BMJ Open

Modifying the STarT Back Tool for use with patients with other musculoskeletal conditions: does it work?

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012445
Article Type:	Research
Date Submitted by the Author:	26-Apr-2016
Complete List of Authors:	Hill, Jonathan; Keele University, Institute of primary care and health sciences Afolabi, Ebenezer ; Keele University, Institute of Primary Care and Health Sciences Lewis, Martyn; Keele University, Institute of Primary Care and Health Sciences Dunn, Kate; Keele University, Institute of Primary Care and Health Sciences Roddy, Edward; Keele University, Institute for Primary Care and Health Sciences van der Windt, Danielle; Keele University, Institute of Primary Care and Health Sciences Foster, Nadine; Keele University, Institute of Primary Care Health Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	General practice / Family practice, Rheumatology, Sports and exercise medicine
Keywords:	Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, prognosis, prospective cohort, STarT Back Tool, Risk stratification

Title Page 26th April 2016

Title: Modifying the STarT Back Tool for use with patients with other musculoskeletal conditions: does it work?

Corresponding Author:

Dr Jonathan C Hill[1]

Senior Lecturer

Institute of primary care and health sciences

Keele University, Staffs

ST5 5BG

UK

01782733900

j.hill@keele.ac.uk

Co-authors: Hill JC, Afolabi EK, Lewis M, Dunn KM, Roddy E, van der Windt DA, Foster NE.

Affiliation:

1. Institute of Primary Care and Health Sciences, Keele University, Staffordshire, ST5 5BG, UK

Key words: Musculoskeletal Pain, Questionnaires, Prognosis, Prospective studies, STarT Back Tool

Manuscript Word count =2900

ABSTRACT: (Word count = 300)

Objectives: The STarT Back Tool has good predictive performance for non-specific low back pain in primary care. We therefore aimed to investigate whether a modified STarT Back Tool predicted outcome with a broader group of musculoskeletal patients, and assessed the consequences of using existing risk-group cut-points across different pain regions.

Setting: Secondary analysis of prospective data from 2 cohorts: 1) outpatient musculoskeletal physiotherapy services (PhysioDirect trial n=1887), and 2) musculoskeletal primary-secondary care interface services (SAMBA study n=1082)

Participants: Patients with back, neck, upper-limb, lower-limb, or multi-site pain with a completed modified STarT Back Tool (baseline) and six-month physical health outcome (SF-36).

Outcomes: Area-Under-the-Receiving-Operator-Curve (AUCs) tested discriminative abilities of the tool's baseline score for identifying poor six-month outcome (SF-36 lower-tertile Physical-Component-Score). Risk-group cut-points were tested using sensitivity and specificity for identifying poor outcome using a) Youden's J statistic, and b) a clinically determined rule that specificity should not fall below 0.7 (false-positive rate less than 30%).

Results: In PhysioDirect and SAMBA poor six-month physical health was 18.5%, 28.2% respectively. Modified STarT Back Tool score AUCs for predicting outcome in back pain were 0.72, 0.79 in PhysioDirect and SAMBA respectively, neck 0.82, 0.88, upper limb 0.79, 0.86, lower limb 0.77, 0.83 and multi-site pain 0.83, 0.82. Differences between pain region AUCs were non-significant. Optimal cut-points to discriminate low and medium/high risk-groups depended on pain region and clinical services.

Conclusion: A modified STarT Back Tool similarly predicts six-month physical health outcome across five musculoskeletal pain regions. However, use of consistent risk-group cut-points was not possible and resulted in poor sensitivity (too many with long-term disability being missed) or specificity (too many with good outcome inaccurately classified as 'at risk') for some pain regions. The draft tool is now being refined and validated within a new programme of research for a broader musculoskeletal population.

Strengths and limitations of this study

- First study to demonstrate that modified STarT Back Tool items are similarly predictive of six-month physical health across different musculoskeletal pain regions.
- Similar findings were demonstrated in 2 large musculoskeletal cohorts (n=1887, n=1082).
- However, existing STarT Back Tool cut-offs were not optimal across all pain regions and resulted in poor sensitivity (too many with long-term disability being missed) or specificity (too many with good outcome inaccurately classified as 'at risk', leading to over treatment)
- A limitation of the study was that the original STarT Back Tool was not included in these two datasets, so a direct comparison between the performance of the original and modified STarT Back Tool versions for patients with low back pain was not possible.

INTRODUCTION

The Keele STarT Back Tool is designed to stratify patients with low back pain according to their risk of future physical disability, in order that prognostic subgroups can receive matched treatment.(1) For example, individuals at a low risk of persistent disabling problems can be reassured and discouraged from receiving unnecessary treatments and investigations, whilst those at high risk can matched to treatment which combines physical and psychological approaches.(2-4) A large randomised trial testing a risk stratification approach (use of the STarT Back Tool and matched treatments) for low back pain in comparison to best current care demonstrated superior clinical and cost outcomes.(5) In addition, an implementation study testing risk stratification for patients with low back pain in routine general practice demonstrated significant improvements in physical function and time off work, sickness certification rates, and reductions in healthcare costs compared to usual non-stratified care.(2) Since low back pain accounts for only 17% of all musculoskeletal pain consultations in primary care,(6) if a similar screening tool could be used for patients with other common pain presentations, such as neck pain and knee pain, then there could be potential for stratified care to make a greater impact for patients and healthcare services.

A previous systematic review of 45 cohort studies(7) reported that prognostic factors are often similar across different musculoskeletal presentations, with 11 factors predicting poor outcome at follow-up for at least two different musculoskeletal pain problems. Other studies have similarly shown that a generic set of baseline factors (pain intensity, episode duration, pain interference, depression and co-morbid pain problems) predicts risk of a poor outcome across different pain regions including back pain, headache, facial pain and knee pain, regardless of the specific location of pain or underlying pathology.(8-12) These studies indicate that it might be possible to utilise the same prognostic factors as those included within the STarT Back Tool to discriminate risk status for a much larger group of musculoskeletal pain patients than those consulting with low back pain. The key benefit of using a single tool to stratify patients with a wide range of musculoskeletal conditions rather than multiple site-specific prognostic screening tools is its simplicity for use in busy clinical practice.

Whilst the likely value and acceptability of extending risk stratification to patients with other common musculoskeletal pain is as yet unknown, evidence suggests that the majority of General Practitioners (GPs) consider prognosis to be important in their clinical decision-making for musculoskeletal treatment.(13) Despite the widespread support for prognostic information, the clinical reality is that predicting outcome in these patients is not always easy and patient's risk status is not typically included within medical records.(14) GPs are not alone in wanting information about patients' likely prognosis over time, as more than 80% of musculoskeletal patients also want prognostic information from their GP, although less than a third actually receive this information.(14) Existing musculoskeletal prognostic tools are available (e.g. Linton and Hallden,(15) and von Korff et al,(16, 17)). However, these prognostic tools were not designed or tested to support clinical decisions in primary care about matched treatments (stratified care); only the STarT Back Tool has been specifically developed and tested to guide patient treatment matching.

The aim of this study was therefore to investigate the performance of a modified STarT Back Tool for predicting future outcome for a broader group of musculoskeletal pain patients. Specific objectives were to compare the predictive performance of a modified STarT Back Tool for patients with musculoskeletal pain in different body regions and assess the consequences (false positive and false negative rates) of using existing STarT Back Tool score cut-points for classifying patients as medium/high risk across different pain regions (neck, back, upper limb, lower limb, and multi-site pain).

METHODS

Design:

This study involved pre-specified further analysis of existing datasets from two prospective cohorts of adults with musculoskeletal conditions consulting in two different services in the NHS, UK.

Patient population:

1) The PhysioDirect trial included 2249 adult musculoskeletal patients taking part in a randomised trial comparing a PhysioDirect service (telephone-based physiotherapy assessment and advice) with usual physiotherapy care.(18-20) Primary outcome data (physical health measured using the SF-36v2 physical component score) at six-month follow-up and baseline modified STarT Back Tool score were available for 1887 patients (84%) and were included in this analysis. The trial was conducted in four NHS community physiotherapy services in four different areas of England (Bristol, Somerset, Stoke-on-Trent, and Cheshire). Adults (aged ≥18 years) who were referred by 94 general practitioners (covering a wide range of geographical areas and populations), or who referred themselves for physiotherapy for a musculoskeletal problem, were eligible for the trial. Patients completed postal questionnaires at baseline and six-months after randomisation. Details about the PhysioDirect patient sample have been published.(18)

2) The SAMBA study was an observational cohort of adults attending an NHS musculoskeletal clinical assessment and treatment service at the primary-secondary care interface.(21, 22) The study population included 2166 patients referred from primary care and subsequently triaged to musculoskeletal and back pain interface clinics in Stoke-on-Trent Primary Care Trust (PCT) over a 12-month period. Primary outcome data at six-month follow-up (physical health measured using the SF-36v2 physical component score) was available for 1174 patients (54%) who formed the study population for this evaluation. All adults (aged ≥18 years) capable of giving written informed consent were eligible to participate in the study. Patients completed study questionnaires before their first appointment during which consent was obtained and six-months after that initial clinic appointment. Details of the SAMBA study sample have been published.(22)

Modifying the STarT Back Tool:

The original STarT Back Tool includes nine items of which five concern psychosocial factors (fear, catastrophising, anxiety, depression, and bothersomeness). Both the PhysioDirect trial and SAMBA study included the STarT Back Tool's psychosocial items within their baseline questionnaires.(1)

These items were used without modification as they were developed from generic tools and are not specific to low back pain. However, the four further items of the original STarT Back Tool that capture three physical factors (referred pain from the back down the leg, co-morbid pain in the neck and shoulder, and physical function with walking and dressing items) are specific to low back pain and therefore these items in their original form needed to be replaced by similar items that were applicable for all musculoskeletal patients. We therefore used proxy items for these outcome domains that were available in both datasets. The STarT Back Tool's two 'function' items (walking and dressing) were replaced by items from the generic EQ-5D(23) ('I have some problems in walking about', Y/N and 'I have some problems washing or dressing myself", Y/N), and we used item 7 from the SF-12(24) ('How much bodily pain have you had?' with positive responses defined as 'extremely' or 'very severe') instead of the original STarT Back Tool item for co-morbid pain in the neck or shoulder. To score the modified STarT Back Tool, responses from these 8 items were summed (range 0-8) for all patients in both datasets. The original STarT Back Tool cut-off of 0-3 positive items was used to classify patients as at low risk and 4 or more as medium or high risk. There were no reference standards for psychological distress in either the PhysioDirect or SAMBA datasets and so in this analysis we did not seek to examine the ability of the modified STarT Back Tool to identify a high risk only group.

Defining the body regions of pain:

Participants were asked to indicate the primary site of their musculoskeletal pain for which they had sought treatment. From this information patients were categorized as having one of the following regional pain problems: neck, back (thoracic or lumbar), upper limb, lower limb, or multi-site pain (pain in more than one region).

Defining poor outcome:

The standardised summary score for the Physical-Component-Score (PCS) of the Short Form 36 Health Survey (SF-36) is population normalised (0 is worst physical health and 100 is best physical health) and was classified by tertiles (<=33, 34-66, >66) as has been used previously (25, 26) with a six-month poor outcome defined using the most severe tertile (<=33). Outcome was defined as poor physical health at six-month follow-up using the SF-36 PCS because this was the most appropriate physical function outcome score available in both studies, and it has demonstrated good validity and responsiveness in this population.(27-29)

Statistical analysis:

All analyses were conducted separately for the two datasets and a descriptive comparison of the modified baseline STarT Back Tool scores (mean and standard deviation [SD]) and proportion with poor six-month outcome (SF-36 PCS <=33) calculated. Descriptive statistics using means and standard deviations were used to examine the modified STarT Back Tool score's distribution and investigate potential floor or ceiling effects (>10% of either lowest or maximum score).(30)

Predictive performance (discrimination) was assessed by calculating ROC curve AUCs for baseline modified STarT Back Tool total scores against six-month poor outcome (dichotomised as poor/good) for each of the five different bodily pain presentations and their equality compared using STATA's 'roccomp' command to establish if AUC differences were statistically significant.

To examine whether the optimal subgroup cut-point on the modified STarT Back Tool total score to discriminate low from medium/high risk for poor six-month outcome was consistent across the five different pain regions and across the two datasets, we used two methods based on sensitivity and specificity of each potential cut-point. Firstly, we used Youden's J Statistic which is calculated as sensitivity + specificity -1 for each potential cut-point and the optimal cut-point is the tool score with the highest value.(31, 32) Secondly, we *a priori* agreed that specificity should not fall below 0.7, as lower values would mean potentially over-treating more than 30% of medium/high risk patients, which was considered an unacceptable level for an efficient matched treatment approach.

RESULTS

Distribution of the modified STarT Back Tool scores in both datasets:

In the PhysioDirect trial sample (n=1887) the 8-item modified STarT Back Tool score at baseline was normally distributed with a mean (SD) of 3.35 (2.09); 8.4% had the lowest score (0) and 2.2% had the maximum score (8). The distribution of primary pain regions was reported by clinicians as: lower limb 31.1%, back 28.7%, upper limb 23.5%, neck 11.8%, and multi-site pain 4.8%. The six-month SF-36 PCS mean (SD) was 43.7 (10.9) with 18.5% having a 'poor outcome' in their physical health at six-month follow-up.

In the SAMBA study sample (n=1082) the 8-item modified STarT Back Tool score at baseline was not normally distributed but had roughly equal numbers of all possible scores with a mean (SD) of 3.95 (2.65); 12.6% had the lowest score (0) and 10.9% had the maximum score (8). The distribution of primary pain sites was reported by patients as: lower limb 30.8%, back 26.7%, upper limb 23.8%, multi-site pain 13.4% and neck 5.4%. The six-month SF-36 PCS mean (SD) was 38.41 (12.76) with 28.2% having a 'poor outcome' in their physical health at six-month follow-up.

Predictive performance of the modified STarT Back Tool score across pain regions in both datasets:

Predictive performance of the modified STarT Back Tool as determined by ROC curve AUCs ranged from 0.72 to 0.83 and was not found to be statistically different across different pain regions in the PhysioDirect trial (p=0.130) and SAMBA study (p= 0.098) (presented in Figures 1 & 2).

Optimal modified STarT Back Tool score cut-offs in both datasets:

Table 1 reports sensitivity, specificity, and the Youden's J statistic for each possible modified STarT Back Tool score cut-point at baseline for each pain region. The results demonstrate that the optimal STarT Back Tool baseline score cut-point for discriminating 'poor outcome' at six-month follow-up was not consistent across pain regions. For example, among (PhysioDirect) patients with neck, back and multi-site pain the optimal STarT Back Tool cut-point for discriminating 'poor outcome' was 5 or

more, whereas this was 4 or more for those with upper limb and lower limb as their primary pain site.

Discussion

This is the first study to demonstrate that a modified STarT Back Tool is similarly predictive of six-month physical health (defined by worst tertile of the SF-36) across different musculoskeletal pain regions. Predictive performance determined by AUCs for the 8-item modified STarT Back Tool total score was in fact slightly higher for neck, upper limb, lower limb and multi-site pain than for back pain, although differences were not statistically significant. The results therefore demonstrate that the prognostic factors included within the STarT Back Tool are predictive of six-month physical health across a range of musculoskeletal pain regions, not just back pain. However, the results demonstrated that the optimal baseline STarT Back Tool score cut-point for identifying individuals with poor physical health outcome was neither consistent across different pain regions, nor across clinical services (community physiotherapy services [PhysioDirect trial] and primary-secondary care interface services [SAMBA study]. This finding was consistent regardless of method used to determine the optimal modified STarT Back Tool score cut-point (Youden's J statistic or an a priori defined maximum false positive rate of 30%). This implies that the existing original STarT Back Tool score cut-point (4 or more out of 9) used to allocate patients with low back pain to the medium/high risk subgroups cannot simply be applied to patients with other musculoskeletal pain presentations or in different clinical services. This is likely to be due to differences in patient characteristics across services such as episode duration, which is known to influence the performance of the original STarT Back Tool.(33) It is also likely that individual modified STarT Back Tool items are not equally applicable to patients with pain in the five regions.(34) For example, the item about walking difficulties is likely to be less relevant and therefore less predictive of physical health outcome for patients with upper limb pain than for those with lower limb or spinal pain.

The findings of this study concur with previous evidence suggesting that the same set of prognostic variables can be used to estimate prognosis of patients with different musculoskeletal pain presentations.(7, 15, 17) The STarT Back Tool uses biopsychosocial constructs known to predict persistent disability among patients with low back pain, such as: difficulty with walking and dressing, pain elsewhere, fear avoidance, pain catastrophising, anxiety and low mood.(1) However, the STarT Back Tool is not just a prognostic index, but is used to stratify patients for different matched treatments. An important issue highlighted by this analysis is that if clinicians simply modify the STarT Back Tool for use with other musculoskeletal pain patients, they are at risk of matching patients to inappropriate treatments. It is also apparent that future translation and validation studies of the STarT Back Tool need to carefully consider adopting the same STarT Back Tool score cut-points as used in the original UK STarT Back Tool study(1) without first testing if these cut-points are appropriate for their own clinical populations. Based on these findings our team has begun to further refine and validate an improved stratification tool – the Keele STarT MSK Tool - which will be specifically designed for use with primary care patients consulting with the five most common musculoskeletal pain presentations in a new programme of research.

The strengths of our analyses reported here include the large sample sizes of both the PhysioDirect and SAMBA studies and the opportunity to examine optimal cut-points in patients with different pain sites and in different NHS musculoskeletal services. An additional strength was that both studies used the same measure of physical health (SF-36), had the same six-month follow-up time-point and included patients whose pain could be classified into the same musculoskeletal pain regions. Given the potential weakness of using the Youden's J Statistic to define optimal cut-points for discriminating between low and medium/high risk, we also used a clinically determined guide (maximum false positive rate), which showed similar inconsistencies in optimal cut-off between regional pain site and clinical setting. One weakness is that the original STarT Back Tool was not included in these two datasets, which meant a direct comparison between the performance of the original and modified versions for patients with low back pain was not possible. The choice of poor outcome at six-months using the lowest tertile on the SF-12 was also relatively arbitrary, but served the purpose of this analysis to compare outcome between different regional pain sites, making the exact definition of poor outcome less critical to the study aims.

The implications from this analysis are that, despite good predictive performance of the modified STarT Back Tool in patients with pain in different regions of the body, clinicians need to cautiously consider the choice of cut-points when using a modified STarT Back Tool for musculoskeletal pain regions other than low back pain. The results suggest existing cut-points may lead to an inefficiency in healthcare resource use, with too many patients with a likely long-term disability being missed, or too many patients with good outcome being inaccurately classified as 'at risk', which may result in over treatment of low risk groups.

Conclusions

A modified version of the STarT Back Tool has similar predictive performance when used for patients with musculoskeletal pain in different body regions. However, the cut-points used to identify patients with a poor physical health outcome at six-month follow-up are not consistent across pain regions or clinical services. Further research is underway to refine and validate a new Keele STarT MSK Tool which will form part of a new stratified care approach to be tested in a randomised controlled trial.

Contributors: JCH, DvdW, NEF, ER and KD conceived and designed the research; NF and ER were responsible for the modified STarT Back Tool being embedded within the PhysioDirect and SAMBA datasets; JCH, EA and ML analysed the data and all authors were involved in the interpretation of the data analysis; JCH, ER, EA, ML, DvdW, KD and NF were involved in the drafting of the manuscript and its revision for important intellectual content, and gave final approval for the manuscript submission.

Funding: This research is funded by: the NIHR Programme Grants for Applied Research programme (Grant reference number: RP-PG-1211-20010), the NIHR Primary Care Career Scientist Award to Professor Nadine Foster [CSA 04/03], and Arthritis Research UK [13413]. The views and opinions expressed within this manuscript do not necessarily reflect those of DH/NIHR. The funding bodies were not involved in the design of the study outlined within this protocol, and had no involvement in the writing and revision of this manuscript. DvdW received funding from an MRC Partnership Grant for the PROGnosis RESearch Strategy (PROGRESS) group (grant reference number: G0902393).

The SAMBA study was supported by an Arthritis Research UK Integrated Clinical Arthritis Centre Grant (17684), the Arthritis Research UK Primary Care Centre Grant (18139), funding secured from Stoke-on-Trent Primary Care Trust (PCT), and service support through the West Midlands North CLRN.

Competing interests: None declared.

Patient consent Obtained: Yes, full written informed consent for all participants.

Ethics approval: This was secondary data analysis of two studies which both obtained ethical approval through written informed consent.

Data sharing statement: Data can be accessed via the Keele data repository at http:// http://www.keele.ac.uk/pchs/publications/datasharingresources/

References

1. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum. 2008;59(5):632-41.
2. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Ann Fam Med. 2014;12(2):102-11.
3. Main CJ, Sowden G, Hill JC, Watson PJ, Hay EM. Integrating physical and psychological approaches to treatment in low back pain: the development and content of the STarT Back trial's 'high-risk' intervention (StarT Back; ISRCTN 37113406). Physiotherapy. 2012;98(2):110-6.
4. Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P, et al. Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). Fam Pract. 2012;29(1):50-62.
5. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet. 2011;378(9802):1560-71.
6. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord. 2010;11:144.
7. Mallen CD, Peat G, Thomas E, Dunn KM, Croft PR. Prognostic factors for musculoskeletal pain in primary care: a systematic review. Br J Gen Pract. 2007;57(541):655-61.
8. Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. Pain. 2005;117(3):304-13.
9. Von Korff M, Dunn KM. Chronic pain reconsidered. Pain. 2008;138(2):267-76.
10. Dunn KM, Croft PR, Main CJ, Von Korff M. A prognostic approach to defining chronic pain: replication in a UK primary care low back pain population. Pain. 2008;135(1-2):48-54.
11. Thomas E, Dunn KM, Mallen C, Peat G. A prognostic approach to defining chronic pain: application to knee pain in older adults. Pain. 2008;139(2):389-97.
12. Muller S, Thomas E, Dunn KM, Mallen CD. A prognostic approach to defining chronic pain across a range of musculoskeletal pain sites. Clin J Pain. 2013;29(5):411-6.
13. Mallen CD, Peat G, Porcheret M, Croft P. The prognosis of joint pain in the older patient: general practitioners' views on discussing and estimating prognosis. Eur J Gen Pract. 2007;13(3):166-8.
14. Mallen CD, Peat G. Discussing prognosis with older people with musculoskeletal pain: a cross-sectional study in general practice. BMC Fam Pract. 2009;10:50.
15. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain. 1998;14(3):209-15.
16. Von Korff M, Shortreed SM, Saunders KW, LeResche L, Berlin JA, Stang P, et al. Comparison of back pain prognostic risk stratification item sets. J Pain. 2014;15(1):81-9.
17. Von Korff M. Tailoring chronic pain care by brief assessment of impact and prognosis: comment on "Point-of-care prognosis for common musculoskeletal pain in older adults". JAMA Intern Med. 2013;173(12):1126-7.
18. Salisbury C, Foster NE, Hopper C, Bishop A, Hollinghurst S, Coast J, et al. A pragmatic randomised controlled trial of the effectiveness and cost-effectiveness of

'PhysioDirect' telephone assessment and advice services for physiotherapy. Health Technol Assess. 2013;17(2):1-157, v-vi.
19. Foster NE, Williams B, Grove S, Gamlin J, Salisbury C. The evidence for and against 'PhysioDirect' telephone assessment and advice services. Physiotherapy. 2011;97(1):78-82.
20. Salisbury C, Foster NE, Bishop A, Calnan M, Coast J, Hall J, et al. 'PhysioDirect' telephone assessment and advice services for physiotherapy: protocol for a pragmatic randomised controlled trial. BMC Health Serv Res. 2009;9:136.
21. Roddy E, Zwierska I, Dawes P, Hider SL, Jordan KP, Packham J, et al. The Staffordshire arthritis, musculoskeletal, and back assessment (SAMBA) study: a prospective observational study of patient outcome following referral to a primary-secondary care musculoskeletal interface service. BMC Musculoskelet Disord. 2010;11:67.
22. Roddy E, Zwierska I, Jordan KP, Dawes P, Hider SL, Packham J, et al. Musculoskeletal clinical assessment and treatment services at the primary-secondary care interface: an observational study. Br J Gen Pract. 2013;63(607):e141-8.
23. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. Med Care. 2002;40(5):442-6.
24. Ware JE, Jr. SF-36 health survey update. Spine (Phila Pa 1976). 2000;25(24):3130-9.
25. Corica F, Corsonello A, Apolone G, Mannucci E, Lucchetti M, Bonfiglio C, et al. Metabolic syndrome, psychological status and quality of life in obesity: the QUOVADIS Study. Int J Obes (Lond). 2008;32(1):185-91.
26. Dorynska A, Pajak A, Kubinova R, Malyutina S, Tamosiunas A, Pikhart H, et al. Socioeconomic circumstances, health behaviours and functional limitations in older persons in four Central and Eastern European populations. Age Ageing. 2012;41(6):728-35.
27. Wittink H, Turk DC, Carr DB, Sukiennik A, Rogers W. Comparison of the redundancy, reliability, and responsiveness to change among SF-36, Oswestry Disability Index, and Multidimensional Pain Inventory. Clin J Pain. 2004;20(3):133-42.
28. Angst F, Verra ML, Lehmann S, Gysi F, Benz T, Aeschlimann A. Responsiveness of the cervical Northern American Spine Society questionnaire (NASS) and the Short Form 36 (SF-36) in chronic whiplash. Clin Rehabil. 2012;26(2):142-51.
29. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB, Kinemax Outcomes G. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. J Bone Joint Surg Am. 2001;83-A(12):1856-64.
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
31. Yin J, Samawi H, Linder D. Improved nonparametric estimation of the optimal diagnostic cut-off point associated with the Youden index under different sampling schemes. Biom J. 2016.
32. Greenhouse SW, Cornfield J, Homburger F. The Youden index: letters to the editor. Cancer. 1950;3(6):1097-101.
33. Morso L, Kongsted A, Hestbaek L, Kent P. The prognostic ability of the STarT Back Tool was affected by episode duration. Eur Spine J. 2016;25(3):936-44.
34. Butera KA, Lentz TA, Beneciuk JM, George SZ. Preliminary Evaluation of a Modified STarT Back Screening Tool Across Different Musculoskeletal Pain Conditions. Phys Ther. 2016.

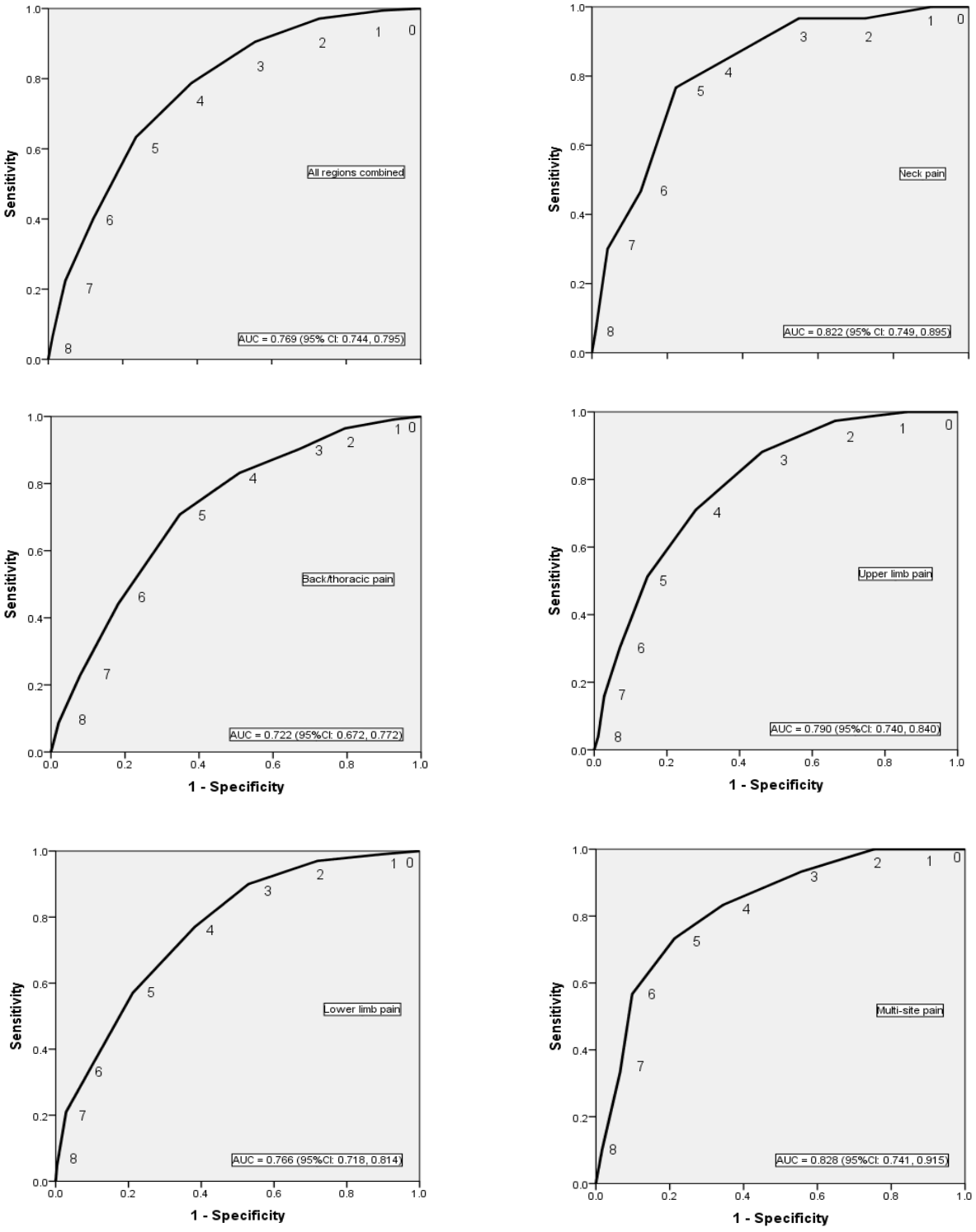

Figure 1. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor outcome (SF-36 PCS <=33) by different pain regions in the PhysioDirect dataset.

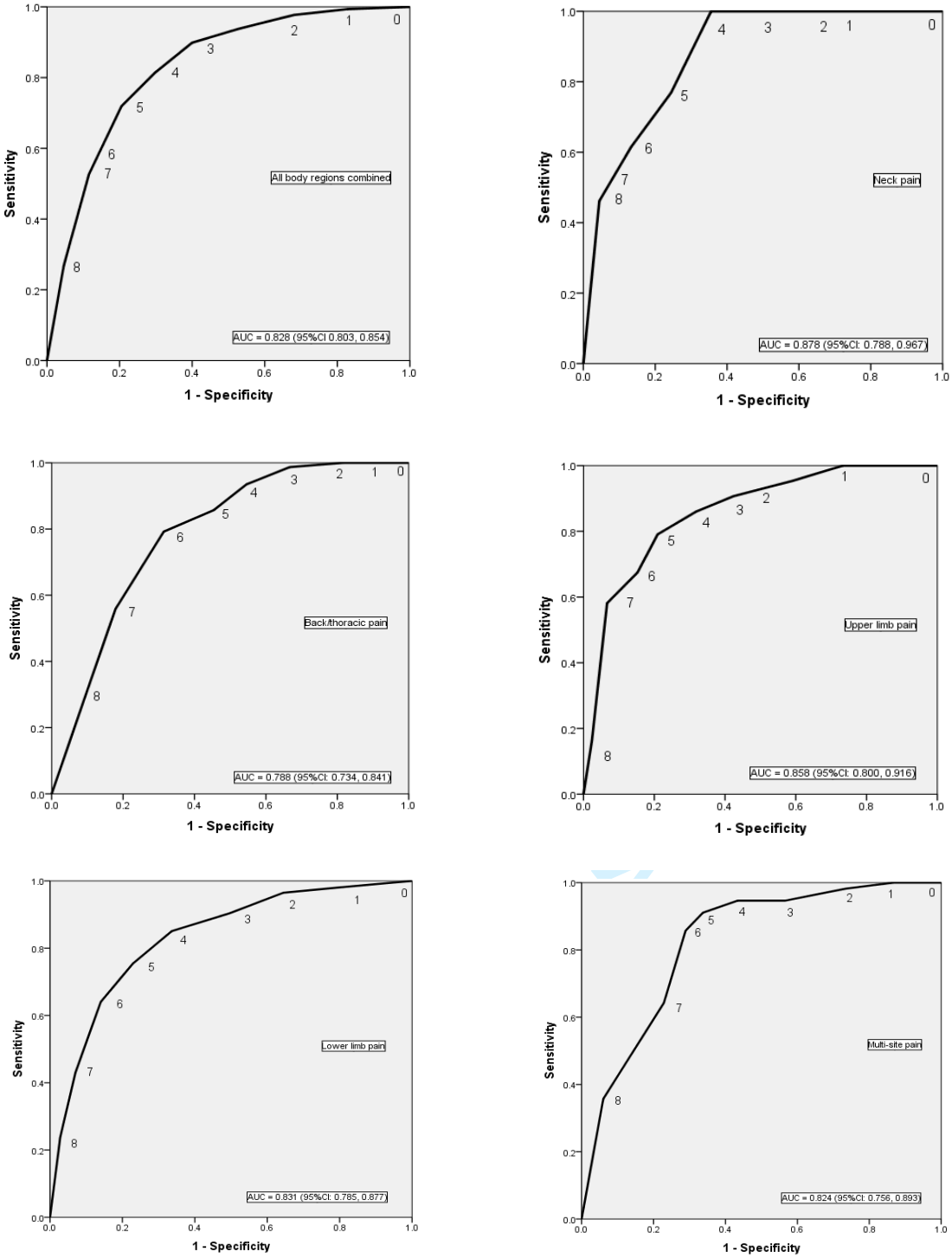

Figure 2. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor outcome (SF-36 PCS <=33) by different pain regions in the SAMBA dataset.

Table 1. Identifying optimal modified STarT Back Tool cut-points for each pain region using a) Youden's J statistic and b) a clinically defined maximum specificity of 0.7.

		PhysioDirect trial data			*SAMBA study data*		
Pain region	**Modified STarT Back Tool cut-point**	**Sens**	**Spec**	**Youden's**	***Sens***	***Spec***	***Youden's***
Neck	3	0.967	0.451	0.418	*1*	*0.556*	*0.556*
Neck	4	0.833	0.668	0.501	*1*	*0.644*	*0.644*
Neck	5	0.767	0.777*	0.544	*0.769*	*0.756**	*0.525*
Neck	6	0.467	0.87	0.337	*0.615*	*0.867*	*0.482*
Neck	7	0.3	0.959	0.259	*0.538*	*0.911*	*0.449*
Neck	8	0.067	0.99	0.057	*0.462*	*0.956*	*0.418*
Back	3	0.903	0.329	0.232	*0.987*	*0.333*	*0.32*
Back	4	0.832	0.491	0.323	*0.935*	*0.454*	*0.389*
Back	5	0.708	0.652	0.36	*0.857*	*0.546*	*0.403*
Back	6	0.442	0.818*	0.26	*0.792*	*0.686*	*0.478*
Back	7	0.23	0.921	0.151	*0.558*	*0.821**	*0.379*
Back	8	0.088	0.979	0.067	*0.273*	*0.913*	*0.186*
Upper limb	3	0.882	0.538	0.42	*0.907*	*0.576*	*0.483*
Upper limb	4	0.711	0.72*	0.431	*0.86*	*0.681*	*0.541*
Upper limb	5	0.513	0.853	0.366	*0.791*	*0.79**	*0.581*
Upper limb	6	0.303	0.929	0.232	*0.674*	*0.848*	*0.522*
Upper limb	7	0.158	0.973	0.131	*0.581*	*0.933*	*0.514*
Upper limb	8	0.039	0.989	0.028	*0.163*	*0.976*	*0.139*
Lower limb	3	0.9	0.47	0.37	*0.904*	*0.505*	*0.409*
Lower limb	4	0.77	0.618	0.388	*0.851*	*0.664*	*0.515*
Lower limb	5	0.57	0.789*	0.359	*0.754*	*0.771**	*0.525*
Lower limb	6	0.36	0.895	0.255	*0.64*	*0.86*	*0.5*
Lower limb	7	0.21	0.971	0.181	*0.43*	*0.93*	*0.36*
Lower limb	8	0.05	0.996	0.046	*0.237*	*0.972*	*0.209*
Multi-site pain	3	0.933	0.443	0.376	*0.946*	*0.434*	*0.38*
Multi-site pain	4	0.833	0.656	0.489	*0.946*	*0.566*	*0.512*
Multi-site pain	5	0.733	0.787*	0.52	*0.911*	*0.663*	*0.574*
Multi-site pain	6	0.567	0.902	0.469	*0.857*	*0.711**	*0.568*
Multi-site pain	7	0.333	0.934	0.267	*0.643*	*0.771*	*0.414*
Multi-site pain	8	0.1	0.984	0.084	*0.357*	*0.94*	*0.297*

Grey shaded row indicates Youden's optimal score cut-point for predicting 6-month outcome

* indicates specificity was >0.7 according to pre-defined clinical criteria

BMJ Open

Modifying the STarT Back Tool for use with patients with other musculoskeletal conditions: does it work?

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012445.R1
Article Type:	Research
Date Submitted by the Author:	29-Jun-2016
Complete List of Authors:	Hill, Jonathan; Keele University, Institute of primary care and health sciences Afolabi, Ebenezer ; Keele University, Institute of Primary Care and Health Sciences Lewis, Martyn; Keele University, Institute of Primary Care and Health Sciences Dunn, Kate; Keele University, Institute of Primary Care and Health Sciences Roddy, Edward; Keele University, Institute for Primary Care and Health Sciences van der Windt, Danielle; Keele University, Institute of Primary Care and Health Sciences Foster, Nadine; Keele University, Institute of Primary Care Health Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	General practice / Family practice, Rheumatology, Sports and exercise medicine
Keywords:	Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, prognosis, prospective cohort, STarT Back Tool, Risk stratification

Title Page 22nd June 2016

Title: Modifying the STarT Back Tool for use with patients with other musculoskeletal conditions: does it work?

Corresponding Author:

Dr Jonathan C Hill[1]

Senior Lecturer

Institute of primary care and health sciences

Keele University, Staffs

ST5 5BG

UK

01782733900

j.hill@keele.ac.uk

Co-authors: Hill JC, Afolabi EK, Lewis M, Dunn KM, Roddy E, van der Windt DA, Foster NE.

Affiliation:

1. Institute of Primary Care and Health Sciences, Keele University, Staffordshire, ST5 5BG, UK

Key words: Musculoskeletal Pain, Questionnaires, Prognosis, Prospective studies, STarT Back Tool

Manuscript Word count =2900

ABSTRACT: (Word count = 300)

Objectives: The STarT Back Tool has good predictive performance for non-specific low back pain in primary care. We therefore aimed to investigate whether a modified STarT Back Tool predicted outcome with a broader group of musculoskeletal patients, and assessed the consequences of using existing risk-group cut-points across different pain regions.

Setting: Secondary analysis of prospective data from 2 cohorts: 1) outpatient musculoskeletal physiotherapy services (PhysioDirect trial n=1887), and 2) musculoskeletal primary-secondary care interface services (SAMBA study n=1082)

Participants: Patients with back, neck, upper-limb, lower-limb, or multi-site pain with a completed modified STarT Back Tool (baseline) and six-month physical health outcome (SF-36).

Outcomes: Area-Under-the-Receiving-Operator-Curve (AUCs) tested discriminative abilities of the tool's baseline score for identifying poor six-month outcome (SF-36 lower-tertile Physical-Component-Score). Risk-group cut-points were tested using sensitivity and specificity for identifying poor outcome using a) Youden's J statistic, and b) a clinically determined rule that specificity should not fall below 0.7 (false-positive rate less than 30%).

Results: In PhysioDirect and SAMBA poor six-month physical health was 18.5%, 28.2% respectively. Modified STarT Back Tool score AUCs for predicting outcome in back pain were 0.72, 0.79 in PhysioDirect and SAMBA respectively, neck 0.82, 0.88, upper limb 0.79, 0.86, lower limb 0.77, 0.83 and multi-site pain 0.83, 0.82. Differences between pain region AUCs were non-significant. Optimal cut-points to discriminate low and medium/high risk-groups depended on pain region and clinical services.

Conclusion: A modified STarT Back Tool similarly predicts six-month physical health outcome across five musculoskeletal pain regions. However, use of consistent risk-group cut-points was not possible and resulted in poor sensitivity (too many with long-term disability being missed) or specificity (too many with good outcome inaccurately classified as 'at risk') for some pain regions. The draft tool is now being refined and validated within a new programme of research for a broader musculoskeletal population.

Strengths and limitations of this study

- First study to demonstrate that modified STarT Back Tool items are similarly predictive of six-month physical health across different musculoskeletal pain regions.
- A limitation of the study was that the original STarT Back Tool was not included in these two datasets, so a direct comparison between the performance of the original and modified STarT Back Tool versions for patients with low back pain was not possible.

INTRODUCTION

The Keele STarT Back Tool is designed to stratify patients with low back pain according to their risk of future physical disability, in order that prognostic subgroups can receive matched treatment.(1) For example, individuals at a low risk of persistent disabling problems can be reassured and discouraged from receiving unnecessary treatments and investigations, whilst those at high risk can matched to treatment which combines physical and psychological approaches.(2-4) A large randomised trial testing a risk stratification approach (use of the STarT Back Tool and matched treatments) for low back pain in comparison to best current care demonstrated superior clinical and cost outcomes.(5) In addition, an implementation study testing risk stratification for patients with low back pain in routine general practice demonstrated significant improvements in physical function and time off work, sickness certification rates, and reductions in healthcare costs compared to usual non-stratified care.(2) Since low back pain accounts for only 17% of all UK primary care musculoskeletal consultations in general practice,(6) if a similar screening tool could be used for patients with other common pain presentations, such as neck pain and knee pain, then there could be potential for stratified care to make a greater impact for patients and healthcare services.

A previous systematic review of 45 cohort studies(7) reported that prognostic factors are often similar across different musculoskeletal presentations, with 11 factors predicting poor outcome at follow-up for at least two different musculoskeletal pain problems. Other studies have similarly shown that a generic set of baseline factors (pain intensity, episode duration, pain interference, depression and co-morbid pain problems) predicts risk of a poor outcome across different pain regions including back pain, headache, facial pain and knee pain, regardless of the specific location of pain or underlying pathology.(8-12) These studies indicate that it might be possible to utilise the same prognostic factors as those included within the STarT Back Tool to discriminate risk status for a much larger group of musculoskeletal pain patients than those consulting with low back pain. The key benefit of using a single tool to stratify patients with a wide range of musculoskeletal conditions rather than multiple site-specific prognostic screening tools is its simplicity for use in busy clinical practice.

Whilst the likely value and acceptability of extending risk stratification to patients with other common musculoskeletal pain is as yet unknown, evidence suggests that the majority of General Practitioners (GPs) consider prognosis to be important in their clinical decision-making for musculoskeletal treatment.(13) Despite the widespread support for prognostic information, the clinical reality is that predicting outcome in these patients is not always easy and patient's risk status is not typically included within medical records.(14) GPs are not alone in wanting information about patients' likely prognosis over time, as more than 80% of musculoskeletal patients also want prognostic information from their GP, although less than a third actually receive this information.(14) Existing musculoskeletal prognostic tools are available (e.g. Linton and Hallden,(15) and von Korff et al,(16, 17)). However, these prognostic tools were not designed or tested to support clinical decisions in primary care about matched treatments (stratified care); only the STarT Back Tool has been specifically developed and tested to guide patient treatment matching.

The aim of this study was therefore to investigate the performance of a modified STarT Back Tool for predicting future physical health outcome for a broader group of musculoskeletal pain patients. Specific objectives were to compare the predictive performance of a modified STarT Back Tool for patients with musculoskeletal pain in different body regions and assess the consequences (false positive and false negative rates) of using existing STarT Back Tool score cut-points for classifying patients as medium/high risk across different pain regions (neck, back, upper limb, lower limb, and multi-site pain).

METHODS

Design:

This study involved pre-specified further analysis of existing datasets from two prospective cohorts of adults with musculoskeletal conditions consulting in two different services in the NHS, UK. Full ethical approval for both these studies was obtained and patients provided written informed consent prior to their research participation.

Patient population:

1) The PhysioDirect trial included 2249 adult musculoskeletal patients taking part in a randomised trial comparing a PhysioDirect service (telephone-based physiotherapy assessment and advice) with usual physiotherapy care.(18-20) Primary outcome data (physical health measured using the SF-36v2 physical component score) at six-month follow-up and baseline modified STarT Back Tool score were available for 1887 patients (84%) and were included in this analysis. The trial was conducted in four NHS community physiotherapy services in four different areas of England (Bristol, Somerset, Stoke-on-Trent, and Cheshire). Adults (aged ≥18 years) who were referred by 94 general practitioners (covering a wide range of geographical areas and populations), or who referred themselves for physiotherapy for a musculoskeletal problem, were eligible for the trial. Patients completed postal questionnaires at baseline and six-months after randomisation. Details about the PhysioDirect patient sample have been published.(18) For the study reported here we used patients from both the control and intervention arms.

2) The SAMBA study was an observational cohort of adults attending an NHS musculoskeletal clinical assessment and treatment service at the primary-secondary care interface.(21, 22) The study population included 2166 patients referred from primary care and subsequently triaged to musculoskeletal and back pain interface clinics in Stoke-on-Trent Primary Care Trust (PCT) over a 12-month period. Primary outcome data at six-month follow-up (physical health measured using the SF-36v2 physical component score) and the modified STarT Back Tool score at baseline was available for 1082 patients (50%) who formed the study population for this evaluation. All adults (aged ≥18 years) capable of giving written informed consent were eligible to participate in the study. Patients completed study questionnaires before their first appointment during which consent was obtained and six-months after that initial clinic appointment. Details of the SAMBA study sample have been published.(22)

Modifying the STarT Back Tool:

The original STarT Back Tool includes nine items of which five concern psychosocial factors (fear, catastrophising, anxiety, depression, and bothersomeness). Both the PhysioDirect trial and SAMBA study included the STarT Back Tool's psychosocial items within their baseline questionnaires.(1) These items were used without modification as they were developed from generic tools and are not specific to low back pain. However, the four further items of the original STarT Back Tool that capture three physical factors (referred pain from the back down the leg, co-morbid pain in the neck and shoulder, and physical function with walking and dressing items) are specific to low back pain and therefore these items in their original form needed to be replaced by similar items that were applicable for all musculoskeletal patients. We therefore used proxy items for these outcome domains that were available in both datasets. The STarT Back Tool's two 'function' items (walking and dressing) were replaced by items from the generic EQ-5D(23) ('I have some problems in walking about', Y/N and 'I have some problems washing or dressing myself", Y/N), and we used item 7 from the SF-12(24) ('How much bodily pain have you had?' with positive responses defined as 'extremely' or 'very severe') instead of the original STarT Back Tool item for co-morbid pain in the neck or shoulder. It was not possible to replace 'referred pain from the back down the leg' with an item that was suitable for all musculoskeletal pain and so this construct of the 'spread of pain' was omitted from the modified tool. To score the modified STarT Back Tool, responses from these 8 items were summed (range 0-8) for all patients in both datasets. The original STarT Back Tool cut-off of 0-3 positive items was used to classify patients as at low risk and 4 or more as medium or high risk. There were no reference standards for psychological distress in either the PhysioDirect or SAMBA datasets and so in this analysis we did not seek to examine the ability of the modified STarT Back Tool to identify a high risk only group.

Defining the body regions of pain:

Participants were asked to indicate the primary site of their musculoskeletal pain for which they had sought treatment. From this information patients were categorized as having one of the following regional pain problems: neck, back (thoracic or lumbar), upper limb, lower limb, or multi-site pain (pain in more than one region).

Defining physical health outcome:

The standardised summary score for the Physical-Component-Score (PCS) of the Short Form 36 Health Survey (SF-36) is population normalised (0 is worst physical health and 100 is best physical health) and was classified by tertiles (<=33, 34-66, >66) as has been used previously (25, 26) with a six-month poor outcome defined using the most severe tertile (<=33). Outcome was defined as poor physical health at six-month follow-up using the SF-36 PCS because this was the most appropriate physical function outcome score available in both studies, and it has demonstrated good validity and responsiveness in this population.(27-29)

Statistical analysis:

All analyses were conducted separately for the two datasets and a descriptive comparison of the modified baseline STarT Back Tool scores (mean and standard deviation [SD]) and proportion with poor six-month physical health outcome (SF-36 PCS <=33) calculated. Descriptive statistics using means and standard deviations were used to examine the modified STarT Back Tool score's distribution and investigate potential floor or ceiling effects (>10% of either lowest or maximum score).(30)

Predictive performance (discrimination) was assessed by calculating ROC curve AUCs for baseline modified STarT Back Tool total scores against six-month poor physical health outcome (dichotomised as poor/good) for each of the five different bodily pain presentations and their equality compared using STATA's 'roccomp' command to establish if AUC differences were statistically significant.

To examine whether the optimal subgroup cut-point on the modified STarT Back Tool total score to discriminate low from medium/high risk for poor six-month physical health outcome was consistent across the five different pain regions and across the two datasets, we used two methods based on sensitivity and specificity of each potential cut-point. Firstly, we used Youden's J Statistic which is calculated as sensitivity + specificity -1 for each potential cut-point and the optimal cut-point is the tool score with the highest value.(31, 32) Secondly, we *a priori* agreed that specificity should not fall below 0.7, as lower values would mean potentially over-treating more than 30% of medium/high risk patients, which was considered an unacceptable level for an efficient matched treatment approach.

In this study we were not able to identify optimal subgroup cut-points on the modified STarT Back Tool to distinguish between medium and high risk patients as there were no reference standards for psychological distress in the two available datasets. The original STarT Back Tool used these reference standards to identify distress 'caseness' at baseline, and identified the optimal cut-point to screen for these distressed 'cases' using a psychological subscale score. Without these reference standards for psychological distress, we were limited to determining optimal subgroup cut-points on the total scale score between low and medium/high risk alone.

RESULTS

Distribution of the modified STarT Back Tool scores in both datasets:

In the PhysioDirect trial sample (n=1887) the 8-item modified STarT Back Tool score at baseline was normally distributed with a mean (SD) of 3.35 (2.09); 8.4% had the lowest score (0) and 2.2% had the maximum score (8). The distribution of primary pain regions was reported by clinicians as: lower limb 31.1%, back 28.7%, upper limb 23.5%, neck 11.8%, and multi-site pain 4.8%. The six-month SF-36 PCS mean (SD) was 43.7 (10.9) with 18.5% having a 'poor outcome' in their physical health at six-month follow-up. The mean age was 48 years old and 60% of the sample were female.

In the SAMBA study sample (n=1082) the 8-item modified STarT Back Tool score at baseline was not normally distributed but had roughly equal numbers of all possible scores with a mean (SD) of 3.95

(2.65); 12.6% had the lowest score (0) and 10.9% had the maximum score (8). The distribution of primary pain sites was reported by patients as: lower limb 30.8%, back 26.7%, upper limb 23.8%, multi-site pain 13.4% and neck 5.4%. The six-month SF-36 PCS mean (SD) was 38.41 (12.76) with 28.2% having a 'poor outcome' in their physical health at six-month follow-up. The mean age was 51 years old and 57% were female.

Predictive performance of the modified STarT Back Tool score across pain regions in both datasets:

Predictive performance of the modified STarT Back Tool as determined by ROC curve AUCs ranged from 0.72 to 0.83 and was not found to be statistically different across different pain regions in the PhysioDirect trial (p=0.130) and SAMBA study (p= 0.098) (presented in Figures 1 & 2).

Optimal modified STarT Back Tool score cut-offs in both datasets:

Table 1 reports sensitivity, specificity, and the Youden's J statistic for each possible modified STarT Back Tool score cut-point at baseline for each pain region. The results demonstrate that the optimal STarT Back Tool baseline score cut-point for discriminating 'poor outcome' at six-month follow-up was not consistent across pain regions. For example, among (PhysioDirect) patients with neck, back and multi-site pain the optimal STarT Back Tool cut-point for discriminating 'poor outcome' was 5, whereas this was 4 for those with upper limb and lower limb as their primary pain site.

Discussion

This is the first study to demonstrate that a modified STarT Back Tool is similarly predictive of six-month physical health (defined by worst tertile of the SF-36) across different musculoskeletal pain regions. Predictive performance determined by AUCs for the 8-item modified STarT Back Tool total score was in fact slightly higher for neck, upper limb, lower limb and multi-site pain than for back pain, although differences were not statistically significant. The results therefore demonstrate that the prognostic factors included within the STarT Back Tool are predictive of six-month physical health across a range of musculoskeletal pain regions, not just back pain. However, the results demonstrated that the optimal baseline STarT Back Tool score cut-point for identifying individuals with poor physical health outcome was neither consistent across different pain regions, nor across clinical services (community physiotherapy services [PhysioDirect trial] and primary-secondary care interface services [SAMBA study]. This finding was consistent regardless of method used to determine the optimal modified STarT Back Tool score cut-point (Youden's J statistic or an a priori defined maximum false positive rate of 30%). This implies that the existing original STarT Back Tool score cut-point (4 or more out of 9) used to allocate patients with low back pain to the medium/high risk subgroups cannot simply be applied to patients with other musculoskeletal pain presentations or in different clinical services. This is likely to be due to differences in patient characteristics across services such as episode duration, which is known to influence the performance of the original STarT Back Tool.(33) It is also likely that individual modified STarT Back Tool items are not equally applicable to patients with pain in the five regions.(34) For example, the item about walking difficulties is likely to be less relevant and therefore less predictive of physical health outcome for patients with upper limb pain than for those with lower limb or spinal pain. A

key message from this study is the value and importance of testing the capabilities of the STarT Back Tool in different settings and patient populations and not presuming that existing primary care subgroup cut-points will be the same in other groups. If wider validity is demonstrated this will help strengthen the case for the general applicability of the tool.

The findings of this study concur with previous evidence suggesting that the same set of prognostic variables can be used to estimate prognosis of patients with different musculoskeletal pain presentations.(7, 15, 17) The STarT Back Tool uses biopsychosocial constructs known to predict persistent disability among patients with low back pain, such as: difficulty with walking and dressing, pain elsewhere, fear avoidance, pain catastrophising, anxiety and low mood.(1) However, the STarT Back Tool is not just a prognostic index, but is used to stratify patients for different matched treatments. An important issue highlighted by this analysis is that if clinicians simply modify the STarT Back Tool for use with other musculoskeletal pain patients, they are at risk of matching patients to inappropriate treatments. It is also apparent that future translation and validation studies of the STarT Back Tool need to carefully consider adopting the same STarT Back Tool score cut-points as used in the original UK STarT Back Tool study(1) without first testing if these cut-points are appropriate for their own clinical populations. Based on these findings our team has begun to further refine and validate an improved stratification tool – the Keele STarT MSK Tool - which will be specifically designed for use with primary care patients consulting with the five most common musculoskeletal pain presentations in a new programme of research. Whilst our study was not able to examine optimal high risk subgroup cut-offs for 'distressed' patients, a previous cross-sectional study [34] in a US physical therapy population has compared the relationships between a modified STarT Back Tool and psychological measures in people with different pain regions. It found that regardless of pain body region higher modified STarT Back Tool scores were associated with higher levels of kinesiophobia, catastrophising, fear avoidance, anxiety, and depressive symptoms. The strengths of our analyses reported here include the large sample sizes of both the PhysioDirect and SAMBA studies and the opportunity to examine optimal cut-points in patients with different pain sites and in different NHS musculoskeletal services. An additional strength was that both studies used the same measure of physical health (SF-36), had the same six-month follow-up time-point and included patients whose pain could be classified into the same musculoskeletal pain regions. Given the potential weakness of using the Youden's J Statistic to define optimal cut-points for discriminating between low and medium/high risk, we also used a clinically determined guide (maximum false positive rate), which showed similar inconsistencies in optimal cut-off between regional pain site and clinical setting. One weakness is that the original STarT Back Tool was not included in these two datasets, which meant a direct comparison between the performance of the original and modified versions for patients with low back pain was not possible. The choice of poor physical health outcome at six-months using the lowest tertile on the SF-12 was also relatively arbitrary, but served the purpose of this analysis to compare outcome between different regional pain sites, making the exact definition of poor outcome less critical to the study aims. It should be noted that the different levels of poor clinical outcome between the PhysioDirect (18.5%) and Samba (28.2%) studies could be due to the different settings and design of these two studies and it is possible this may have influenced the findings.

The implications from this analysis are that, despite good predictive performance of the modified STarT Back Tool in patients with pain in different regions of the body, clinicians need to cautiously consider the choice of cut-points when using a modified STarT Back Tool for musculoskeletal pain regions other than low back pain. The results suggest existing cut-points may lead to an inefficiency in healthcare resource use, with too many patients with a likely long-term disability being missed, or too many patients with good physical health outcome being inaccurately classified as 'at risk', which may result in over treatment of low risk groups.

Conclusions

A modified version of the STarT Back Tool has similar predictive performance when used for patients with musculoskeletal pain in different body regions. However, the cut-points used to identify patients with a poor physical health outcome at six-month follow-up are not consistent across pain regions or clinical services. Further research is underway to refine and validate a new Keele STarT MSK Tool which will form part of a new stratified care approach to be tested in a randomised controlled trial.

Author Contribution: JCH, DvdW, NEF, ER and KD conceived and designed the research; NF and ER were responsible for the modified STarT Back Tool being embedded within the PhysioDirect and SAMBA datasets; JCH, EA and ML analysed the data and all authors were involved in the interpretation of the data analysis; JCH, ER, EA, ML, DvdW, KD and NF were involved in the drafting of the manuscript and its revision for important intellectual content, and gave final approval for the manuscript submission.

Funding: This research is funded by: the NIHR Programme Grants for Applied Research programme (Grant reference number: RP-PG-1211-20010), the NIHR Primary Care Career Scientist Award to Professor Nadine Foster [CSA 04/03], and Arthritis Research UK [13413]. Professor Foster holds an NIHR Research Professorship (NIHR-RP-2011-015). The views and opinions expressed within this manuscript do not necessarily reflect those of DH/NIHR. The funding bodies were not involved in the design of the study outlined within this protocol, and had no involvement in the writing and revision of this manuscript. DvdW received funding from an MRC Partnership Grant for the PROGnosis RESearch Strategy (PROGRESS) group (grant reference number: G0902393).

The SAMBA study was supported by an Arthritis Research UK Integrated Clinical Arthritis Centre Grant (17684), the Arthritis Research UK Primary Care Centre Grant (18139), funding secured from Stoke-on-Trent Primary Care Trust (PCT), and service support through the West Midlands North CLRN.

Competing interests: None declared.

Patient consent Obtained: Yes, full written informed consent for all participants.

Ethics approval: This was secondary data analysis of two studies which both obtained ethical approval through written informed consent.

Data sharing statement: Data can be accessed via the Keele data repository at http:// http://www.keele.ac.uk/pchs/publications/datasharingresources/

Acknowledgements:

We thank the patients and clinical teams that participated in the two studies.

References

1. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum. 2008;59(5):632-41.
2. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Ann Fam Med. 2014;12(2):102-11.
3. Main CJ, Sowden G, Hill JC, Watson PJ, Hay EM. Integrating physical and psychological approaches to treatment in low back pain: the development and content of the STarT Back trial's 'high-risk' intervention (StarT Back; ISRCTN 37113406). Physiotherapy. 2012;98(2):110-6.
4. Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P, et al. Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). Fam Pract. 2012;29(1):50-62.
5. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet. 2011;378(9802):1560-71.
6. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord. 2010;11:144.
7. Mallen CD, Peat G, Thomas E, Dunn KM, Croft PR. Prognostic factors for musculoskeletal pain in primary care: a systematic review. Br J Gen Pract. 2007;57(541):655-61.
8. Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. Pain. 2005;117(3):304-13.
9. Von Korff M, Dunn KM. Chronic pain reconsidered. Pain. 2008;138(2):267-76.
10. Dunn KM, Croft PR, Main CJ, Von Korff M. A prognostic approach to defining chronic pain: replication in a UK primary care low back pain population. Pain. 2008;135(1-2):48-54.
11. Thomas E, Dunn KM, Mallen C, Peat G. A prognostic approach to defining chronic pain: application to knee pain in older adults. Pain. 2008;139(2):389-97.
12. Muller S, Thomas E, Dunn KM, Mallen CD. A prognostic approach to defining chronic pain across a range of musculoskeletal pain sites. Clin J Pain. 2013;29(5):411-6.
13. Mallen CD, Peat G, Porcheret M, Croft P. The prognosis of joint pain in the older patient: general practitioners' views on discussing and estimating prognosis. Eur J Gen Pract. 2007;13(3):166-8.
14. Mallen CD, Peat G. Discussing prognosis with older people with musculoskeletal pain: a cross-sectional study in general practice. BMC Fam Pract. 2009;10:50.
15. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain. 1998;14(3):209-15.
16. Von Korff M, Shortreed SM, Saunders KW, LeResche L, Berlin JA, Stang P, et al. Comparison of back pain prognostic risk stratification item sets. J Pain. 2014;15(1):81-9.
17. Von Korff M. Tailoring chronic pain care by brief assessment of impact and prognosis: comment on "Point-of-care prognosis for common musculoskeletal pain in older adults". JAMA Intern Med. 2013;173(12):1126-7.
18. Salisbury C, Foster NE, Hopper C, Bishop A, Hollinghurst S, Coast J, et al. A pragmatic randomised controlled trial of the effectiveness and cost-effectiveness of

'PhysioDirect' telephone assessment and advice services for physiotherapy. Health Technol Assess. 2013;17(2):1-157, v-vi.
19. Foster NE, Williams B, Grove S, Gamlin J, Salisbury C. The evidence for and against 'PhysioDirect' telephone assessment and advice services. Physiotherapy. 2011;97(1):78-82.
20. Salisbury C, Foster NE, Bishop A, Calnan M, Coast J, Hall J, et al. 'PhysioDirect' telephone assessment and advice services for physiotherapy: protocol for a pragmatic randomised controlled trial. BMC Health Serv Res. 2009;9:136.
21. Roddy E, Zwierska I, Dawes P, Hider SL, Jordan KP, Packham J, et al. The Staffordshire arthritis, musculoskeletal, and back assessment (SAMBA) study: a prospective observational study of patient outcome following referral to a primary-secondary care musculoskeletal interface service. BMC Musculoskelet Disord. 2010;11:67.
22. Roddy E, Zwierska I, Jordan KP, Dawes P, Hider SL, Packham J, et al. Musculoskeletal clinical assessment and treatment services at the primary-secondary care interface: an observational study. Br J Gen Pract. 2013;63(607):e141-8.
23. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. Med Care. 2002;40(5):442-6.
24. Ware JE, Jr. SF-36 health survey update. Spine (Phila Pa 1976). 2000;25(24):3130-9.
25. Corica F, Corsonello A, Apolone G, Mannucci E, Lucchetti M, Bonfiglio C, et al. Metabolic syndrome, psychological status and quality of life in obesity: the QUOVADIS Study. Int J Obes (Lond). 2008;32(1):185-91.
26. Dorynska A, Pajak A, Kubinova R, Malyutina S, Tamosiunas A, Pikhart H, et al. Socioeconomic circumstances, health behaviours and functional limitations in older persons in four Central and Eastern European populations. Age Ageing. 2012;41(6):728-35.
27. Wittink H, Turk DC, Carr DB, Sukiennik A, Rogers W. Comparison of the redundancy, reliability, and responsiveness to change among SF-36, Oswestry Disability Index, and Multidimensional Pain Inventory. Clin J Pain. 2004;20(3):133-42.
28. Angst F, Verra ML, Lehmann S, Gysi F, Benz T, Aeschlimann A. Responsiveness of the cervical Northern American Spine Society questionnaire (NASS) and the Short Form 36 (SF-36) in chronic whiplash. Clin Rehabil. 2012;26(2):142-51.
29. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB, Kinemax Outcomes G. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. J Bone Joint Surg Am. 2001;83-A(12):1856-64.
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
31. Yin J, Samawi H, Linder D. Improved nonparametric estimation of the optimal diagnostic cut-off point associated with the Youden index under different sampling schemes. Biom J. 2016.
32. Greenhouse SW, Cornfield J, Homburger F. The Youden index: letters to the editor. Cancer. 1950;3(6):1097-101.
33. Morso L, Kongsted A, Hestbaek L, Kent P. The prognostic ability of the STarT Back Tool was affected by episode duration. Eur Spine J. 2016;25(3):936-44.
34. Butera KA, Lentz TA, Beneciuk JM, George SZ. Preliminary Evaluation of a Modified STarT Back Screening Tool Across Different Musculoskeletal Pain Conditions. Phys Ther. 2016.

Figure 1. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the PhysioDirect dataset.

Figure 2. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the SAMBA dataset.

Table 1. Identifying optimal modified STarT Back Tool cut-points for each pain region using a) Youden's J statistic and b) a clinically defined maximum specificity of 0.7.

		PhysioDirect trial data			*SAMBA study data*		
Pain region	**Modified STarT Back Tool cut-point**	**Sens**	**Spec**	**Youden's**	***Sens***	***Spec***	***Youden's***
Neck	3	0.967	0.451	0.418	*1*	*0.556*	*0.556*
Neck	4	0.833	0.668	0.501	*1*	*0.644*	*0.644*
Neck	5	0.767	0.777*	0.544	*0.769*	*0.756**	*0.525*
Neck	6	0.467	0.87*	0.337	*0.615*	*0.867**	*0.482*
Neck	7	0.3	0.959*	0.259	*0.538*	*0.911**	*0.449*
Neck	8	0.067	0.99*	0.057	*0.462*	*0.956**	*0.418*
Back	3	0.903	0.329	0.232	*0.987*	*0.333*	*0.32*
Back	4	0.832	0.491	0.323	*0.935*	*0.454*	*0.389*
Back	5	0.708	0.652	0.36	*0.857*	*0.546*	*0.403*
Back	6	0.442	0.818*	0.26	*0.792*	*0.686*	*0.478*
Back	7	0.23	0.921*	0.151	*0.558*	*0.821**	*0.379*
Back	8	0.088	0.979*	0.067	*0.273*	*0.913**	*0.186*
Upper limb	3	0.882	0.538	0.42	*0.907*	*0.576*	*0.483*
Upper limb	4	0.711	0.72*	0.431	*0.86*	*0.681*	*0.541*
Upper limb	5	0.513	0.853*	0.366	*0.791*	*0.79**	*0.581*
Upper limb	6	0.303	0.929*	0.232	*0.674*	*0.848**	*0.522*
Upper limb	7	0.158	0.973*	0.131	*0.581*	*0.933**	*0.514*
Upper limb	8	0.039	0.989*	0.028	*0.163*	*0.976**	*0.139*
Lower limb	3	0.9	0.47	0.37	*0.904*	*0.505*	*0.409*
Lower limb	4	0.77	0.618	0.388	*0.851*	*0.664*	*0.515*
Lower limb	5	0.57	0.789*	0.359	*0.754*	*0.771**	*0.525*
Lower limb	6	0.36	0.895*	0.255	*0.64*	*0.86**	*0.5*
Lower limb	7	0.21	0.971*	0.181	*0.43*	*0.93**	*0.36*
Lower limb	8	0.05	0.996*	0.046	*0.237*	*0.972**	*0.209*
Multi-site pain	3	0.933	0.443	0.376	*0.946*	*0.434*	*0.38*
Multi-site pain	4	0.833	0.656	0.489	*0.946*	*0.566*	*0.512*
Multi-site pain	5	0.733	0.787*	0.52	*0.911*	*0.663*	*0.574*
Multi-site pain	6	0.567	0.902*	0.469	*0.857*	*0.711**	*0.568*
Multi-site pain	7	0.333	0.934*	0.267	*0.643*	*0.771**	*0.414*
Multi-site pain	8	0.1	0.984*	0.084	*0.357*	*0.94**	*0.297*

Grey shaded row indicates Youden's optimal score cut-point for predicting 6-month outcome

* indicates specificity was >0.7 according to pre-defined clinical criteria

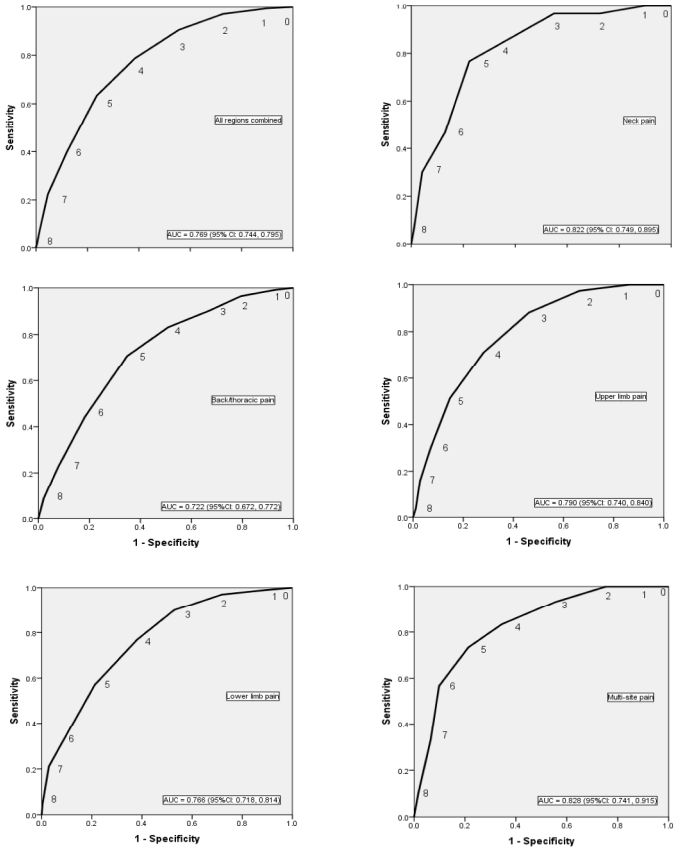

Figure 1. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the PhysioDirect dataset.
presented in Figures 1 & 2
210x297mm (300 x 300 DPI)

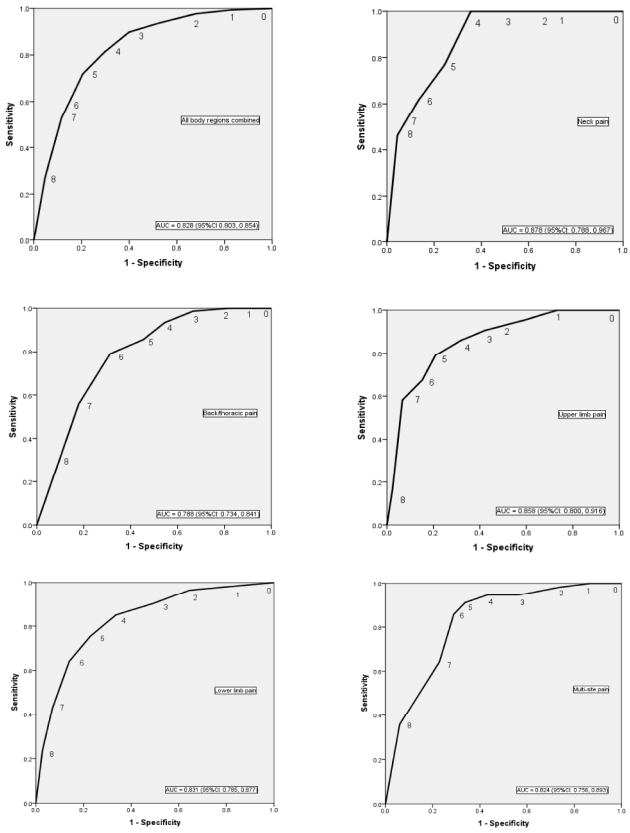

Figure 2. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the SAMBA dataset.
presented in Figures 1 & 2
210x297mm (300 x 300 DPI)

BMJ Open

Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012445.R2
Article Type:	Research
Date Submitted by the Author:	09-Sep-2016
Complete List of Authors:	Hill, Jonathan; Keele University, Institute of primary care and health sciences Afolabi, Ebenezer ; Keele University, Institute of Primary Care and Health Sciences Lewis, Martyn; Keele University, Institute of Primary Care and Health Sciences Dunn, Kate; Keele University, Institute of Primary Care and Health Sciences Roddy, Edward; Keele University, Institute for Primary Care and Health Sciences van der Windt, Danielle; Keele University, Institute of Primary Care and Health Sciences Foster, Nadine; Keele University, Institute of Primary Care Health Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	General practice / Family practice, Rheumatology, Sports and exercise medicine
Keywords:	Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, prognosis, prospective cohort, STarT Back Tool, Risk stratification

Title Page 9th September 2016

Title: Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data.

Corresponding Author:

Dr Jonathan C Hill[1]

Senior Lecturer

Institute of primary care and health sciences

Keele University, Staffs

ST5 5BG

UK

01782733900

j.hill@keele.ac.uk

Co-authors: Hill JC, Afolabi EK, Lewis M, Dunn KM, Roddy E, van der Windt DA, Foster NE.

Affiliation:

1. Institute of Primary Care and Health Sciences, Keele University, Staffordshire, ST5 5BG, UK

Key words: Musculoskeletal Pain, Questionnaires, Prognosis, Prospective studies, STarT Back Tool

Manuscript Word count =2900

ABSTRACT: (Word count = 300)

Objectives: The STarT Back Tool has good predictive performance for non-specific low back pain in primary care. We therefore aimed to investigate whether a modified STarT Back Tool predicted outcome with a broader group of musculoskeletal patients, and assessed the consequences of using existing risk-group cut-points across different pain regions.

Setting: Secondary analysis of prospective data from 2 cohorts: 1) outpatient musculoskeletal physiotherapy services (PhysioDirect trial n=1887), and 2) musculoskeletal primary-secondary care interface services (SAMBA study n=1082)

Participants: Patients with back, neck, upper-limb, lower-limb, or multi-site pain with a completed modified STarT Back Tool (baseline) and six-month physical health outcome (SF-36).

Outcomes: Area-Under-the-Receiving-Operator-Curve (AUCs) tested discriminative abilities of the tool's baseline score for identifying poor six-month outcome (SF-36 lower-tertile Physical-Component-Score). Risk-group cut-points were tested using sensitivity and specificity for identifying poor outcome using a) Youden's J statistic, and b) a clinically determined rule that specificity should not fall below 0.7 (false-positive rate less than 30%).

Results: In PhysioDirect and SAMBA poor six-month physical health was 18.5%, 28.2% respectively. Modified STarT Back Tool score AUCs for predicting outcome in back pain were 0.72, 0.79 in PhysioDirect and SAMBA respectively, neck 0.82, 0.88, upper limb 0.79, 0.86, lower limb 0.77, 0.83 and multi-site pain 0.83, 0.82. Differences between pain region AUCs were non-significant. Optimal cut-points to discriminate low and medium/high risk-groups depended on pain region and clinical services.

Conclusion: A modified STarT Back Tool similarly predicts six-month physical health outcome across five musculoskeletal pain regions. However, use of consistent risk-group cut-points was not possible and resulted in poor sensitivity (too many with long-term disability being missed) or specificity (too many with good outcome inaccurately classified as 'at risk') for some pain regions. The draft tool is now being refined and validated within a new programme of research for a broader musculoskeletal population.

Strengths and limitations of this study

- First study to demonstrate that modified STarT Back Tool items are similarly predictive of six-month physical health across different musculoskeletal pain regions.
- A limitation of the study was that the original STarT Back Tool was not included in these two datasets, so a direct comparison between the performance of the original and modified STarT Back Tool versions for patients with low back pain was not possible.

INTRODUCTION

The Keele STarT Back Tool is designed to stratify patients with low back pain according to their risk of future physical disability, in order that prognostic subgroups can receive matched treatment.(1) For example, individuals at a low risk of persistent disabling problems can be reassured and discouraged from receiving unnecessary treatments and investigations, whilst those at high risk can matched to treatment which combines physical and psychological approaches.(2-4) A large randomised trial testing a risk stratification approach (use of the STarT Back Tool and matched treatments) for low back pain in comparison to best current care demonstrated superior clinical and cost outcomes.(5) In addition, an implementation study testing risk stratification for patients with low back pain in routine general practice demonstrated significant improvements in physical function and time off work, sickness certification rates, and reductions in healthcare costs compared to usual non-stratified care.(2) Since low back pain accounts for only 17% of all UK primary care musculoskeletal consultations in general practice,(6) if a similar screening tool could be used for patients with other common pain presentations, such as neck pain and knee pain, then there could be potential for stratified care to make a greater impact for patients and healthcare services.

A previous systematic review of 45 cohort studies(7) reported that prognostic factors are often similar across different musculoskeletal presentations, with 11 factors predicting poor outcome at follow-up for at least two different musculoskeletal pain problems. Other studies have similarly shown that a generic set of baseline factors (pain intensity, episode duration, pain interference, depression and co-morbid pain problems) predicts risk of a poor outcome across different pain regions including back pain, headache, facial pain and knee pain, regardless of the specific location of pain or underlying pathology.(8-12) These studies indicate that it might be possible to utilise the same prognostic factors as those included within the STarT Back Tool to discriminate risk status for a much larger group of musculoskeletal pain patients than those consulting with low back pain. The key benefit of using a single tool to stratify patients with a wide range of musculoskeletal conditions rather than multiple site-specific prognostic screening tools is its simplicity for use in busy clinical practice.

Whilst the likely value and acceptability of extending risk stratification to patients with other common musculoskeletal pain is as yet unknown, evidence suggests that the majority of General Practitioners (GPs) consider prognosis to be important in their clinical decision-making for musculoskeletal treatment.(13) Despite the widespread support for prognostic information, the clinical reality is that predicting outcome in these patients is not always easy and patient's risk status is not typically included within medical records.(14) GPs are not alone in wanting information about patients' likely prognosis over time, as more than 80% of musculoskeletal patients also want prognostic information from their GP, although less than a third actually receive this information.(14) Existing musculoskeletal prognostic tools are available (e.g. Linton and Hallden,(15) and von Korff et al,(16, 17)). However, these prognostic tools were not designed or tested to support clinical decisions in primary care about matched treatments (stratified care); only the STarT Back Tool has been specifically developed and tested to guide patient treatment matching.

The aim of this study was therefore to investigate the performance of a modified STarT Back Tool for predicting future physical health outcome for a broader group of musculoskeletal pain patients. Specific objectives were to compare the predictive performance of a modified STarT Back Tool for patients with musculoskeletal pain in different body regions and assess the consequences (false positive and false negative rates) of using existing STarT Back Tool score cut-points for classifying patients as medium/high risk across different pain regions (neck, back, upper limb, lower limb, and multi-site pain).

METHODS

Design:

This study involved pre-specified further analysis of existing datasets from two prospective cohorts of adults with musculoskeletal conditions consulting in two different services in the NHS, UK. Full ethical approval for both these studies was obtained and patients provided written informed consent prior to their research participation.

Patient population:

1) The PhysioDirect trial included 2249 adult musculoskeletal patients taking part in a randomised trial comparing a PhysioDirect service (telephone-based physiotherapy assessment and advice) with usual physiotherapy care.(18-20) Primary outcome data (physical health measured using the SF-36v2 physical component score) at six-month follow-up and baseline modified STarT Back Tool score were available for 1887 patients (84%) and were included in this analysis. The trial was conducted in four NHS community physiotherapy services in four different areas of England (Bristol, Somerset, Stoke-on-Trent, and Cheshire). Adults (aged ≥18 years) who were referred by 94 general practitioners (covering a wide range of geographical areas and populations), or who referred themselves for physiotherapy for a musculoskeletal problem, were eligible for the trial. Patients completed postal questionnaires at baseline and six-months after randomisation. Details about the PhysioDirect patient sample have been published.(18) For the study reported here we used patients from both the control and intervention arms.

2) The SAMBA study was an observational cohort of adults attending an NHS musculoskeletal clinical assessment and treatment service at the primary-secondary care interface.(21, 22) The study population included 2166 patients referred from primary care and subsequently triaged to musculoskeletal and back pain interface clinics in Stoke-on-Trent Primary Care Trust (PCT) over a 12-month period. Primary outcome data at six-month follow-up (physical health measured using the SF-36v2 physical component score) and the modified STarT Back Tool score at baseline was available for 1082 patients (50%) who formed the study population for this evaluation. All adults (aged ≥18 years) capable of giving written informed consent were eligible to participate in the study. Patients completed study questionnaires before their first appointment during which consent was obtained and six-months after that initial clinic appointment. Details of the SAMBA study sample have been published.(22)

Modifying the STarT Back Tool:

The original STarT Back Tool includes nine items of which five concern psychosocial factors (fear, catastrophising, anxiety, depression, and bothersomeness). Both the PhysioDirect trial and SAMBA study included the STarT Back Tool's psychosocial items within their baseline questionnaires.(1) These items were used without modification as they were developed from generic tools and are not specific to low back pain. However, the four further items of the original STarT Back Tool that capture three physical factors (referred pain from the back down the leg, co-morbid pain in the neck and shoulder, and physical function with walking and dressing items) are specific to low back pain and therefore these items in their original form needed to be replaced by similar items that were applicable for all musculoskeletal patients. We therefore used proxy items for these outcome domains that were available in both datasets. The STarT Back Tool's two 'function' items (walking and dressing) were replaced by items from the generic EQ-5D(23) ('I have some problems in walking about', Y/N and 'I have some problems washing or dressing myself", Y/N), and we used item 7 from the SF-12(24) ('How much bodily pain have you had?' with positive responses defined as 'extremely' or 'very severe') instead of the original STarT Back Tool item for co-morbid pain in the neck or shoulder. It was not possible to replace 'referred pain from the back down the leg' with an item that was suitable for all musculoskeletal pain and so this construct of the 'spread of pain' was omitted from the modified tool. To score the modified STarT Back Tool, responses from these 8 items were summed (range 0-8) for all patients in both datasets. The original STarT Back Tool cut-off of 0-3 positive items was used to classify patients as at low risk and 4 or more as medium or high risk. There were no reference standards for psychological distress in either the PhysioDirect or SAMBA datasets and so in this analysis we did not seek to examine the ability of the modified STarT Back Tool to identify a high risk only group. We believe that there is a strong clinical rationale for identifying musculoskeletal cases that are 'at risk' of a poor prognosis, which reflects the combined medium and high risk subgroup. In our previous IMPaCT Back study(2) implementing risk stratification in general practice, the clinicians used a 6-item STarT Back Tool which only discriminated between low risk and a combined medium/high risk group to decide which patients to refer or not to physiotherapy. In that study the physiotherapists who received 'at risk' patients then used the full 9-item STarT Back Tool to discriminate the distressed patients that needed a psychologically informed physiotherapy treatment approach.

Defining the body regions of pain:

Participants were asked to indicate the primary site of their musculoskeletal pain for which they had sought treatment. From this information patients were categorized as having one of the following regional pain problems: neck, back (thoracic or lumbar), upper limb, lower limb, or multi-site pain (pain in more than one region).

Defining physical health outcome:

The standardised summary score for the Physical-Component-Score (PCS) of the Short Form 36 Health Survey (SF-36) is population normalised (0 is worst physical health and 100 is best physical health) and was classified by tertiles (<=33, 34-66, >66) as has been used previously (25, 26) with a

six-month poor outcome defined using the most severe tertile (<=33). Outcome was defined as poor physical health at six-month follow-up using the SF-36 PCS because this was the most appropriate physical function outcome score available in both studies, and it has demonstrated good validity and responsiveness in this population.(27-29)

Statistical analysis:

All analyses were conducted separately for the two datasets and a descriptive comparison of the modified baseline STarT Back Tool scores (mean and standard deviation [SD]) and proportion with poor six-month physical health outcome (SF-36 PCS <=33) calculated. Descriptive statistics using means and standard deviations were used to examine the modified STarT Back Tool score's distribution and investigate potential floor or ceiling effects (>10% of either lowest or maximum score).(30)

Predictive performance (discrimination) was assessed by calculating ROC curve AUCs for baseline modified STarT Back Tool total scores against six-month poor physical health outcome (dichotomised as poor/good) for each of the five different bodily pain presentations and their equality compared using STATA's 'roccomp' command to establish if AUC differences were statistically significant.

To examine whether the optimal subgroup cut-point on the modified STarT Back Tool total score to discriminate low from medium/high risk for poor six-month physical health outcome was consistent across the five different pain regions and across the two datasets, we used two methods based on sensitivity and specificity of each potential cut-point. Firstly, we used Youden's J Statistic which is calculated as sensitivity + specificity -1 for each potential cut-point and the optimal cut-point is the tool score with the highest value.(31, 32) Secondly, we *a priori* agreed that specificity should not fall below 0.7, as lower values would mean potentially over-treating more than 30% of medium/high risk patients, which was considered an unacceptable level for an efficient matched treatment approach.

In this study we were not able to identify optimal subgroup cut-points on the modified STarT Back Tool to distinguish between medium and high risk patients as there were no reference standards for psychological distress in the two available datasets. The original STarT Back Tool used these reference standards to identify distress 'caseness' at baseline, and identified the optimal cut-point to screen for these distressed 'cases' using a psychological subscale score. Without these reference standards for psychological distress, we were limited to determining optimal subgroup cut-points on the total scale score between low and medium/high risk alone.

RESULTS

Distribution of the modified STarT Back Tool scores in both datasets:

In the PhysioDirect trial sample (n=1887) the 8-item modified STarT Back Tool score at baseline was normally distributed with a mean (SD) of 3.35 (2.09); 8.4% had the lowest score (0) and 2.2% had the maximum score (8). The distribution of primary pain regions was reported by clinicians as: lower

limb 31.1%, back 28.7%, upper limb 23.5%, neck 11.8%, and multi-site pain 4.8%. The six-month SF-36 PCS mean (SD) was 43.7 (10.9) with 18.5% having a 'poor outcome' in their physical health at six-month follow-up. The mean age was 48 years old and 60% of the sample were female.

In the SAMBA study sample (n=1082) the 8-item modified STarT Back Tool score at baseline was not normally distributed but had roughly equal numbers of all possible scores with a mean (SD) of 3.95 (2.65); 12.6% had the lowest score (0) and 10.9% had the maximum score (8). The distribution of primary pain sites was reported by patients as: lower limb 30.8%, back 26.7%, upper limb 23.8%, multi-site pain 13.4% and neck 5.4%. The six-month SF-36 PCS mean (SD) was 38.41 (12.76) with 28.2% having a 'poor outcome' in their physical health at six-month follow-up. The mean age was 51 years old and 57% were female.

Predictive performance of the modified STarT Back Tool score across pain regions in both datasets:

Predictive performance of the modified STarT Back Tool as determined by ROC curve AUCs ranged from 0.72 to 0.83 and was not found to be statistically different across different pain regions in the PhysioDirect trial (p= 0.098) and SAMBA study (p=0.130) (presented in Figures 1 & 2).

Optimal modified STarT Back Tool score cut-offs in both datasets:

Table 1 reports sensitivity, specificity, and the Youden's J statistic for each possible modified STarT Back Tool score cut-point at baseline for each pain region. The results demonstrate that the optimal STarT Back Tool baseline score cut-point for discriminating 'poor outcome' at six-month follow-up was not consistent across pain regions. For example, among (PhysioDirect) patients with neck, back and multi-site pain the optimal STarT Back Tool cut-point for discriminating 'poor outcome' was 5, whereas this was 4 for those with upper limb and lower limb as their primary pain site.

Discussion

This is the first study to demonstrate that a modified STarT Back Tool is similarly predictive of six-month physical health (defined by worst tertile of the SF-36) across different musculoskeletal pain regions. Predictive performance determined by AUCs for the 8-item modified STarT Back Tool total score was in fact slightly higher for neck, upper limb, lower limb and multi-site pain than for back pain, although differences were not statistically significant. The results therefore demonstrate that the prognostic factors included within the STarT Back Tool are predictive of six-month physical health across a range of musculoskeletal pain regions, not just back pain. However, the results demonstrated that the optimal baseline STarT Back Tool score cut-point for identifying individuals with poor physical health outcome was neither consistent across different pain regions, nor across clinical services (community physiotherapy services [PhysioDirect trial] and primary-secondary care interface services [SAMBA study]. This finding was consistent regardless of method used to determine the optimal modified STarT Back Tool score cut-point (Youden's J statistic or an a priori defined maximum false positive rate of 30%). This implies that the existing original STarT Back Tool score cut-point (4 or more out of 9) used to allocate patients with low back pain to the medium/high risk subgroups cannot simply be applied to patients with other musculoskeletal pain

presentations or in different clinical services. This is likely to be due to differences in patient characteristics across services such as episode duration, which is known to influence the performance of the original STarT Back Tool.(33) It is also likely that individual modified STarT Back Tool items are not equally applicable to patients with pain in the five regions.(34) For example, the item about walking difficulties is likely to be less relevant and therefore less predictive of physical health outcome for patients with upper limb pain than for those with lower limb or spinal pain. A key message from this study is the value and importance of testing the capabilities of the STarT Back Tool in different settings and patient populations and not presuming that existing primary care subgroup cut-points will be the same in other groups. If wider validity is demonstrated this will help strengthen the case for the general applicability of the tool.

The findings of this study concur with previous evidence suggesting that the same set of prognostic variables can be used to estimate prognosis of patients with different musculoskeletal pain presentations.(7, 15, 17) The STarT Back Tool uses biopsychosocial constructs known to predict persistent disability among patients with low back pain, such as: difficulty with walking and dressing, pain elsewhere, fear avoidance, pain catastrophising, anxiety and low mood.(1) However, the STarT Back Tool is not just a prognostic index, but is used to stratify patients for different matched treatments. An important issue highlighted by this analysis is that if clinicians simply modify the STarT Back Tool for use with other musculoskeletal pain patients, they are at risk of matching patients to inappropriate treatments. It is also apparent that future translation and validation studies of the STarT Back Tool need to carefully consider adopting the same STarT Back Tool score cut-points as used in the original UK STarT Back Tool study(1) without first testing if these cut-points are appropriate for their own clinical populations. Based on these findings our team has begun to further refine and validate an improved stratification tool – the Keele STarT MSK Tool - which will be specifically designed for use with primary care patients consulting with the five most common musculoskeletal pain presentations in a new programme of research. Whilst our study was not able to examine optimal high risk subgroup cut-offs for 'distressed' patients, a previous cross-sectional study [34] in a US physical therapy population has compared the relationships between a modified STarT Back Tool and psychological measures in people with different pain regions. It found that regardless of pain body region higher modified STarT Back Tool scores were associated with higher levels of kinesiophobia, catastrophising, fear avoidance, anxiety, and depressive symptoms. The strengths of our analyses reported here include the large sample sizes of both the PhysioDirect and SAMBA studies and the opportunity to examine optimal cut-points in patients with different pain sites and in different NHS musculoskeletal services. An additional strength was that both studies used the same measure of physical health (SF-36), had the same six-month follow-up time-point and included patients whose pain could be classified into the same musculoskeletal pain regions. Given the potential weakness of using the Youden's J Statistic to define optimal cut-points for discriminating between low and medium/high risk, we also used a clinically determined guide (maximum false positive rate), which showed similar inconsistencies in optimal cut-off between regional pain site and clinical setting. One weakness is that the original STarT Back Tool was not included in these two datasets, which meant a direct comparison between the performance of the original and modified versions for patients with low back pain was not possible. The choice of poor

physical health outcome at six-months using the lowest tertile on the SF-12 was also relatively arbitrary, but served the purpose of this analysis to compare outcome between different regional pain sites, making the exact definition of poor outcome less critical to the study aims. It should be noted that the different levels of poor clinical outcome between the PhysioDirect (18.5%) and Samba (28.2%) studies could be due to the different settings and design of these two studies and it is possible this may have influenced the findings.

The implications from this analysis are that, despite good predictive performance of the modified STarT Back Tool in patients with pain in different regions of the body, clinicians need to cautiously consider the choice of cut-points when using a modified STarT Back Tool for musculoskeletal pain regions other than low back pain. The results suggest existing cut-points may lead to an inefficiency in healthcare resource use, with too many patients with a likely long-term disability being missed, or too many patients with good physical health outcome being inaccurately classified as 'at risk', which may result in over treatment of low risk groups.

Conclusions

A modified version of the STarT Back Tool has similar predictive performance when used for patients with musculoskeletal pain in different body regions. However, the cut-points used to identify patients with a poor physical health outcome at six-month follow-up are not consistent across pain regions or clinical services. Further research is underway to refine and validate a new Keele STarT MSK Tool which will form part of a new stratified care approach to be tested in a randomised controlled trial.

Author Contribution: JCH, DvdW, NEF, ER and KD conceived and designed the research; NF and ER were responsible for the modified STarT Back Tool being embedded within the PhysioDirect and SAMBA datasets; JCH, EA and ML analysed the data and all authors were involved in the interpretation of the data analysis; JCH, ER, EA, ML, DvdW, KD and NF were involved in the drafting of the manuscript and its revision for important intellectual content, and gave final approval for the manuscript submission.

Funding: This paper presents independent research (part) funded by the National Institute for Health Research (NIHR) Programme Grants for Applied Research programme (Grant reference number: RP-PG-1211-20010), the NIHR Primary Care Career Scientist Award to Professor Nadine Foster [CSA 04/03], and Arthritis Research UK [13413]. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Professor Foster holds an NIHR Research Professorship (NIHR-RP-2011-015) which also supports Jonathan Hill. The views and opinions expressed within this manuscript do not necessarily reflect those of DH/NIHR. The funding bodies were not involved in the design of the study outlined within this protocol, and had no involvement in the writing and revision of this manuscript. DvdW received

funding from an MRC Partnership Grant for the PROGnosis RESearch Strategy (PROGRESS) group (grant reference number: G0902393). The SAMBA study was supported by an Arthritis Research UK Integrated Clinical Arthritis Centre Grant (17684), the Arthritis Research UK Primary Care Centre Grant (18139), funding secured from Stoke-on-Trent Primary Care Trust (PCT), and service support through the West Midlands North CLRN.

Competing interests: None declared.

Patient consent Obtained: Yes, full written informed consent for all participants.

Ethics approval: This was secondary data analysis of two studies which both obtained ethical approval through written informed consent. For PhysioDirect approval came from Southmead NHS Research Ethics Committee, reference 08/H0102/95 and for the SAMBA study from the South Staffordshire NHS Research Ethics Committee reference 07/H1203/86.

Data sharing statement: Data can be accessed via the Keele data repository at http:// http://www.keele.ac.uk/pchs/publications/datasharingresources/

Acknowledgements:

We thank the patients and clinical teams that participated in the two studies.

References

1. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum. 2008;59(5):632-41.
2. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Ann Fam Med. 2014;12(2):102-11.
3. Main CJ, Sowden G, Hill JC, Watson PJ, Hay EM. Integrating physical and psychological approaches to treatment in low back pain: the development and content of the STarT Back trial's 'high-risk' intervention (StarT Back; ISRCTN 37113406). Physiotherapy. 2012;98(2):110-6.
4. Sowden G, Hill JC, Konstantinou K, Khanna M, Main CJ, Salmon P, et al. Targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study (ISRCTN 55174281). Fam Pract. 2012;29(1):50-62.
5. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet. 2011;378(9802):1560-71.
6. Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. BMC Musculoskelet Disord. 2010;11:144.
7. Mallen CD, Peat G, Thomas E, Dunn KM, Croft PR. Prognostic factors for musculoskeletal pain in primary care: a systematic review. Br J Gen Pract. 2007;57(541):655-61.
8. Von Korff M, Miglioretti DL. A prognostic approach to defining chronic pain. Pain. 2005;117(3):304-13.
9. Von Korff M, Dunn KM. Chronic pain reconsidered. Pain. 2008;138(2):267-76.
10. Dunn KM, Croft PR, Main CJ, Von Korff M. A prognostic approach to defining chronic pain: replication in a UK primary care low back pain population. Pain. 2008;135(1-2):48-54.
11. Thomas E, Dunn KM, Mallen C, Peat G. A prognostic approach to defining chronic pain: application to knee pain in older adults. Pain. 2008;139(2):389-97.
12. Muller S, Thomas E, Dunn KM, Mallen CD. A prognostic approach to defining chronic pain across a range of musculoskeletal pain sites. Clin J Pain. 2013;29(5):411-6.
13. Mallen CD, Peat G, Porcheret M, Croft P. The prognosis of joint pain in the older patient: general practitioners' views on discussing and estimating prognosis. Eur J Gen Pract. 2007;13(3):166-8.
14. Mallen CD, Peat G. Discussing prognosis with older people with musculoskeletal pain: a cross-sectional study in general practice. BMC Fam Pract. 2009;10:50.
15. Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain. 1998;14(3):209-15.
16. Von Korff M, Shortreed SM, Saunders KW, LeResche L, Berlin JA, Stang P, et al. Comparison of back pain prognostic risk stratification item sets. J Pain. 2014;15(1):81-9.
17. Von Korff M. Tailoring chronic pain care by brief assessment of impact and prognosis: comment on "Point-of-care prognosis for common musculoskeletal pain in older adults". JAMA Intern Med. 2013;173(12):1126-7.
18. Salisbury C, Foster NE, Hopper C, Bishop A, Hollinghurst S, Coast J, et al. A pragmatic randomised controlled trial of the effectiveness and cost-effectiveness of

'PhysioDirect' telephone assessment and advice services for physiotherapy. Health Technol Assess. 2013;17(2):1-157, v-vi.
19. Foster NE, Williams B, Grove S, Gamlin J, Salisbury C. The evidence for and against 'PhysioDirect' telephone assessment and advice services. Physiotherapy. 2011;97(1):78-82.
20. Salisbury C, Foster NE, Bishop A, Calnan M, Coast J, Hall J, et al. 'PhysioDirect' telephone assessment and advice services for physiotherapy: protocol for a pragmatic randomised controlled trial. BMC Health Serv Res. 2009;9:136.
21. Roddy E, Zwierska I, Dawes P, Hider SL, Jordan KP, Packham J, et al. The Staffordshire arthritis, musculoskeletal, and back assessment (SAMBA) study: a prospective observational study of patient outcome following referral to a primary-secondary care musculoskeletal interface service. BMC Musculoskelet Disord. 2010;11:67.
22. Roddy E, Zwierska I, Jordan KP, Dawes P, Hider SL, Packham J, et al. Musculoskeletal clinical assessment and treatment services at the primary-secondary care interface: an observational study. Br J Gen Pract. 2013;63(607):e141-8.
23. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. Med Care. 2002;40(5):442-6.
24. Ware JE, Jr. SF-36 health survey update. Spine (Phila Pa 1976). 2000;25(24):3130-9.
25. Corica F, Corsonello A, Apolone G, Mannucci E, Lucchetti M, Bonfiglio C, et al. Metabolic syndrome, psychological status and quality of life in obesity: the QUOVADIS Study. Int J Obes (Lond). 2008;32(1):185-91.
26. Dorynska A, Pajak A, Kubinova R, Malyutina S, Tamosiunas A, Pikhart H, et al. Socioeconomic circumstances, health behaviours and functional limitations in older persons in four Central and Eastern European populations. Age Ageing. 2012;41(6):728-35.
27. Wittink H, Turk DC, Carr DB, Sukiennik A, Rogers W. Comparison of the redundancy, reliability, and responsiveness to change among SF-36, Oswestry Disability Index, and Multidimensional Pain Inventory. Clin J Pain. 2004;20(3):133-42.
28. Angst F, Verra ML, Lehmann S, Gysi F, Benz T, Aeschlimann A. Responsiveness of the cervical Northern American Spine Society questionnaire (NASS) and the Short Form 36 (SF-36) in chronic whiplash. Clin Rehabil. 2012;26(2):142-51.
29. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB, Kinemax Outcomes G. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. J Bone Joint Surg Am. 2001;83-A(12):1856-64.
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
31. Yin J, Samawi H, Linder D. Improved nonparametric estimation of the optimal diagnostic cut-off point associated with the Youden index under different sampling schemes. Biom J. 2016.
32. Greenhouse SW, Cornfield J, Homburger F. The Youden index: letters to the editor. Cancer. 1950;3(6):1097-101.
33. Morso L, Kongsted A, Hestbaek L, Kent P. The prognostic ability of the STarT Back Tool was affected by episode duration. Eur Spine J. 2016;25(3):936-44.
34. Butera KA, Lentz TA, Beneciuk JM, George SZ. Preliminary Evaluation of a Modified STarT Back Screening Tool Across Different Musculoskeletal Pain Conditions. Phys Ther. 2016.

Figure 1. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the PhysioDirect dataset.

Figure 2. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the SAMBA dataset.

Table 1. Identifying optimal modified STarT Back Tool cut-points for each pain region using a) Youden's J statistic and b) a clinically defined maximum specificity of 0.7.

		PhysioDirect trial data			*SAMBA study data*		
Pain region	**Modified STarT Back Tool cut-point**	**Sens**	**Spec**	**Youden's**	***Sens***	***Spec***	***Youden's***
Neck	3	0.967	0.451	0.418	*1*	*0.556*	*0.556*
Neck	4	0.833	0.668	0.501	*1*	*0.644*	*0.644*
Neck	5	0.767	0.777*	0.544	*0.769*	*0.756**	*0.525*
Neck	6	0.467	0.87*	0.337	*0.615*	*0.867**	*0.482*
Neck	7	0.3	0.959*	0.259	*0.538*	*0.911**	*0.449*
Neck	8	0.067	0.99*	0.057	*0.462*	*0.956**	*0.418*
Back	3	0.903	0.329	0.232	*0.987*	*0.333*	*0.32*
Back	4	0.832	0.491	0.323	*0.935*	*0.454*	*0.389*
Back	5	0.708	0.652	0.36	*0.857*	*0.546*	*0.403*
Back	6	0.442	0.818*	0.26	*0.792*	*0.686*	*0.478*
Back	7	0.23	0.921*	0.151	*0.558*	*0.821**	*0.379*
Back	8	0.088	0.979*	0.067	*0.273*	*0.913**	*0.186*
Upper limb	3	0.882	0.538	0.42	*0.907*	*0.576*	*0.483*
Upper limb	4	0.711	0.72*	0.431	*0.86*	*0.681*	*0.541*
Upper limb	5	0.513	0.853*	0.366	*0.791*	*0.79**	*0.581*
Upper limb	6	0.303	0.929*	0.232	*0.674*	*0.848**	*0.522*
Upper limb	7	0.158	0.973*	0.131	*0.581*	*0.933**	*0.514*
Upper limb	8	0.039	0.989*	0.028	*0.163*	*0.976**	*0.139*
Lower limb	3	0.9	0.47	0.37	*0.904*	*0.505*	*0.409*
Lower limb	4	0.77	0.618	0.388	*0.851*	*0.664*	*0.515*
Lower limb	5	0.57	0.789*	0.359	*0.754*	*0.771**	*0.525*
Lower limb	6	0.36	0.895*	0.255	*0.64*	*0.86**	*0.5*
Lower limb	7	0.21	0.971*	0.181	*0.43*	*0.93**	*0.36*
Lower limb	8	0.05	0.996*	0.046	*0.237*	*0.972**	*0.209*
Multi-site pain	3	0.933	0.443	0.376	*0.946*	*0.434*	*0.38*
Multi-site pain	4	0.833	0.656	0.489	*0.946*	*0.566*	*0.512*
Multi-site pain	5	0.733	0.787*	0.52	*0.911*	*0.663*	*0.574*
Multi-site pain	6	0.567	0.902*	0.469	*0.857*	*0.711**	*0.568*
Multi-site pain	7	0.333	0.934*	0.267	*0.643*	*0.771**	*0.414*
Multi-site pain	8	0.1	0.984*	0.084	*0.357*	*0.94**	*0.297*

Grey shaded row indicates Youden's optimal score cut-point for predicting 6-month outcome

* indicates specificity was >0.7 according to pre-defined clinical criteria

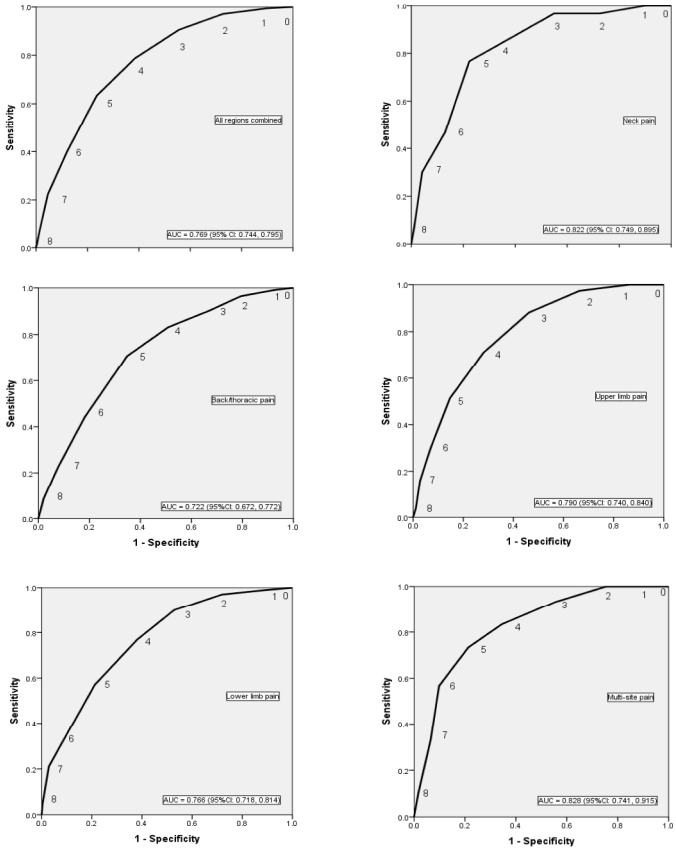

Figure 1. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the PhysioDirect dataset.
presented in Figures 1 & 2
210x297mm (300 x 300 DPI)

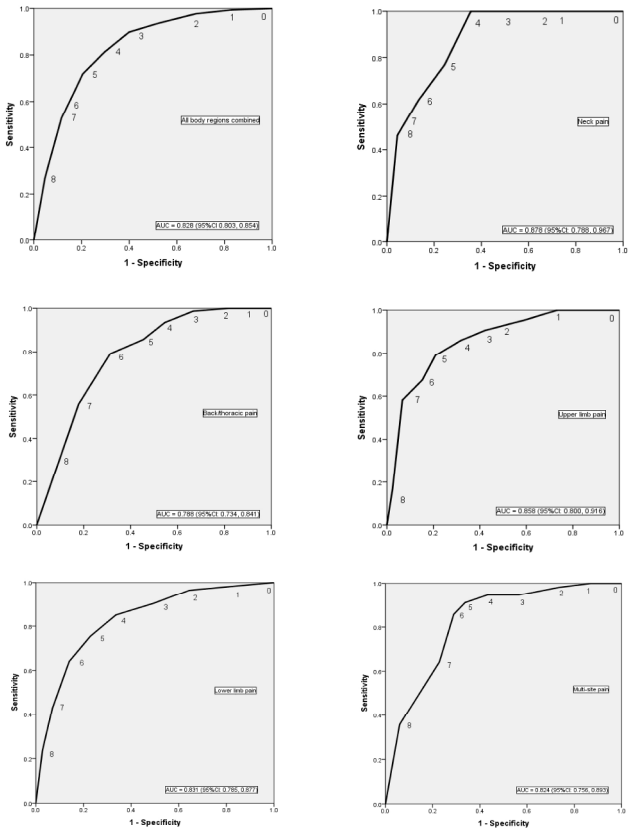

Figure 2. Receiver operating characteristic (ROC) curves for overall modified STarT Back tool scores against 6-month poor physical health outcome (SF-36 PCS <=33) by different pain regions in the SAMBA dataset.
presented in Figures 1 & 2
210x297mm (300 x 300 DPI)