as part of an RCT 31".
…….

5. Whilst the Startback tool has not been used to inform participant management the intervention arm of the trial specifically targeted psychosocial yellow flags for poor prognosis of low back pain. This approach covers the same domains as the StartBack Tool and might reduce any apparent effect of its use as a screening tool by addressing psychosocial risk factors. Without the main paper being available it is difficult to know how these factors might interact

Authors 'response: We agree on this point. The main paper was just recently accepted for publication in Pain (Sennehed CP et al 2018). Neither ÖMPSQ or SBT were used to target interventions. ÖMSPQ was used for screening for inclusion (≥40) and SBT was administered only for the purpose of psychometric testing. We agree that there is potential for an interaction between how the type of treatment given and the predictive abilities of the SBT. Given that the SBT's prognostic stratification is based on modifiable factors, there is a decrease in the tool's predictive ability in treatment contexts where those risk factors are effectively targeted (Morsø L et al 2013, Field J et al 2012). We appreciate that it was an overlap but we don´t think this will reduce the ability to perform the validation of the SBT in this study. We will highlight in the paper that the SBT was not used to direct and target interventions specifically. We also put an acknowledgement of this in the discussion.

References:
Morsø L, Kent P, Albert HB, Hill JC, Kongsted A, Manniche C. The predictive and external validity of the STarT Back Tool in Danish primary care. Eur Spine J 2013;22(8):1859–67.

Field J, Newell D. Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. Chiropr Man Therap 2012;20(1):17.

Authors' action: We have added some text, please see Discussion, page 19, line 392-395:
"In this study, the physiotherapists did not target treatment based on SBT. However, we accept that some of the constructs within the SBT may have been addressed by the intervention provided which may have affected SBTs ability to predict the above mentioned outcomes"

…….

6. That there are only full data on those included in the RCT makes interpretation of the final results extremely challenging. Fundamentally, we are being asked to believe that the StartBack Tool predicts outcome in terms of health related quality of life when we have not, and cannot be, presented with the baseline data for health related quality of life, and back pain severity, for the whole cohort. Thus we do not know if what is being reported for differences are those engendered by future effects of StartBack status at baseline or they are simply reflecting the relationship that was already present at baseline. It would not be surprising in pain severity, health related quality of life, and StartBack Status were associated at baseline. The authors need to work out how best to address this problem that

seriously impedes the interpretation of their findings.

Authors 'response: Thank you for this comment. We do agree that this is a problem. There are limitations with not having baseline data for the whole population. We are sorry but we only have access to baseline data for patients included in the RCT (n=160).
Although we don´t have baseline data on health related quality of life and work ability for the whole study population, we do have the data for most participants (2/3 of the population, 160/238) and for your information we have added a table with these data below:

Table. X. Health related quality of life and work ability at baseline for patients included in the RCT - total population and stratified by SBT risk groups, n=160.

SBT risk group Follow-up measure	Total population n=160	Low n=41	Medium n=92	High n=27
Health related quality of life[1]; median (range)	0.66 (-0.08-1)	0.73 (-0.003-0.88)	0.64 (-0.07-1)	0.21 (-0.08-0.73)
EQ-5D <0.6, n (%)	64 (41)	8 (20)	39 (43)	17 (65)
Work ability[2]; median (range)	6 (0-10)	7 (0-10)	6 (0-10)	5 (0-9)
WAS <8, n (%)	114 (73)	24 (59)	68 (76)	22 (82)

SBT, STarT Back Tool; EQ-5D, EuroQol five-dimension; WAS, Work Ability Score
[1]3 missing (1 person from each risk group)
[2]Where 0 equates to "completely unable to work" and 10 equates to "work ability at its best", 3 missing (1 from low risk, 2 from medium risk group)

Authors' action: We have addressed this problem in the discussion, as a limitation of the results from the multiple regression analyses. Please see Discussion, page 16, line 309-311:
"When recommending tools for use in primary care settings, preferably they should have been validated in large trials within this specific setting. However, as is the case with this study of the SBT, information from smaller studies is still of scientific value."

…….

7. The authors may wish to reflect rather more on the discriminant ability of the startback tool for predicting poor outcome when compared to other measures when compared , for example, to clinician assessment or baseline pain severity and relevance for clinical practice. It is suggested here that StartBack is marginally superior to other measures but AUCs of 0.68 and 0.73. Many may feel these data do not indicate a sufficiently high level of discrimination for this to be worthwhile.

Authors 'response: We agree that AUC values are not very high, but still around 0.7, which is considered as being acceptable (Hosmer Jr D, Lemeshaw A, 2000).
In this setting, we found these results of the AUC values and we are aware of that it is not a high level of discrimination. We know that we have to be cautious and not to exaggerate the findings of the AUC

values which correspond to the discriminative ability of the SBT. As we mention in the discussion (page 17, line 351-353), a variation of values of the AUC are expected and depends on, for example the characteristics of the population and possible explanations might be cultural and differences in treatment. Another possible explanation in variation of AUC values may be that a ROC curve analysis requires dichotomization of outcomes and the definitions of poor outcome may also have affected the results (please see page 17, line 353-355).

References:
Hosmer Jr D, Lemeshaw A, Sturdivant RX. Introduction to the Logistic Regression Model. Applied Logistic Regression 2000.

Authors' action:
We have added:" The AUC values are not very high, but still around 0.7, which is considered as acceptable 46, please see Discussion, page 17, line 337-338.

This is discussed more in detail in the discussion as mentioned above. Please see Discussion, page 17, line 351-355.

…….

8. A minor point is that in the background it I suggested that evidence based guidelines suggest use of brief questionnaires for risk stratification. To be specific it is UK NICE Guidance that recommends this and the second reference is this sentence does not refer to evidence based guidance

Authors 'response: Thank you for point on this. We agree.
Authors' action: We have changed and rephrased: "There are recommendations for the use of screening methods in health care to identify patients in early stages with the purpose to guide them to the best treatment 15-17 and also for enhancing return to work 18 19. The UK Nice guidance recommend using brief questionnaires to identify individuals of poor outcomes and stratify care20 but there is a lack of such tools that can be used in primary care", please see Introduction, page 4, line 75-79.
…….

Reviewer #3 (Prawit Janwantanakul):
Reviewer reports:

General comments
==========

This prospective cohort study aimed to evaluate the SBT's predictive validity for HRQoL and work ability outcomes at long-term follow-up in a population with acute/subacute back and/or neck pain. The topic is of interest in the field and the paper is very well-written. My main concern is about the inclusion of those with neck pain with and without low back pain into the study because the SBT is developed and validated to predict future disability due to low back pain. This may affect the internal validity of the findings and the authors did not state clearly on this issue.

Authors 'response: Thank you for your encouraging response. We have made changes in the manuscript and our response to each of your comments are described below. Please note that changes in the manuscript are marked in red.

Yes, we are aware of that SBT is developed for patients with low back pain. This is the first time, to our knowledge, that SBT is also applied to patients with neck pain. We believe that this is relevant,

since many patients in primary care apply for both neck and back pain.

Thank you for highlighting that we have not been enough clear about this issue. The SBT was recently validated against the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute/subacute back and/or neck pain in a Swedish primary care setting (as described earlier for reviewer #1, question number 1). It is actually the same population as we used in this study.

Authors' action: We have added: "The SBT is cross-culturally adapted and validated in Swedish24 and recently also for a population with both back and neck pain in primary care 25", please see Introduction, page 4, line 87-88.

…….

Specific comments

==========

1. INTRODUCTION

1a) The authors mentioned that the SBT is developed and validated to predict future disability due to low back pain of any duration. What is the justification for studying in a population with neck pain? Are there any limitations for translating findings of previous studies in LBP to a population with NP?

Authors 'response: The reason for studying SBT also for a population with neck pain is that both back pain and neck pain are highly prevalent in the general population and both are common reasons for seeking treatment in primary health care. It is also common to have neck pain together with back pain (Leijon O et al 2009, Nyman T et al 2007) and the SBT includes questions on modifiable physical and psychosocial risk factors which are, to a large extent, similar for both back and neck pain patients (Linton SJ 2000).

References:

Leijon O, Wahlstrom J, Mulder M. Prevalence of self-reported neck-shoulder-arm pain and concurrent low back pain or psychological distress: time-trends in a general population, 1990-2006. Spine 2009;34(17):1863-8. doi: http://dx.doi.org/10.1097/BRS.0b013e3181ab3397

Nyman T, Grooten WJ, Wiktorin C, Liwing J, Norrman L. Sickness absence and concurrent low back and neck-shoulder pain: results from the MUSIC-Norrtalje study.[Erratum appears in Eur Spine J. 2007 May;16(5):639-40]. European Spine Journal 2007;16(5):631-8.

Linton SJ. A review of psychological risk factors in back and neck pain. Spine 2000;25(9):1148-56.

Authors' action: We have added text and two new references (Leijon et al 2009, Nyman T et al 2007) to clarify that also concurrent BP and NP is common and therefore relevant to study: "To have concurrent BP and NP is also common11 and increases the risk of work disability further in the long-term12." Please see Introduction, page 4, line 71-72.

……

1b) Comments on using SBT, which is developed for LBP, in the population of NP would be useful. How appropriate is it?

Authors 'response: This is the first time that the SBT is applied to a population with neck pain. In a

study of concurrent validity (Forsbrand et al 2017) we did stratified analyses on pain site and found that there were no differences between the total study population, patients that had back pain only or patients with back and neck pain/neck pain only.

Reference:
Forsbrand M, Grahn B, Hill JC, Petersson IF, Sennehed CP, Stigmar K. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. BMC musculoskeletal disorders 2017;18(1):89.

Authors' action: We have added the reference above (Forsbrand M et al 2017); "The SBT is cross-culturally adapted and validated in Swedish25 and recently also for a population with both back and neck pain in primary care 26", please see Introduction, page 4, line 87-88.

There is also a comment on this in the discussion: "SBTs concurrent validity has earlier been studied for patients with back and/or neck pain 26 and a modified SBT have been tested to predict physical health outcome, using the SF-36 58 but this was the first time the predictive validity of the SBT was studied for the outcomes of HRQoL and work ability for individuals with both back and neck pain", please see Discussion, page 19, line 384-387

……

2. METHODS
2a) How was non-specific LBP and NP confirmed?

Authors 'response: Thank you for your question. We agree that this is not enough clear for the readers. Patients that applied through self-referral for BP or NP were all eligible to be included in the clinical trial, and also this validation study. At time for inclusion/exclusion the BP or NP were not further confirmed.

Patients that were pregnant, had severe pathology ("Red flags") or were not able to understand the Swedish language were not eligible to participate.
In the paper of the RCT study (Sennehed CP et al 2018) the word "non-specific" is not used and therefore we decided to omit this word and rewrite the text instead, please see below. We hope that the study population will be better described now.

Authors' action:
We have rewritten and clarified some text regarding participants and the procedure, please see Methods, page 5, line 103-109:
"Patients that all applied for physiotherapy treatment on self-referral due to an episode of acute/subacute (<12 weeks) non-specific back and/or neck pain, who were not currently on sick leave or had been on sick leave for less than 60 days and who had been working ≥4 consecutive weeks last year were asked to participate. It could be either a first episode or a recurrent episode of back and/or neck pain after a period of at least three months of no substantial pain. Patients that were pregnant, had severe pathology ("red flags") 32 or were not able to understand the Swedish language were not eligible to participate.

……

2b) It is not clear to me how patients were identified as having acute or subacute LBP and NP. Did you ask about previous episodes?

Authors ‘response: We asked them about how long they have had their current episode of back and/or neck pain. All patients with a new episode of BP or NP lasting for no more than 12 weeks, were eligible for inclusion. As mentioned in the rewritten text above, it could be either a first episode or a recurrent episode of back and/or neck pain after a period of at least three months of no substantial pain.

Authors’ action: We have clarified this in “Participants and procedure”, please see the answer on question 2a above.
“Patients that all applied for physiotherapy treatment on self-referral due to an episode of acute/subacute (<12 weeks) non-specific back and/or neck pain, who were not currently on sick leave or had been on sick leave for less than 60 days and who had been working ≥4 consecutive weeks last year were asked to participate. It could be either a first episode or a recurrent episode of back and/or neck pain after a period of at least three months of no substantial pain. (Methods, page 5, line 103-109)

…….

2c) Did you have exclusion criteria for the study?
Authors ‘response: In this psychometric study we identified patients alongside the inclusion to a clinical trial. All patients that applied for physiotherapy due to acute or subacute back and/or neck pain, who were not currently on sick leave or had been on sick leave for less than 60 days and who had been working ≥4 consecutive weeks last year, were not pregnant, had signs of severe illness (red flags) or did not understand Swedish were asked to participate. All these patients completed the SBT and formed study population for this psychometric study (n=329). We excluded patients that were older than 67 years or younger than 18 years, had incomplete SBT and also those who did not want to participate in the follow-up (see fig 1) and the final sample included 238 patients.

Authors’ action: We have added this in the Methods, please see page 5, line 108-109: “Patients that were pregnant, had severe pathology (“red flags”) 32 or were not able to understand the Swedish language were not eligible to participate.”
and at page 6, line 113-115: “Patients that were older than 67 years or younger than 18 years (n=3), declined participation (n=4), had any missing item on the SBT (n=11) or those who were lost to follow-up (n=73) were excluded.”

…….

2d) A statement about the completion of follow up is required.

Authors ‘response: Thank you for making us aware of this.
Authors’ action: We have described how the follow-ups were completed, please Methods see page 5, line 112 to page 6, line 115: “In all, 329 patients completed the SBT questionnaire and formed the population of this psychometric study. Patients that were older than 67 years or younger than 18 years (n=3), declined participation (n=4), had any missing item on the SBT (n=11) or those who were lost to follow-up (n=73) were excluded. “

…….

2e) The authors used a nonparametric approach which was chosen based on the distribution of the data. Please specify the reason for using a nonparametric statistic.

Authors ‘response: We used a non-parametric approach because the majority of our data came from

questionnaires with order statistics which are based on the ranks of observations. The data was not considered as normally distributed because of the ordered, categorical nature of the data. Such statistics play a central role in many non-parametric approaches.

Authors' action: No actions taken.

…….

3. RESULTS

3a) No description of the Kruskal Wallis test and Mann Whitney U test for studying if there were any differences between the SBT risk groups on follow-up data on poor or good HRQoL and work ability.

Authors 'response: Thank you for pointing this out. We agree that we have not been enough clear about this.
The results of the Kruskal Wallis test for HRQoL scores on follow-up is found in Results, page 11, line 238-239: "There were statistically significant differences in the distribution of HRQoL scores (n=238) between the SBT low, medium and high risk groups at long-term follow-up ($p<0.001$)" .
The results of the Kruskal Wallis test for work ability scores is found in Results, page 11, line 241-243: "We also found differences in the distribution of work ability (WAS) scores (n=235) between the SBT low, medium and high risk groups at long-term follow-up ($p<0.001$)".

We only used Mann Whitneys test to confirm potential differences (medians) between the three SBT risk groups (low vs medium, low vs high and medium vs high) and it was therefore not necessary to present these results in the manuscript. We have now clarified about how we thought about it in the Methods, please see below.

(We used Chi-square test for trend to confirm potential differences concerning "poor" or "good" HRQoL and work ability, dichotomous outcome)

Authors' action:
We have clarified and added some text in Methods, page 8, line 176-178: "The Kruskal Wallis test was used to study if there were any differences between the SBT risk groups on follow-up data on poor or good HRQoL and work ability (median), respectively. Potential differences were confirmed with Mann Whitney U-test."

To make the results of the Kruskal Wallis test clearer, we have added the "p-values of Kruskal Wallis test" also in table 2, please see Table 2, page 12 in the Results.
…….

3b) It is unclear how the medians in Table 2 are derived. The authors did not clearly state how their outcomes were collected during the follow up.

Authors 'response: For the patients that were included in the clinical trial (n=160), the follow-up outcomes were collected at a planned physiotherapy visit at 12-month through self-reports. For patients that were not included in the trial (due to ÖMPSQ <40) we sent postal questionnaires. Concerning the medians we have clarified how they were derived, please see below.

Authors' action: We have added text on this, please see Methods, page 6, line 120-123:
"RCT patients (n=160) received either structured physiotherapy treatment with a workplace intervention (RCT intervention) or structured physiotherapy without a workplace intervention (RCT control)31 and were followed up at the planned 12-months (median 12 months, range 11-19) follow-

up. Not RCT patients received usual primary care and were followed up by postal questionnaires."
We have clarified how the medians were derived, please see Methods, page 8, line 176-178: "The Kruskal Wallis test was used to study if there were any differences between the SBT risk groups on follow-up data on poor or good HRQoL and work ability (median), respectively. "

We have added the word "also" in Methods, page 7, line 155:
"Health Related Quality of Life was also dichotomized into "poor" HRQoL (EQ-5D <0.6) and "good" HRQoL…."
….and in Methods, page 8, line 164: "Work ability was also dichotomized using a previously published cut-off score 43 into "poor" work ability (WAS<8 points) and "good" work ability (WAS ≥8 points).

…….

3c) I wonder whether the authors had information about duration from the onset of symptoms to the first encounter with a clinician. Would it affect the findings of the study?

Authors 'response: When including patients to the clinical trial we were aiming at including patients with no long-standing BP or NP. All patients were asked about the duration of their problems at the first visit to the physiotherapist, and patients that had pain that have lasted for more than 12 weeks were not asked to participate.

Authors' action: We have added a few words to make this clearer in the Methods, page 5, line 103-106: "Patients that all applied for physiotherapy treatment on self-referral due to an episode of acute/subacute (<12 weeks) non-specific back and/or neck pain, who were not currently on sick leave or had been on sick leave for less than 60 days and who had been working ≥4 consecutive weeks last year were asked to participate".

…….

4. DISCUSSION

4a) Discussion regarding the effect of the broad variation in time to follow up on the findings would be useful.

Authors 'response: We agree; this can be discussed more in depth. The variation in follow-up time encompasses only patients that were not included in the clinical trial. They were sent postal questionnaires, and also reminders if they did not answer. In the regression model we saw that the time for follow-up did not have an impact on the outcome.

Authors' action: We have added some text about the broad variation in time to follow up in the Discussion, page 16, line 316-322:
"The time to follow-up varied between patients in our study which may have influenced the results. The optimal time point for identifying patients at risk of developing persistent back pain may vary and is a forum for discussion 49. In our study, two third of the study population (n=160) were in the RCT and were followed-up at a planned physiotherapy visit at 12 months. For not RCT patients (n=78) the ambition was also to follow-up at 12 months but these patients were followed-up with postal questionnaires and due to practical reasons there were a wider variation on the time for follow-up. This is of course a limitation, but did not have impact on the results in the regression analyses."

…….

4b) What about the clinical implications of the findings? How can clinicians use the findings in their day-to-day work?

Authors 'response: We are happy to address the clinical implications. There is need for short questionnaires that are easy to distribute and interpret, especially in primary care. The results from this study shows that SBT can be used as a prognostic tool, both for patients with back pain and neck pain, with the aim to predict HRQoL and work ability. These outcomes are highly relevant in this group of patients, that correspond to about 25 % of all sick leave in Sweden and is one of the most common patients in primary care.

Authors' action:
We have clarified in the Discussion, page 19, line 396-400: "The results of this study suggest that the SBT can be used as a prognostic tool in primary care for subgroup identification of acute/subacute back and/or neck pain patients at risk of poor long-term HRQoL and/or work ability outcome. The information about important risk factors may help clinicians in primary care to develop personalized treatment strategies which are a priority in research65."

……

Reviewer# 4 (Achim Elfering):
Reviewer reports:

This interesting study reports the prognostic validity of the STarT Back Tool in prediction of health-related quality of life and work ability in a large sample of primary care patients across 35 health care centers. The writing of the study is clear and well-developped. I have only a few points author(s) should consider ina revision:

1. To investigate a varying number of patients from 35 health care centers means to have a multilevel-data structure. Author(s) should report whether health-related quality of life and work ability differed systematically between 35 health care centers (Intra-class correlations). If so, a multilevel-regression approach is more adequate than the current approach.

Authors 'response: Thank you for your encouring response. We have made changes in the manuscript and our response to each of your comments are described below. Please note that changes in the manuscript are marked in red.

We understand your question but since we identified our study population alongside a cluster randomized trial, all intervention units and all reference units, respectively were merged together. There were differences in how many patients that were included at the different units, so to compare for systematic differences in HRQoL and work ability at baseline, between the units, would have been improper. In this study we also used patients that were not included in the trial (n=78) and we have no baseline data on HRQoL or work ability for them.

Authors' action: No actions taken.
……
2. Some data were reported separately for patients having back pain or « patients reporting neck pain plus back pain patients with neck or shoulder pain (NP) of without back pain » (cf. Table 1). This separation is unclear to me. Moreover, it would be more clear and interesting when the prognostic validity of the STarT Back Tool would have been tested not only for the total of patients but also separately for patients reporting solely LBP, solely neck and shoulder pain, and those patients who reported LBP and neck and shoulder pain.

Authors 'response: We understand that this might be a little bit confusing. In this study we do not have access to ICD-10 diagnosis for all patients, and this also justifies why we used the phrasing "pain site". Pain sites are based on self-reports from the patients on question number 2 in SBT (which is about having neck pain). This question enabled us to identify two groups: a) back pain only and b) back pain with neck pain/neck pain only. This was the only way that was possible. We considered to also include pain site in the regression analyses but as many patients have not only BP or NP, they have often concurrent BP and NP, we decided to look at the group "as one group" and to not separate BP patients from NP patients. This makes the results of this study applicable to a common clinical situation. This topic is also described and discussed in a previous study where we used the same study population as in this study (Forsbrand et al 2017).
Reference:
Forsbrand M, Grahn B, Hill JC, Petersson IF, Sennehed CP, Stigmar K. Comparison of the Swedish STarT Back Screening Tool and the Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire in patients with acute or subacute back and neck pain. BMC musculoskeletal disorders 2017;18(1):89.

Authors' action: We have added a few words on this in the discussion, page 16, lines 327-329: "In this study we included both patients with neck pain and back pain. Since this group of patients often have concurrent pain from the back or neck50, we decided to not include this in the regression analysis".

…….

3. Time of follow-up after primary care did vary much. Author(s) should discuss time in more depth (see as an examples for a systematic investigation of follow-up time after primary care : Melloh, M., Elfering, A., Käser, A., Rolli Salathé, C., Barz, T., Röder, C., & Theis, J.-C. (2015). What is the best time point to identify patients at risk of developing persistent low back pain? Journal of Back and Musculoskeletal Rehabilitation, 28(2), 267-276. doi: 10.3233/BMR-140514

Authors 'response: Thank you for your advice and reference. In the article the authors conclude that 12 weeks is an appropriate time for follow-up to predict persistent LBP at 6 months. However, all patients in our study had acute or subacute pain when they applied for care and that means that they have had pain up to 12 weeks. It seems to me that to screen patients for poor outcome at this stage (around 12 weeks) would be an appropriate time point according to this article.

In our study two third of the study population (n=162) that were included in the clinical trial were followed-up at a planned physiotherapy visit at 12 months. For the patients that were excluded from the trial (n=78) the ambition was also to follow-up at 12 months. These patients were followed-up with a postal questionnaire and due to practical reasons there were a variation on the time for follow-up. This is of course a limitation, but did not have an impact of the results in the regression analyses.
Authors' action:
We have added your suggested reference (Melloh M et al 2014) and addressed this in the Discusson, page 16, line 316-322: "The time to follow-up varied between patients in our study which may have influenced the results. The optimal time point for identifying patients at risk of developing persistent back pain may vary and is still a forum for discussion 49. In our study, two third of the study population (n=160) were in the RCT and were followed-up at a planned physiotherapy visit at 12 months. For not RCT patients (n=78) the ambition was also to follow-up at 12 months but these patients were followed-up with postal questionnaires and due to practical reasons there were a wider variation on the time for follow-up. This is of course a limitation, but did not have impact on the results in the regression analyses."

……

4. Working ability was assessed, but to me it was unclear whether the sample indeed included only individuals who were employed or had an active working status at baseline.

Authors 'response: We agree; this is not clear. As we have mentioned in the manuscript we only have access to limited baseline data for patients not included in the clinical trial (n=78), but for those included (n=160) we have more information. The vast majority of the patients in this group were employed (96%) and we have no reason to believe that the group that did not meet the inclusion criteria for the RCT (<40 points ÖMPSQ-short) differed from that. To be included in the study, patients must have been working ≥4 consecutive weeks last year which means that they had a link to the labor market or were at work the last year.
Authors' action: We have clarified and added text on this in the Methods, page 5, line 103-106: "Patients that all applied for physiotherapy treatment on self-referral due to an episode of acute/subacute (<12 weeks) non-specific back and/or neck pain, who were not currently on sick leave or had been on sick leave for less than 60 days and who had been working ≥4 consecutive weeks last year were asked to participate".

……

5. Two thirds of participants were also included in a RCT (intervention or control group). One third of participants did not. Author(s) tested whether RCT participation predicted health-related quality of life and work ability. Author(s) should also test whether RCT participation interacted with STarT Back Tool in prediction of health-related quality of life and work ability.

Authors 'response: Thank you for your comment on this. It has been unclear how the variables in the regression analysis was assessed (see comments from Reviewer #1, question number 4). We have now added a table on this (Table 3) and we hope that it will be clearer for you and for the readers how we used RCT participation in the regression analysis.
We built a multiple regression model where all independent variables were put into the model at the same time and we tested how the different variables interacted with STarT Back Tool in prediction of health related quality of life and work ability. Please see Table 3, page 13 and Methods -Statistical analysis, page 8, line 183-184 and page 9, line 185-187.

Authors' action: We have clarified how the independent variables (including RCT participation) were assessed and treated in the regression analysis in a new table (Table 3) in the Results section, page 13.

VERSION 2 – REVIEW

REVIEWER	Martin Underwood Warwick CTU, UK
REVIEW RETURNED	29-Apr-2018

GENERAL COMMENTS	No further comments

REVIEWER	Professor Dr. Prawit Janwantanakul Department of Physical Therapy, Faculty of Allied Health Sciences, Chulalongkorn University, Bangkok, Thailand
REVIEW RETURNED	04-May-2018

GENERAL COMMENTS	None.

REVIEWER	Victor Hoe Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, MALAYSIA
REVIEW RETURNED	15-May-2018

GENERAL COMMENTS	Thank you for your positive response to my comments and suggestions. All the issues highlighted have been addressed satisfactorily. I have no further comments.

REVIEWER	Achim Elfering University of Bern, Switzerland
REVIEW RETURNED	16-May-2018

GENERAL COMMENTS	All my points were addressed.