



## Programme Grants for Applied Research

Volume 11 • Issue 4 • June 2023

ISSN 2050-4322

# Stratified primary care for adults with musculoskeletal pain: the STarT MSK research programme including RCTs

*Nadine E Foster, Kate M Dunn, Joanne Protheroe, Jonathan C Hill, Martyn Lewis, Benjamin Saunders, Sue Jowett, Susie Hennings, Paul Campbell, Kieran Bromley, Bernadette Bartlam, Opeyemi Babatunde, Simon Wathall, Raymond Oppong, Jesse Kigozi and Adrian Chudyk*





# Stratified primary care for adults with musculoskeletal pain: the STarT MSK research programme including RCTs

Nadine E Foster<sup>1,2</sup>, Kate M Dunn<sup>2</sup>, Joanne Protheroe<sup>2</sup>, Jonathan C Hill<sup>2\*</sup>, Martyn Lewis<sup>2,3</sup>, Benjamin Saunders<sup>2</sup>, Sue Jowett<sup>4</sup>, Susie Hennings<sup>3</sup>, Paul Campbell<sup>2,5</sup>, Kieran Bromley<sup>2,3</sup>, Bernadette Bartlam<sup>2,6,7</sup>, Opeyemi Babatunde<sup>2</sup>, Simon Wathall<sup>2,3</sup>, Raymond Oppong<sup>4</sup>, Jesse Kigozi<sup>4</sup> and Adrian Chudyk<sup>2</sup>

<sup>1</sup>STARS Education and Research Alliance, Surgical Treatment and Rehabilitation Service (STARS), The University of Queensland and Metro North Health, Brisbane, QLD, Australia

<sup>2</sup>Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, UK

<sup>3</sup>Keele Clinical Trials Unit, School of Medicine, Keele University, Keele, UK

<sup>4</sup>Health Economics Unit, Institute of Applied Health, University of Birmingham, Birmingham, UK

<sup>5</sup>Department of Research and Innovation, Midlands Partnership NHS Foundation Trust, St George's Hospital, Stafford, UK

<sup>6</sup>Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

<sup>7</sup>Community and Primary Care Research Group, Faculty of Health: Medicine, Dentistry and Human Sciences, University of Plymouth, Plymouth, UK

\*Corresponding author

**Declared competing interests of authors:** All authors report grant funding from the Versus Arthritis Centre of Excellence. Nadine E Foster reports Clinical Trials Unit funding from the National Institute for Health and Care Research (NIHR) (September 2018 to August 2021), membership of the Primary Care, Community and Preventive Interventions Panel (PCCPI) Topic Identification, Development and Evaluation (TIDE) Panel, NIHR Health Technology Assessment (HTA) Programme (2013–15) and funding support from an NIHR Research Professorship. Sue Jowett reports personal fees from Pfizer (Pfizer Inc., New York, NY, USA) and membership of the HTA Funding Committee Policy Group (February 2019–November 20) and HTA Clinical Evaluation and Trials Committee (November 2016–20).



Published June 2023  
DOI: 10.3310/FBVX4177

This report should be referenced as follows:

Foster NE, Dunn KM, Protheroe J, Hill JC, Lewis M, Saunders B, *et al.* Stratified primary care for adults with musculoskeletal pain: the STarT MSK research programme including RCTs. *Programme Grants Appl Res* 2023;**11**(4). <https://doi.org/10.3310/FBVX4177>



# Programme Grants for Applied Research

ISSN 2050-4322 (Print)

ISSN 2050-4330 (Online)

Programme Grants for Applied Research (PGfAR) was launched in 2013 and is indexed by Europe PMC, NCBI Bookshelf, DOAJ, Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and Scopus® (Elsevier, Amsterdam, Netherlands).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) ([www.publicationethics.org/](http://www.publicationethics.org/)).

Editorial contact: [journals.library@nihr.ac.uk](mailto:journals.library@nihr.ac.uk)

The full PGfAR archive is freely available to view online at [www.journalslibrary.nihr.ac.uk/pgfar](http://www.journalslibrary.nihr.ac.uk/pgfar).

## Criteria for inclusion in the *Programme Grants for Applied Research* journal

Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

## Programme Grants for Applied Research programme

The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health and Care Research (NIHR), was established in 2006 to fund collaborative, multidisciplinary programmes of applied research to solve health and social care challenges. Findings are expected to provide evidence that lead to clear and identifiable patient benefits, in the relatively near future.

PGfAR is researcher led and does not specify topics for research; however, the research must be in an area of priority or need for the NHS and the social care sector of the Department of Health and Social Care, with particular emphasis on health and social care areas that cause significant burden, where other research funders may not be focused, or where insufficient funding is available.

The programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director. For more information about the PGfAR programme please visit the website: <https://www.nihr.ac.uk/explore-nihr/funding-programmes/programme-grants-for-applied-research.htm>

## This report

The research reported in this issue of the journal was funded by PGfAR as project number RP-PG-1211-20010. The contractual start date was in June 2014. The final report began editorial review in February 2021 and was accepted for publication in June 2022. As the funder, the PGfAR programme agreed the research questions and study designs in advance with the investigators. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PGfAR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, CCF, PGfAR or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the PGfAR programme or the Department of Health and Social Care.

Copyright © 2023 Foster *et al.* This work was produced by Foster *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by NIHR Journals Library ([www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)), produced by Prepress, final files produced by Newgen Digitalworks Pvt Ltd, Chennai, India ([www.newgen.co](http://www.newgen.co)).

## NIHR Journals Library Editor-in-Chief

---

**Dr Cat Chatfield** Director of Health Services Research UK

## NIHR Journals Library Editors

---

**Professor Andrée Le May** Chair of NIHR Journals Library Editorial Group (HSDR, PGfAR, PHR journals) and Editor-in-Chief of HSDR, PGfAR, PHR journals

**Dr Peter Davidson** Interim Chair of HTA and EME Editorial Board. Consultant Advisor, School of Healthcare Enterprise and Innovation, University of Southampton, UK

**Professor Matthias Beck** Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

**Dr Tessa Crilly** Director, Crystal Blue Consulting Ltd, UK

**Dr Eugenia Cronin** Consultant in Public Health, Delta Public Health Consulting Ltd, UK

**Ms Tara Lamont** Senior Adviser, School of Healthcare Enterprise and Innovation, University of Southampton, UK

**Dr Catriona McDaid** Reader in Trials, Department of Health Sciences, University of York, UK

**Professor William McGuire** Professor of Child Health, Hull York Medical School, University of York, UK

**Professor Geoffrey Meads** Emeritus Professor of Wellbeing Research, University of Winchester, UK

**Professor James Raftery** Professor of Health Technology Assessment, School of Healthcare Enterprise and Innovation, University of Southampton, UK

**Dr Rob Riemsma** Consultant Advisor, School of Healthcare Enterprise and Innovation, University of Southampton, UK

**Professor Helen Roberts** Professor of Child Health Research, Child and Adolescent Mental Health, Palliative Care and Paediatrics Unit, Population Policy and Practice Programme, UCL Great Ormond Street Institute of Child Health, London, UK

**Professor Jonathan Ross** Professor of Sexual Health and HIV, University Hospital Birmingham, UK

**Professor Helen Snooks** Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of editors: [www.journalslibrary.nihr.ac.uk/about/editors](http://www.journalslibrary.nihr.ac.uk/about/editors)

**Editorial contact:** [journals.library@nihr.ac.uk](mailto:journals.library@nihr.ac.uk)



# Abstract

## Stratified primary care for adults with musculoskeletal pain: the STarT MSK research programme including RCTs

Nadine E Foster<sup>1,2</sup>, Kate M Dunn<sup>2</sup>, Joanne Protheroe<sup>2</sup>, Jonathan C Hill<sup>2\*</sup>, Martyn Lewis<sup>2,3</sup>, Benjamin Saunders<sup>2</sup>, Sue Jowett<sup>4</sup>, Susie Hennings<sup>3</sup>, Paul Campbell<sup>2,5</sup>, Kieran Bromley<sup>2,3</sup>, Bernadette Bartlam<sup>2,6,7</sup>, Opeyemi Babatunde<sup>2</sup>, Simon Wathall<sup>2,3</sup>, Raymond Oppong<sup>4</sup>, Jesse Kigozi<sup>4</sup> and Adrian Chudyk<sup>2</sup>

<sup>1</sup>STARS Education and Research Alliance, Surgical Treatment and Rehabilitation Service (STARS), The University of Queensland and Metro North Health, Brisbane, QLD, Australia

<sup>2</sup>Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, UK

<sup>3</sup>Keele Clinical Trials Unit, School of Medicine, Keele University, Keele, UK

<sup>4</sup>Health Economics Unit, Institute of Applied Health, University of Birmingham, Birmingham, UK

<sup>5</sup>Department of Research and Innovation, Midlands Partnership NHS FoundationTrust, St George's Hospital, Stafford, UK

<sup>6</sup>Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

<sup>7</sup>Community and Primary Care Research Group, Faculty of Health: Medicine, Dentistry and Human Sciences, University of Plymouth, Plymouth, UK

\*Corresponding author [j.hill@keele.ac.uk](mailto:j.hill@keele.ac.uk)

**Background:** Usual primary care for patients with musculoskeletal pain varies widely and treatment outcomes are suboptimal. Stratified care involves targeting treatments according to patient subgroups, in the hope of maximising treatment benefit and reducing potential harm or unnecessary interventions. This programme developed a new prognostic stratified primary care approach, where treatments are matched to a patient's risk of future persistent pain and disability based on a prognostic tool, and compared this with usual care.

**Objectives:** In four linked work packages, we refined and validated a prognostic tool [the Keele STarT MSK (Subgrouping for Targeted Treatment for Musculoskeletal pain) Tool] to identify risk of poor outcome and defined cut-off scores to distinguish patient risk subgroups (work package 1); defined and agreed new matched treatment options for each risk subgroup and developed a support package for delivery of stratified care (work package 2); tested the feasibility of delivering the stratified approach through a pilot randomised controlled trial and externally validated the prognostic tool (work package 3); and tested the effectiveness of the approach by comparing the clinical effectiveness and cost-effectiveness of stratified primary care with that of usual care through a cluster randomised controlled trial with embedded health economic and qualitative studies (work package 4).

**Setting:** General practices and linked musculoskeletal services in the West Midlands of England, UK.

**Participants:** Adults registered with participating practices consulting with back, neck, shoulder, knee or multisite musculoskeletal pain, and clinicians involved in managing these patients.

**Design:** The programme included the following work packages: work package 1 – a prospective cohort study in 12 practices; work package 2 – an evidence synthesis, consensus group workshops and qualitative studies; work package 3 – a cluster feasibility and pilot trial in eight practices; and work

package 4 – a main cluster randomised controlled trial in 24 practices, with health economic analyses and process evaluation.

**Interventions:** Stratified care comprised training general practitioners to use the tool and match patients to treatment options depending on their risk subgroup. Usual care comprised usual non-stratified primary care without formal stratification tools.

**Main outcome measures:** Cohort primary end points included function (Short Form questionnaire-36 items physical component score) and pain intensity (numerical rating scale). The trial primary end point for patient outcomes was pain intensity (monthly for 6 months) (0–10 numerical rating scale). An audit of primary care electronic medical records evaluated the impact of stratified care on clinical decision-making regarding patient management.

**Results:** Work package 1 – the cohort study ( $n = 1890$  patients) refined and validated a new 10-item tool with which to stratify patients with the five most common musculoskeletal pain presentations. The tool subgroups patients into three strata with different characteristics and prognoses. Work package 2: 17 treatment options were recommended – four for patients at low risk, 10 for patients at medium risk and 15 for patients at high risk. Work package 3: the feasibility and pilot trial included 524 patients, and the learning led to amendments to several tool items and a reduced set of treatments (14 in total). Work package 4: in the main trial, 1211 patients consented to data collection (534 in stratified care, 677 in usual care). Stratified primary care did not lead to statistically significant differences in the primary patient outcome of pain intensity [stratified care mean 4.4 (standard deviation 2.3) vs. usual care mean 4.6 (standard deviation 2.4); adjusted mean difference  $-0.16$ , 95% confidence interval  $-0.65$  to  $0.34$ ;  $p = 0.535$ ]. Where differences were observed, these were largely isolated to patients at high risk of poor outcome (the smallest subgroup), in favour of stratified care. Positive differences were, however, observed in general practitioner clinical decision-making, with increased provision of written self-management information and referrals to physiotherapy, plus reductions in prescription medication. The economic evaluation demonstrated that costs of care were similar across trial arms (£6.85, 95% confidence interval  $-\text{£}107.82$  to  $\text{£}121.54$  more for stratified care), with incremental quality-adjusted life-years of 0.0041 (95% confidence interval  $-0.0013$  to 0.0094), representing a net quality-adjusted life-year gain. Stratified care was associated with an incremental cost-effectiveness ratio of £1670 per additional quality-adjusted life-year gained. At a willingness-to-pay threshold ( $\lambda$ ) of £20,000 per quality-adjusted life-year, the incremental net monetary benefit was £132 and the probability of stratified care being cost-effective was approximately 73%. The very small differences suggest caution in the interpretation of this result. The qualitative findings revealed that general practitioners felt stratified care had a positive role in informing clinical decision-making, helped them to give greater attention to psychosocial issues and take a more functional approach, and facilitated negotiations with patients about treatment options such as imaging.

**Limitations:** The randomised controlled trial was not powered to detect differences between stratified and usual care for patients in each risk subgroup (low, medium and high) nor with each different musculoskeletal pain presentation. The stratified care electronic medical record template ‘fired’ only once per patient.

**Conclusions:** The Keele STarT MSK Tool is a valid instrument with which to discriminate between, and predict outcomes of, primary care patients with musculoskeletal pain. Although the randomised trial showed no significant benefit in patient-reported outcomes compared with usual care, some aspects of clinical decision-making improved and the approach was cost-effective.

**Future work:** The Keele STarT MSK Tool has been shared with over 1000 tool license requestees, leading to other work. Trial data sets have also led to other work, developing personalised prognostic models for back and neck pain patients (the European Union-funded Back-UP project). The challenge remains how to improve outcomes for primary care patients with musculoskeletal pain.

**Trial registration:** This trial is registered as ISRCTN15366334.

**Funding:** This project was funded by the National Institute for Health and Care Research (NIHR) Programme Grants for Applied Research programme and will be published in full in *Programme Grants for Applied Research*; Vol. 11, No. 4. See the NIHR Journals Library website for further project information.

# Contents

List of tables	xiii
List of figures	xv
List of boxes	xvii
List of abbreviations	xix
Plain language summary	xxi
Scientific summary	xxiii
<b>Synopsis</b>	<b>1</b>
Background and rationale for the programme	1
Aims of the research programme	1
<b>Work package 1</b>	<b>3</b>
Research aim and objectives	3
Methods	3
Analysis	4
External validity	4
Key findings	5
<i>Objective 1: confirm the validity and optimise predictive performance of the Keele STarT MSK Tool</i>	5
<i>Objective 2: determine the screening tool risk strata cut-off points</i>	5
<i>Objective 3: estimate the proportions of patients classified at low, medium and high risk of poor outcome and describe their characteristics</i>	6
<i>Objective 4: describe health-care resource use by all patients and in each risk stratum</i>	6
Alterations to initial plans and work package limitations	6
Conclusions	7
Interrelationship with other parts of the programme	7
<b>Work package 2</b>	<b>9</b>
Research aims and objectives	9
Methods and key findings for each study	9
<i>Evidence synthesis</i>	9
<i>Qualitative focus groups and interviews</i>	10
<i>Consensus on matched treatment options</i>	11
<i>Development and specification of an electronic template and training/support package</i>	12
Alterations to initial plans and work package limitations	13
Interrelationship with other parts of the programme	14
<b>Work package 3</b>	<b>15</b>
Research aims and objectives	15
Methods	15
Nested qualitative study	16

Key findings	16
<i>Objective 1: estimating participant recruitment and follow-up rates</i>	16
<i>Objective 2: evidence of selection bias</i>	17
<i>Objective 3: assessing general practitioner fidelity to the stratified care intervention</i>	17
<i>Objective 4: descriptive analyses of general practitioner decision-making and patient self-reported outcomes</i>	17
<i>Nested qualitative study key findings</i>	17
Alterations to initial plans and work package limitations	18
Interrelationship with other parts of the programme	19
<b>Work package 4</b>	<b>21</b>
Research aims and objectives	21
Methods	21
Data collection processes	22
<i>Objective 1: determine the comparative clinical effectiveness of stratified care compared with usual primary care for patient outcomes</i>	22
<i>Objective 2: investigate general practitioner fidelity to delivery of stratified care and the impact on clinical decision-making</i>	22
<i>Objective 3: undertake an economic evaluation of stratified care versus usual care</i>	22
<i>Objective 4: conduct a nested qualitative study to understand how stratified care was perceived and operationalised by clinicians and patients</i>	22
<i>Sample size and analysis</i>	22
<i>Health economics evaluation</i>	22
<i>Nested qualitative study</i>	23
Key findings	23
<i>Objective 1: determine the comparative clinical effectiveness of stratified care compared with usual primary care for patient outcomes</i>	23
<i>Objective 2: investigate general practitioner fidelity to delivery of stratified care and the impact on clinical decision-making</i>	24
<i>Objective 3: undertake an economic evaluation of stratified care compared with usual care</i>	24
<i>Objective 4: conduct a nested qualitative study to understand how stratified care was perceived and operationalised by clinicians and patients</i>	25
Alterations to initial plans and work package limitations	26
Conclusions	27
Interrelationship with other parts of the programme	27
<b>Patient and public involvement and engagement</b>	<b>29</b>
Work package 1	29
Work package 2	29
Work package 3	29
Work package 4	30
<b>Overall programme conclusions</b>	<b>31</b>
Implications for practice	33
Recommendations for future research	34
<b>Acknowledgements</b>	<b>35</b>
<b>References</b>	<b>39</b>

<b>Appendix 1</b> Tables and figures from work package 1	<b>43</b>
<b>Appendix 2</b> Tables and figures from work package 2	<b>47</b>
<b>Appendix 3</b> Tables and figures from work package 3	<b>57</b>
<b>Appendix 4</b> Stratified care intervention developed in work package 4	<b>63</b>
<b>Appendix 5</b> Summary health economic findings from main trial	<b>65</b>
<b>Appendix 6</b> Figures and tables from main trial in work package 4	<b>81</b>



# List of tables

<b>TABLE 1</b> Baseline characteristics of participants in the cohort study	43
<b>TABLE 2</b> Characteristics and outcomes in the KAPS cohort population, overall and within subgroups defined by the Keele STarT MSK Tool	44
<b>TABLE 3</b> Health-care utilisation over 6 months (complete cases)	45
<b>TABLE 4</b> Outline schedule for GP training session	47
<b>TABLE 5</b> Example monthly report showing peer-to-peer comparisons of GP engagement with recruitment template in work package 3	58
<b>TABLE 6</b> Example monthly report showing peer to peer comparisons of general clinical decision-making from intervention general practices in work package 3	59
<b>TABLE 7</b> Summary of participant self-reported measures used in work package 3	60
<b>TABLE 8</b> Descriptive and incremental health outcomes over 6 months for the base-case analysis and the complete case analyses. Values are mean (SD) scores unless stated otherwise	71
<b>TABLE 9</b> Resource components analysed in the trial. Values are mean (SD) resource per patient, by treatment group, for patients providing health care utilisation data at 6 months (n=978), unless stated otherwise	72
<b>TABLE 10</b> Cost components analysed in the trial. Values are mean (SD) costs per patient, by treatment group, for patients providing health care utilisation data at 6 months (n=978), unless stated otherwise	73
<b>TABLE 11</b> Productivity costs analysed in the trial. Values are mean (SD) costs per patient, by treatment group, for patients providing work outcomes data at 6 months	74
<b>TABLE 12</b> Cost-Utility analysis using the net-benefit regression approach <sup>†</sup>	75
<b>TABLE 13</b> Cost-effectiveness results for the within-trial economic analysis with 6-month horizon	76
<b>TABLE 14</b> Subgroup analysis: Costs and health outcomes mean (SD) scores by STarT MSK risk subgroups (Imputed analysis)	78
<b>TABLE 15</b> Cost-effectiveness outcomes for STarT MSK risk subgroups (Imputed analysis)	79
<b>TABLE 16</b> Characteristics of participants vs. non-participants in the main trial	85
<b>TABLE 17</b> General practice characteristics per trial arm in the main trial	86
<b>TABLE 18</b> Baseline participant characteristics, per trial arm in the main trial	86

<b>TABLE 19</b> Pain intensity scores (0 to 10-point NRS) and between-arm differences in the main trial	<b>90</b>
<b>TABLE 20</b> Number (%) achieving the MCIC for pain intensity NRS (at least 1-point change)	<b>91</b>
<b>TABLE 21</b> Pain intensity (0 to 10-point NRS) scores per patient risk subgroup (clinical version)	<b>92</b>
<b>TABLE 22</b> Pain intensity (0 to 10-point NRS) scores per subgroup of musculoskeletal pain site	<b>93</b>
<b>TABLE 23</b> Pain intensity (0 to 10-point NRS) scores per musculoskeletal pain region	<b>94</b>
<b>TABLE 24</b> Secondary outcome measures at 6 months, per trial arm in the main trial	<b>94</b>
<b>TABLE 25</b> Number and proportion of invited patients for whom GPs recorded an intention to provide a recommended matched treatment (stratified care arm) based on trial EMR template	<b>96</b>
<b>TABLE 26</b> Number and proportion of invited patients for whom the GP recorded selection of a recommended matched treatment (based on GP-entered information onto the trial-specific EMR template)	<b>96</b>
<b>TABLE 27</b> Number and proportion of trial participants for whom the GP recorded selection of a recommended matched treatment (stratified care arm) (based on GP-entered information onto the trial-specific electronic template)	<b>97</b>
<b>TABLE 28</b> Number and proportion of trial participants for whom the GP recorded selection of a recommended matched treatment (stratified care arm) (based on GP-entered information onto the trial-specific electronic template)	<b>98</b>
<b>TABLE 29</b> Comparison of GP decision-making per arm (0–182 days) taken from the anonymised EMR audit	<b>99</b>



# List of figures

<b>FIGURE 1</b> The Keele STarT MSK research programme pathway diagram	<b>2</b>
<b>FIGURE 2</b> The EMIS templates for matched treatment options	<b>49</b>
<b>FIGURE 3</b> The Keele STarT MSK Tool developed in work package 3 (clinician-completed version)	<b>62</b>
<b>FIGURE 4</b> Cost Effectiveness Acceptability Curve (Base-case)	<b>76</b>
<b>FIGURE 5</b> Incremental net benefits (INB) of SC vs UC (95% CI) (Base-case)	<b>77</b>
<b>FIGURE 6</b> Cost Effectiveness Acceptability Curve (Complete-case) analysis	<b>77</b>
<b>FIGURE 7</b> Incremental net benefits (INB) of SC vs UC (95% CI) complete case	<b>77</b>
<b>FIGURE 8</b> Main trial flowchart from work package 4	<b>81</b>
<b>FIGURE 9</b> Overall pain intensity (NRS) scores per trial arm, over time	<b>82</b>
<b>FIGURE 10</b> Pain intensity (NRS) scores per trial arm for patients at low risk of poor outcome, over time	<b>82</b>
<b>FIGURE 11</b> Pain intensity (NRS) scores per trial arm for patients at medium risk of poor outcome, over time	<b>82</b>
<b>FIGURE 12</b> Pain intensity (NRS) scores per trial arm for patients at high risk of poor outcome, over time	<b>83</b>
<b>FIGURE 13</b> Pain intensity (NRS) scores per trial arm for patients with back pain	<b>83</b>
<b>FIGURE 14</b> Pain intensity (NRS) scores per trial arm for patients with neck pain	<b>83</b>
<b>FIGURE 15</b> Pain intensity (NRS) scores per trial arm for patients with shoulder pain	<b>84</b>
<b>FIGURE 16</b> Pain intensity (NRS) scores per trial arm for patients with knee pain	<b>84</b>
<b>FIGURE 17</b> Pain intensity (NRS) scores per trial arm for patients with multisite pain	<b>84</b>



# List of boxes

<b>BOX 1</b> Clinician support package developed in work package 2	<b>47</b>
<b>BOX 2</b> Per-protocol recommended matched treatment options for each risk subgroup developed in work package 3	<b>57</b>
<b>BOX 3</b> Per-protocol recommended matched treatment options for each risk subgroup used in work package 4	<b>63</b>



# List of abbreviations

AMSTAR	A MeaSurement Tool to Assess systematic Reviews	OR	odds ratio
BNF	<i>British National Formulary</i>	OTC	over the counter
CHEERS	Consolidated Health Economic Evaluation Reporting Standards	PCS	physical component score
CI	confidence interval	PPIE	patient and public involvement and engagement
COM-B	capability, opportunity, motivation and behaviour	PROMIS	Patient-Reported Outcomes Measurement Information System
EHR	electronic health record	PSC	Programme Steering Committee
EMR	electronic medical record	PSS	Personal Social Services
EQ-5D-5L	EuroQol-5 Dimensions, five-level version	QALY	quality-adjusted life-year
GP	general practitioner	RCT	randomised controlled trial
GRADE	Grading of Recommendations Assessment, Development and Evaluation	RMDQ – original version	Roland Morris Disability Questionnaire
IRR	incident rate ratio	RUG	Research User Group
KAPS	Keele Aches and Pains Study	SD	standard deviation
MCIC	minimum clinically important change	SF-36	Short Form questionnaire-36 items
NGT	Nominal Group Technique	SMD	standardised mean difference
NIHR	National Institute for Health and Care Research	SMS	short message service
NPT	Normalisation Process Theory	STarT MSK	Subgrouping for Targeted Treatment for Musculoskeletal pain
NRS	numerical rating scale	TDF	Theoretical Domains Framework
NSAID	non-steroidal anti-inflammatory drug	TENS	transcutaneous electrical nerve stimulation



## Plain language summary

**A**round 20% of adults consult their general practitioner each year with a musculoskeletal problem. Musculoskeletal problems represent the largest group of long-term conditions for which patients consult their general practitioner. Although many patients have mild symptoms that can be self-managed with support, some have more serious or persistent problems that need treatments such as physiotherapy. Care varies widely because it is not easy to identify who should be supported to self-manage and who needs more treatment.

In this programme we adapted and finalised a short questionnaire [the Keele STarT MSK (Subgrouping for Targeted Treatment for Musculoskeletal pain) Tool] to sort patients consulting their general practitioner with back, neck, shoulder, knee or multisite pain into one of three groups: those at low, medium and high risk of a poor outcome (for example, patients still experiencing high levels of pain or disability at 6-month follow-up). We recommended treatment options that general practitioners could provide, or offer referral to, for each group of patients and gave general practitioners training to use the new approach. We consulted patients, general practitioners and physiotherapists about this new approach and tested whether or not it led to better outcomes for patients and whether or not it was good value for money for the NHS.

Results showed that the Keele STarT MSK Tool works well to identify patients at low, medium and high risk of poor outcome. The approach had several positive impacts on general practitioner clinical decision-making and received positive feedback from patients and clinicians. However, matching groups of patients to the available treatments recommended in this programme did not lead to consistently better patient outcomes than those of patients receiving usual care. We found stratified care had greater potential for the group of patients at high risk.

The challenge remains to improve primary care treatments in ways that lead to better outcomes for patients with musculoskeletal pain.





# Scientific summary

## Background

Musculoskeletal pain such as back, neck, shoulder, knee and multisite pain is common and costly in terms of burden on individuals, the NHS and society. For some people these musculoskeletal problems are painful but short-lived; however, for others the painful episode fails to resolve or recurs, impacting their ability to work and leading to extensive NHS and societal costs. Most patients with musculoskeletal problems are managed in primary care, where 20% of the registered practice population will consult their general practitioner (GP) annually with a musculoskeletal problem, accounting for one in six GP consultations. There is limited evidence to guide how best to direct the right patient to the right treatment in ways that improve patient outcomes such as pain and disability, and make the best use of health-care resource. The sheer number of patients makes it inappropriate and unsustainable to offer more intensive and expensive treatments to all. Building on a previously successful model of prognostic stratified care for patients in primary care with low back pain, the aims of this programme were to adapt, finalise and test a prognostic stratified primary care model for a much larger group of patients with the five most common musculoskeletal pain presentations.

## Objectives

The programme addressed the following objectives:

- on the tool to subgroup patients into prognostic strata –
  - validate and optimise the predictive performance of the Keele STarT MSK (Subgrouping for Targeted Treatment for Musculoskeletal pain) Tool
  - determine the tool risk strata cut-off points based on optimal predictive values and suitability for matched treatment options
  - estimate the proportions of patients at low, medium and high risk of poor outcome and describe their characteristics
  - describe current health-care resource use by all patients and in each risk stratum.
- on the recommended matched treatment options –
  - summarise current best evidence on available treatments for the five most common musculoskeletal pain presentations in primary care
  - explore patients' and GPs' views on the acceptability of prognostic stratified care and the anticipated barriers to and facilitators of its use in clinical practice
  - agree, through expert consensus, the most appropriate matched treatment options that should be recommended for patients in each risk subgroup
  - develop and specify the training and support package to support delivery of stratified primary care.
- on the feasibility of a future main cluster randomised controlled trial (RCT) and of delivery of stratified primary care –
  - estimate participant recruitment and follow-up rates for the main cluster RCT
  - examine evidence of selection bias between trial arms and between participants and non-participants

- assess GP fidelity to the stratified care intervention (use of the stratification tool and matched treatment options)
- conduct secondary descriptive analyses of GP decision-making and patient outcomes.
- on the clinical effectiveness and cost-effectiveness of stratified primary care compared with usual care (usual care) –
  - determine the comparative clinical effectiveness of stratified care versus usual primary care for patient outcomes
  - investigate GP fidelity to delivery of stratified care and the impact on clinical decision-making
  - undertake an economic evaluation of stratified care versus usual care
  - conduct a nested qualitative study to understand how stratified care was perceived and operationalised by clinicians and patients.

## Methods

A series of studies across four work packages was carried out, involving different research methods:

- work package 1 – a prospective longitudinal cohort study and interviews with patients and clinicians
- work package 2 – an evidence synthesis, qualitative focus groups and interviews, consensus group workshops, and development of an electronic template and training/support package to support GPs to deliver stratified care
- work package 3 – a feasibility and pilot RCT with linked qualitative interviews
- work package 4 – a main cluster RCT including analyses of general practice medical record data, health economic analyses and qualitative interviews.

## Results

### *Work package 1 – the Keele Aches and Pains Study (KAPS)*

A total of 1890 patients responded to the invitation and consented to participate (mean age 58 years, 61% female). Subsequently, 1425 participants returned questionnaires at 2 months' follow-up (response rate of 75.8%), and 1452 provided data at 6 months' follow-up (response rate of 78.7%). The amended 10-item Keele STarT MSK Tool has a scale range of 0–12 (0 = lowest risk of poor outcome, 12 = highest risk of poor outcome). The final model fit ( $R^2$ ) and discrimination (c-statistic) in the Keele Aches and Pains Study (KAPS) data set at 6 months' follow-up was 0.422 and 0.839 for physical function, respectively, and 0.430 and 0.822 for pain intensity, respectively; there was also acceptable performance across the five musculoskeletal pain presentations. In the external validation data set, the final model fit ( $R^2$ ) was 0.224 and discrimination (c-statistic) was 0.725 for pain intensity. The cut-off points determined to provide the best combination of sensitivity, specificity, predictive values and likelihood ratios, in combination with suitability for the recommended matched treatments (identified in work package 2), overall and across pain sites, were 0–4 for low risk, 5–8 for medium risk and 9–12 for high risk. The KAPS cohort participants were classified as 25% at low risk, 42% at medium risk and 33% at high risk of poor outcome. There were clear and consistent differences between risk subgroups on key variables, health-care utilisation and associated costs, confirming the discriminative ability of the Keele STarT MSK Tool.

### *Work package 2 – study to agree matched treatment options and support package for delivery*

The evidence synthesis showed that primary care patients with musculoskeletal pain can be managed effectively with non-pharmacological treatments such as self-management advice, exercise therapy and psychosocial interventions, and that pharmacological treatments provide, at best, short-term benefits only. The qualitative focus groups and interviews with patients and clinicians identified four key themes (the acceptability of clinical decision-making guided by stratified care, the impact on the therapeutic

relationship, embedding a prognostic approach within a biomedical model and practical issues in using stratified care). For GPs and patients to see stratified care as useful, it must add to existing knowledge and skills, not undermine GPs' clinical autonomy nor disrupt therapeutic rapport. The need for integration into consultations with minimal disruption was highlighted. In the consensus study, the three Nominal Group Technique workshops with multidisciplinary groups of clinicians led to a total of 17 treatment options being recommended. These were summarised and incorporated into a bespoke electronic health record (EHR) template along with the Keele STarT MSK Tool, in order to make delivery of stratified care as easy as possible for GPs, and a training/support package for GPs was developed, ready for use with GPs participating in the feasibility and pilot trial in the next work package.

### ***Work package 3 – the Treatment for Aches and Pains Study (TAPS) pilot trial***

In work package 3's feasibility and pilot RCT, 524 participants (42% of those invited) consented to provide questionnaire outcome data (stratified care  $n = 231$ , usual care  $n = 293$ ). Anonymised EHR data were available for 1281 patients (529 in stratified care practices and 752 in usual care practices). Although follow-up rates over 6 months were high, the length of time taken to recruit participants (28 weeks) was over twice as long as expected (12 weeks). GP fidelity to use of the stratification tool was only 40% of eligible consultations (compared with a target of at least 50%), and both GPs and patients identified the need for several items in the tool to be amended. However, in those with whom the tool was used, over 80% were recorded as having been matched to a recommended treatment option. Key changes were therefore needed prior to the main trial; thus, this feasibility and pilot trial became an external pilot trial (rather than the originally intended internal pilot trial). There was no evidence of selection bias and, therefore, no changes were made to identification or recruitment procedures for the main trial. Given the learning from this pilot trial, amendments were made to the stratification tool (revision of the language used in four of the tool items) and matched treatment options (a total of 14 were recommended).

### ***Work package 4 – the Treatment for Aches and Pains Study (TAPS) main trial***

In the main cluster RCT in work package 4, 1211 patients from 24 general practices (12 per arm) participated in self-report data collection (534 in stratified care and 677 in usual care); the participants had a mean age of 60 years and 58.9% were female. Mean pain intensity at the point-of-consultation was 6.73 (6.77 in stratified care, 6.70 in usual care). A total of 1178 (97%) participants provided at least one pain intensity response over the 6 months' follow-up [515 (96%) in stratified care, 663 (98%) in usual care] and 80.9% responded to the follow-up questionnaire at 6 months (77.9% in stratified care, 83.3% in usual care).

The main analysis showed no statistically significant differences in pain intensity over 1–6 months comparing stratified care with usual care, with mean values of 4.4 [standard deviation (SD) 2.3] for stratified care and 4.6 (SD 2.4) for usual care [adjusted mean difference  $-0.16$ , 95% confidence interval (CI)  $-0.65$  to  $0.34$ ;  $p = 0.535$ ]. Most sensitivity analyses showed no statistically significant between-arm differences, despite showing consistent slightly favourable results for stratified care. Subgroup analyses showed some between-arm mean differences with a greater difference (although statistically non-significant) in patients at high risk versus those at low risk, and between those with shoulder and knee pain compared with those with neck and back pain. There were no statistically significant differences in secondary clinical outcomes at 6 months, except for a significant improvement in shoulder pain and function and higher satisfaction with care in the stratified care arm compared with the usual care arm.

The health economic evaluation showed that the costs of care were very similar in the two arms of the trial: mean cost of stratified care was £356.36 (SD £864.01) compared with a mean cost of £343.44 (SD £942.92) for usual care. The adjusted incremental cost of stratified care compared with usual care over the 6 months was £6.85 (95% CI  $-\text{£}107.82$  to  $\text{£}121.54$ ), with incremental quality-adjusted life-years (QALYs) of 0.0041 (95% CI  $-0.0013$ ,  $0.0094$ ), representing a net QALY gain. Stratified care was associated with a cost-per-QALY gain of £1670. At a willingness-to-pay threshold ( $\lambda$ ) of £20,000 per

QALY, the incremental net monetary benefit was £132 and the probability of stratified care being cost-effective was approximately 73%.

The anonymised EHR audit data were available for 2494 patients across all 24 practices and demonstrated several important impacts from stratified care on GP treatment decision-making, including increased provision of written information (58% in the stratified care arm vs. 26% in the usual care arm), physiotherapy referral (63.3% in the stratified care arm vs. 9.9% in the usual care arm), and simple over-the-counter analgesics (16.7% in the stratified care arm vs. 6.3% in the usual care arm). Prescribing of short-term courses of strong opioids increased (20.3% in the stratified care arm vs. 1.0% in the usual care arm), but not long-term opioids. GPs in stratified care practices completed the risk-stratification tool in 29.76% (1056/3548) of possible consultations and reported selecting an appropriate risk-matched treatment option in over three-quarters of patients (77.2% low risk, 77.8% medium risk and 76.7% high risk).

## Patient and public involvement

Patient and public involvement and engagement (PPIE) representatives contributed at all stages and to all work packages. In work package 1 they made suggestions on patient-facing materials that improved acceptability and understanding, and gave patient/public perspectives on the draft Keele STarT MSK Tool's face validity. PPIE representative views influenced the presentation of evidence for the work package 2 consensus study, and aided the choice of patient websites and leaflets recommended for GP use (in matched treatment options). In work package 3, PPIE perspectives improved study design and patient-facing materials. PPIE representatives helped the team understand and interpret feasibility/pilot trial findings (quantitative and qualitative), and influenced main trial plans. PPIE input was valuable in wording and ordering the clinician-completed version of the Keele STarT MSK Tool. In work package 4, PPIE input led to improvements in wording and formatting of study documents, and advised on acceptable methods for GPs seeking patient consent. PPIE views were also important in the interpretation of study findings.

## Conclusions

The Keele STarT MSK Tool is a valid instrument with which to discriminate between, and predict outcomes of, primary care patients with musculoskeletal pain. However, matching groups of patients to the available treatment options recommended in this programme did not lead to consistently better clinical outcomes than those receiving usual care. Although the randomised trial showed no significant benefit in patient-reported outcomes compared to usual care, some aspects of clinical decision-making improved and there was only a marginal increase in cost.

## Implications for health care

The Keele STarT MSK Tool is a valid tool with which to identify patients with different levels of risk of persistent pain. The tool provides additional systematic information about an individual patient's prognosis that can help clinicians to direct patients to the most appropriate treatments. The approach of using one tool for this wide range of patients has the major benefit of simplicity for clinical practice, removing the complexity that would result from multiple, pain site-specific screening tools. The main trial showed that although matching patients to treatment options based on their risk subgroup did not lead to superior patient-reported outcomes for the overall trial comparison, there were important improvements in some aspects of clinical decision-making about treatments. Costs of risk-based stratified care were similar to usual care, with marginal additional benefits. The main trial results are partly explained by a loss in fidelity in terms of the delivery of stratified care by participating GPs, likely

explained by the additional burden of time to deliver stratified care in consultations with patients. The stratified care approach with the EHR template comprising the tool and 14 treatment options may have been too complex to deliver with high fidelity. The challenge remains to improve primary care treatments in ways that lead to better outcomes for patients with musculoskeletal pain.

## Implications for research

Stratified care involves matching subgroups of patients to treatments in ways that improve clinical outcomes, reduce unnecessary or harmful treatments and make better use of health-care resource. This programme demonstrated that it is possible to use one brief (10-item) stratification tool to accurately identify the prognosis of patients with musculoskeletal pain using simple self-report information. This required the GP to ask patients questions and record their responses, adding time to consultations. Future research needs to identify ways to use more routinely collected data about patients with musculoskeletal pain so that prognostic subgroup information can be provided to clinicians in more time-efficient ways. The stratified care EHR template in this programme fired only once per patient, so as not to burden GPs, and although this led to important changes in some aspects of clinical decision-making, patients had an average of 4 to 5 musculoskeletal pain-related consultations over 6 months' follow-up. There was no electronic reminder of stratified care during these consultations. Therefore, future research that tests ways to continually 'nudge' clinical decision-making in the right direction is needed. For most patients with musculoskeletal pain, this will require efforts to reduce medication prescriptions and instead support self-management and ensure access to non-pharmacological treatments.

## Trial registration

This trial is registered as ISRCTN15366334.

## Funding

This project was funded by the National Institute for Health and Care Research (NIHR) Programme Grants for Applied Research programme and will be published in full in *Programme Grants for Applied Research*; Vol. 11, No. 4. See the NIHR Journals Library website for further project information.



# Synopsis

## Background and rationale for the programme

Musculoskeletal pain such as back pain, neck, shoulder, knee and multisite pain is common and costly in terms of the burden on individuals, the NHS and society.<sup>1</sup> For some people these musculoskeletal problems are painful but short-lived; however, for others the painful episode fails to resolve or recurs, impacting day-to-day function including ability to work and leading to extensive NHS and societal costs.

Most patients with musculoskeletal problems are managed in primary care, where 20% of the registered practice population will consult their general practitioner (GP) annually with a musculoskeletal problem, accounting for 1 in 6 GP consultations.<sup>2</sup> There is limited evidence to guide how best to direct the right patient to the right treatment in primary care in ways that improve patient outcomes such as pain and function, and make the best use of health-care resource. The sheer number of patients with musculoskeletal pain and the wide variation in their prognosis means that it not appropriate to offer intensive or expensive treatments to all.<sup>3</sup>

Stratified care involves targeting treatments according to patient subgroups, in the hope of maximising treatment benefit and reducing potential harm or unnecessary interventions.<sup>4</sup> Building on a previously successful model of prognostic stratified care for patients in primary care with low back pain,<sup>5-7</sup> the aims of this programme were to adapt, finalise and test a prognostic stratified primary care model for a much larger group of patients with the five most common musculoskeletal pain presentations.

## Aims of the research programme

The aims of this programme were to adapt, finalise and test a prognostic stratified primary care model for primary care patients consulting with the five most common pain presentations (back, neck, shoulder, knee or multisite pain), comparing it with usual primary care. A series of studies involving different methods (a prospective longitudinal cohort study, interviews with patients and clinicians, an evidence synthesis, consensus group workshops, the development of tools and training to support GPs to use the stratified care approach, and clinical trials including analyses of general practice medical record data and health economic analyses) were carried out to address the following research questions:

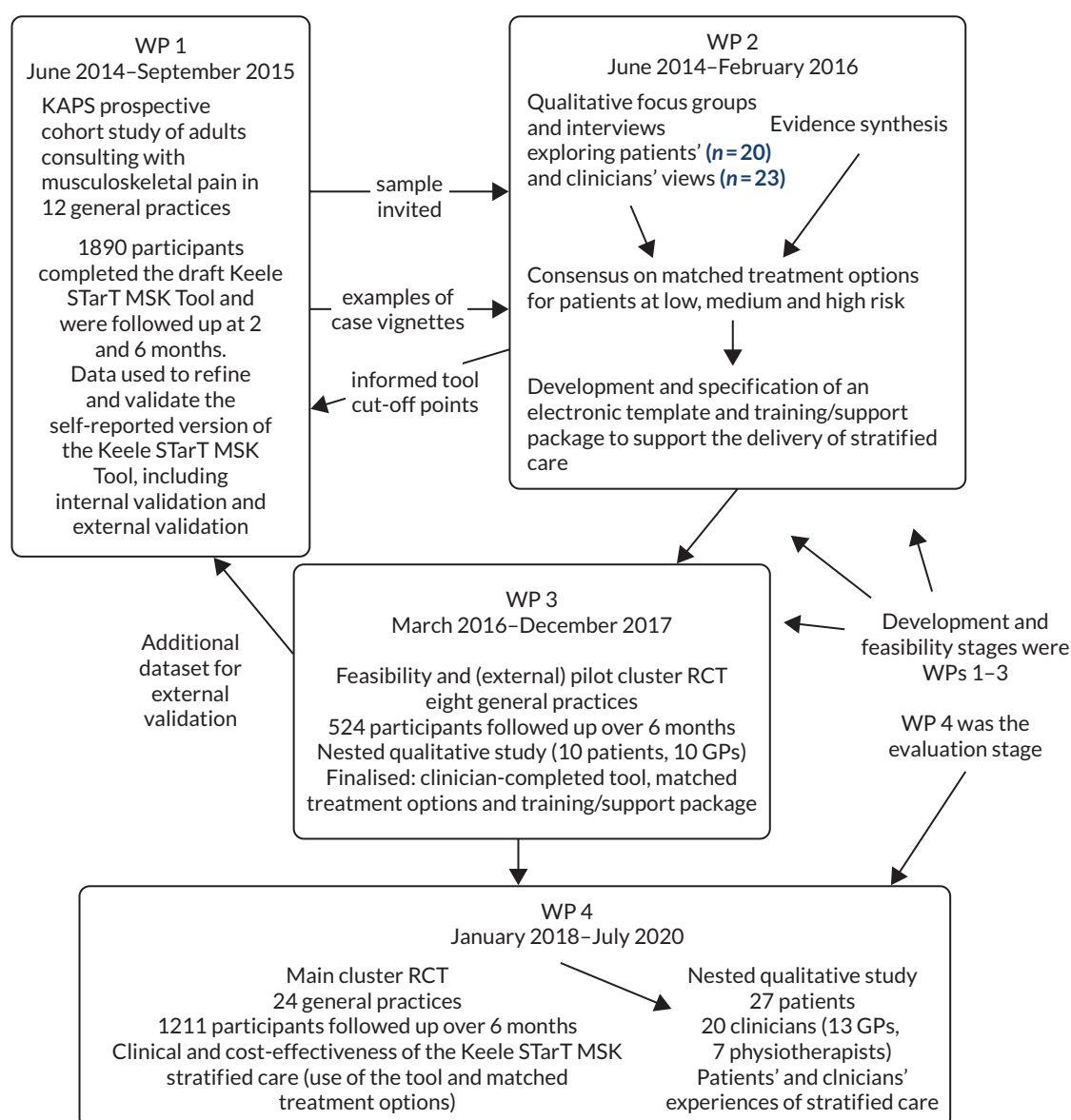
- Can one prognostic tool [the Keele STarT MSK Tool (Subgrouping for Targeted Treatment for Musculoskeletal pain)] identify the risk of poor outcome for a wide range of patients with the most common musculoskeletal pain presentations in primary care, and does it discriminate subgroups at low, medium and high risk of persistent disabling pain?
- What are the most appropriate treatment options that should be recommended for patients in each risk subgroup?
- What are the views and experiences of patients and clinicians about using a prognostic stratified care approach in the management of musculoskeletal pain?
- What is the feasibility of (1) delivering the stratified care intervention in primary care and (2) conducting a large randomised controlled trial (RCT) to test this new model of care?
- In adults with the most common musculoskeletal pain presentations in primary care, does prognostic stratified care (involving use of the Keele STarT MSK Tool and recommended matched treatments) lead to better patient outcomes, greater cost-effectiveness and different clinical decision-making than usual non-stratified primary care?



In addition, a stratified care clinician training and support package comprising electronic templates, a stratified care tool and recommended treatment options to support delivery of stratified primary care was developed.

The research was carried out between June 2014 and September 2020.

The inter-relationships between the four work packages are summarised in [Figure 1](#) (below). The prospective cohort study in work package 1 provided the sampling frame for the qualitative interviews in work package 2. The feasibility and pilot RCT in work package 3 provided the data set for the external validation of the Keele STarT MSK Tool in work package 1.



**FIGURE 1** The Keele STarT MSK research programme pathway diagram. KAPS, Keele Aches and Pains Study.



# Work package 1

Previously, we showed that stratified care involving targeting treatments according to patient subgroups based on the STarT Back tool<sup>5</sup> was clinically effective and cost-effective for patients with low back pain.<sup>6</sup> Preliminary analyses of a modified version of the STarT Back tool amended to better suit patients with a broader range of musculoskeletal pain presentations (e.g. back, neck, upper limb, lower limb or multisite pain) showed promise, but highlighted that modifications to this draft Keele STarT MSK Tool were required because risk varied for patients with pain at different body sites.<sup>8</sup> In addition, the new Keele STarT MSK Tool needed to be validated with the target patient group: those consulting in general practice with musculoskeletal pain.

## Research aim and objectives

The aim of this work package was to refine and validate an instrument, the Keele STarT MSK Tool, designed to identify risk of poor outcome in primary care patients with the most common musculoskeletal pain presentations.

The four specific objectives were as follows:

1. confirm the validity and optimise the predictive performance of the Keele STarT MSK Tool
2. determine the screening tool risk strata cut-off points based on optimal predictive values and suitability for matched treatment options
3. estimate the proportions of patients classified at low, medium and high risk of poor outcome and describe their characteristics
4. describe current health-care resource use by all patients and in each risk stratum.

## Methods

We carried out a prospective cohort study of consecutive consulters in UK general practice [named the Keele Aches and Pains Study (KAPS)]. The protocol for the study has been published.<sup>9</sup>

Patients aged  $\geq 18$  years presenting with one or more of the five most common musculoskeletal pain presentations (back, neck, shoulder, knee or multisite pain) were identified at 12 general practices in the West Midlands in England. Participants were recruited between July 2014 and February 2015. Patients were excluded if there were indications of potentially serious underlying pathology (such as cancer or infection), they had urgent care needs, they were vulnerable or if they were unable to communicate in English. We mailed out information about the study and baseline questionnaire and, for those providing consent, we sent further questionnaires at 2 and 6 months' follow-up and analysed their primary care electronic medical records (EMRs). We estimated that we would need to identify approximately 3000 eligible patients to recruit 1800 participants at baseline (based on a 60% response rate). We anticipated that this would provide 1250 patients at 6 months follow-up, including 125 patients at high risk of poor outcome (the smallest subgroup), which would provide adequate power for validation of the draft Keele STarT MSK Tool.

Our main outcome measures were physical function [Short Form questionnaire-36 items (SF-36) physical component score (PCS)], pain intensity [0–10 numerical rating scale (NRS)], pain interference [Patient-Reported Outcomes Measurement Information System (PROMIS) pain interference scale] and health-related quality of life [EuroQol-5 Dimensions, five-level version (EQ-5D-5L)], measured at 2 and 6 months' follow-up. For dichotomous outcomes, poor outcome on the PCS was defined as scores of  $< 37.17$  at 2 months and  $< 39.61$  at 6 months, based on lower tertiles from an independent study of

UK primary care musculoskeletal pain patients.<sup>10</sup> Poor outcome on pain intensity was defined as NRS scores of  $\geq 5$  points.<sup>11</sup> Secondary outcomes are detailed in our protocol paper by Campbell *et al.* (2016).<sup>9</sup> The baseline questionnaires also included the draft Keele STarT MSK Tool (identical to the modified STarT Back tool) presented in our paper by Hill *et al.* (2016).<sup>8</sup> The responses to the KAPS baseline and follow-up questionnaires are provided in [Appendix 1](#).

## Analysis

Predictive performance was determined using linear regression of the association between baseline tool score and PCS and pain intensity at 2 and 6 months' follow-up. Performance was assessed based on model fit ( $R^2$ ) and discrimination (c-statistic) and calibration (calibration slope and Hosmer–Lemeshow test).

Item redundancy and weighting was investigated within multiple linear regression models for estimating PCS score at 2 and 6 months, and pain intensity at 2 and 6 months. If items did not add significant predictive performance and/or if average standardised beta weight was small (i.e.  $< 0.05$ ) in most analyses, then the item was deemed statistically redundant. The research team, in consultation with members of our patient and public involvement and engagement (PPIE) group, considered face validity in decisions about the removal of statistically redundant items.

The tool cut-off point for identifying the high-risk subgroup (compared with those at medium/low risk) was based on classification on the full score most likely to attain positive predictive values and specificity  $\geq 0.8$ , and positive likelihood ratio  $\geq 5$ , for predicting pain and function at 2 and 6 months' follow-up. The tool cut-off point for categorising the low-risk subgroup (compared with those at medium/high risk) was based on the classification most likely to achieve negative predictive values and sensitivity  $\geq 0.80$ , and negative likelihood ratio  $\leq 0.2$ .<sup>12-14</sup> All decisions about tool cut-off points were based on statistical information in the sample overall and within pain sites, plus suitability for matched treatments.

Health-care utilisation data were collected from patient questionnaires and medical record review at 6 months' follow-up. Information included primary and secondary care contacts, prescribed medications, treatments, tests and investigations. Unit costs were obtained primarily from standard national sources such as NHS Reference costs,<sup>15</sup> *Unit Costs of Health and Social Care*<sup>16</sup> and the *British National Formulary* (BNF),<sup>17</sup> and applied to resource use data.

Descriptive statistics were used to summarise data on health-care utilisation and a total health-care cost per patient over 6 months was also estimated. Mean total costs for each patient risk subgroup were calculated and non-parametric bootstrapping (1000 replications) was used to estimate bias-corrected CIs around differences in mean costs.

## External validity

Independent testing was carried out within the Keele STarT MSK feasibility and pilot trial data set in work package 3, with the Keele STarT MSK Tool completed during GP–patient consultations, and using pain intensity at 6 months' follow-up as the outcome (see *Work package 3* and the pilot trial findings publication by Hill *et al.*<sup>18</sup>). Discriminant and predictive validity were investigated using model fit and discrimination as above. Descriptive analysis of outcomes within risk strata of the final Keele STarT MSK Tool were investigated.

## Key findings

Overall, 4720 patients visited their GP about back, neck, knee, shoulder or multisite pain and were invited to participate in the cohort study. A total of 2057 patients responded (43.6% response rate), and 1890 consented to participate (40.2% adjusted response rate owing to incomplete/ineligible questionnaires and refusals). The mean age of participants was 58 years (range 18–96 years), and 61% were female. Response rates at 2 and 6 months were 75.8% ( $n = 1425$ ) and 78.7% ( $n = 1452$ ), respectively.

The mean baseline physical function (PCS) score was 36.2 [standard deviation (SD) 10.1], mean baseline pain intensity was 5.3 (SD 2.4) and 22% of the sample reported having pain for less than 3 months. Multisite pain was the most commonly reported reason for consulting the GP, followed by low back pain and knee pain (further baseline characteristics are shown in [Appendix 1, Table 1](#)). Response rates to the 2- and 6-month follow-up questionnaires were 75.8% and 78.7%, respectively. At 2 and 6 months, mean PCS scores rose to 38.1 and 38.6, respectively (indicating modest improvements in physical function); 47.7% ( $n = 560$ ) at 2 months and 53.4% ( $n = 581$ ) at 6 months were categorised as having a poor outcome on the PCS. Mean pain intensity reduced to 4.4 at 2 months and 4.1 at 6 months, with 45.6% ( $n = 582$ ) and 42.3% ( $n = 482$ ) categorised as having a poor pain outcome, respectively.

### **Objective 1: confirm the validity and optimise predictive performance of the Keele STarT MSK Tool**

We investigated whether or not adding, removing or replacing items from the draft tool led to improvements in predictive performance and face validity. A list of candidate items was included in the baseline questionnaire; these covered domains including vitality/fatigue, comorbidity, coping, sleep problems, previous treatment success, pain interference, pain self-efficacy, pain persistence, pain-related depression and fear-avoidance beliefs. Three candidate items were used to replace items in the draft tool because they improved model fit ( $R^2$ ) from 0.334 to 0.405 and discrimination (c-statistic) from 0.804 to 0.815 against physical function (PCS) at 6 months follow-up, and were perceived to improve face validity.

We then examined this amended version of the Keele STarT MSK Tool within the independent sample of 524 patients in the feasibility and pilot RCT described in *Work package 3* (mean age 61 years, 61% female). Analyses indicated unacceptable reductions in model fit ( $R^2$  0.149) and discrimination (c-statistic 0.685). We, therefore, returned to the KAPS data set, and identified additional items that further improved the tool. Investigation of item redundancy, in combination with assessment of face validity by the team led to the removal of two items. This led to a final 10-item Keele STarT MSK Tool with scale range 0–12 (0 = lowest risk of poor outcome, 12 = highest risk of poor outcome). Each item is scored 0 or 1 except the item on pain intensity, which is scored on a subscale of 0 to 3, indicating increasing severity of pain; this weighting reflects a higher standardised beta coefficient compared with other items in the final model. The final model fit at 6 months' follow-up was 0.422 and discrimination 0.839 for physical function (PCS), and 0.430 and 0.822 for pain intensity; there was acceptable performance across the five musculoskeletal pain presentations. Multiple imputation indicated that tool performance was robust to missing data. The final model also resulted in improved model fit ( $R^2$  0.224) and discrimination (c-statistic 0.725) for pain intensity in the external data set from work package 3.

### **Objective 2: determine the screening tool risk strata cut-off points**

The cut-off points determined to provide the best combination of sensitivity, specificity, predictive values and likelihood ratios, in combination with suitability for the recommended matched treatments (identified in work package 2), overall and across pain sites, were 0–4 for low risk, 5–8 for medium risk, and 9–12 for high risk, on the full scale.

### **Objective 3: estimate the proportions of patients classified at low, medium and high risk of poor outcome and describe their characteristics**

The KAPS cohort participants were classified as 25% at low risk, 42% at medium risk and 33% at high risk. There were clear and consistent differences between risk subgroups on key variables at both baseline and follow-up. For example, mean baseline physical function (PCS) scores were 45.8 for patients at low risk, 36.8 for those at medium risk and 28.4 for those at high risk, with mean 6-month follow-up scores of 48.0, 39.0 and 30.2, respectively. Pain intensity showed comparable differences, with mean scores at 6 months of 1.9 for the low-risk subgroup, 4.0 for the medium-risk subgroup and 6.2 for the high-risk subgroup. There were also clear differences in the other key outcomes, pain interference (PROMIS pain interference scale) and health-related quality of life (EQ-5D-5L), as well as differences on the secondary outcome measures. Further details are presented in [Appendix 1, Table 2](#). These patterns were still evident when the data were examined by site of pain presentation (back, neck, shoulder, knee and multisite pain).

### **Objective 4: describe health-care resource use by all patients and in each risk stratum**

Overall, there was a mean of 1.44 GP visits for musculoskeletal pain per participant over the 6-month follow-up period (not including the index consultation); this ranged from a mean of 0.66 visits in the low-risk subgroup, to 1.40 in medium risk, to 2.22 in the high-risk subgroup. The risk subgroups differed on all areas of health-care utilisation, for example the mean number of prescriptions ranged from 1.80 per person in the low-risk subgroup to 10.45 per person in the high-risk subgroup over the 6-month period. Further details are given in [Appendix 1, Table 3](#).

Total mean health-care costs per participant over the 6 months were £132.92 (SD £167.88), £279.32 (SD £462.98) and £476.07 (SD £716.44) for patients in the low-, medium- and high-risk subgroups, respectively.

## **Alterations to initial plans and work package limitations**

We anticipated 60% participation among those invited to participate in the KAPS cohort study, based on previous similar studies, but only 40% consented to participate. With approval from the Programme Steering Committee (PSC), we extended recruitment by 3 months to achieve the required sample, but the lower initial response rate may have introduced further bias into the characteristics of the study sample. This is most likely to be reflected in the proportions of participants in each subgroup. However, it is unlikely to strongly affect the data analyses and findings because these were internal comparisons and there was still sufficient variation within the sample, and sufficient numbers, to carry out the analyses.

All the key outcomes were available within our purposely designed development data set (the KAPS prospective cohort) but of those outcomes, only pain intensity was available in the independent validation data set (the feasibility and pilot trial in work package 3). As the tool appeared robust across outcomes in the cohort data set, it seems unlikely that it would show substantially different findings in the trial data set for outcomes of physical function (as measured by the PCS), but this cannot be empirically demonstrated.

The decision was made to examine the performance of the tool developed in the cohort study within the independent pilot trial data set; this took place after the publication of the KAPS cohort protocol paper but was agreed with the PSC. This decision was taken because the predictive performance of the initial refined tool was not as high as required. After examination of the tool in the external data set indicated sub-optimal performance, the requirement for candidate items to be potentially modifiable by treatment was removed because the team wanted to examine the potential for including non-treatment-modifiable factors in the tool such as 'pain duration', which was included in the final version to increase its predictive abilities.

A limitation of work package 1 as planned was that although we conducted the refinement and validation of the Keele STarT MSK Tool as a patient-completed set of questions, its use as completed by clinicians during a clinical consultation was not investigated.

## Conclusions

A new tool was refined and validated, the Keele STarT MSK Tool, with which to subgroup adults consulting with back, neck, shoulder, knee or multisite pain in primary care into those at low, medium and high risk of poor outcome. The tool is available on request from [www.keele.ac.uk/startmsk/startmskresearch/](http://www.keele.ac.uk/startmsk/startmskresearch/). It clearly and simply allocates patients to subgroups with distinct characteristics, different health-care usage and different prognoses, and its performance is acceptable upon independent validation. This study confirms that generic prognostic factors can be combined in one simple stratification tool and that this tool can be used to identify the risk of poor outcome (low, medium or high) in a wide range of patients consulting with different musculoskeletal pain presentations in primary care.

## Interrelationship with other parts of the programme

The Keele STarT MSK Tool was refined and validated in this work package, ready for use in work package 3 (feasibility and pilot RCT) as one of two components of the new stratified care intervention. The data from work package 3 was used to examine the external validity of the tool, prior to work package 4 (Keele STarT MSK trial). Participants in the work package 1 KAPS cohort formed the sampling frame for (1) invitation to participate in the qualitative interviews in work package 2 and (2) selection of example patient case vignettes to aid the development of consensus on matched treatments in work package 2. In addition, the recommended matched treatment options generated in work package 2 contributed to decisions about the cut-off points on the tool used to allocate patients to low-, medium- or high-risk subgroups.



## Work package 2

In parallel with the prospective cohort study (KAPS) in work package 1, work package 2 comprised three studies (an evidence synthesis, qualitative interviews and a consensus study) that together informed the final stage in this work package: the development and specification of the stratified care electronic template and GP training and support package to support the delivery of stratified primary care. These studies have been published in Babatunde *et al.*,<sup>19</sup> Saunders *et al.*<sup>20</sup> and Protheroe *et al.*<sup>21</sup>

### Research aims and objectives

The aim of this work package was to define and agree the matched treatment options for patients in each of the three risk subgroups and develop a training and support package for the delivery of stratified primary care.

The specific objectives were as follows:

- summarise current best evidence on available treatments for the five most common musculoskeletal pain presentations in primary care
- explore patients' and GPs' views on the acceptability of prognostic stratified care, and the anticipated barriers to and facilitators of its use in clinical practice
- agree, through expert consensus, the most appropriate matched treatment options that should be recommended for patients in each risk subgroup
- develop and specify a training and support package to support the delivery of stratified primary care.

### Methods and key findings for each study

#### Evidence synthesis

##### Methods

A systematic search and appraisal of evidence about the effectiveness of first-line available treatments for the five most common musculoskeletal pain presentations was conducted. The evidence synthesis followed a pyramidal approach using national clinical guidelines, policy documents, clinical evidence pathways and systematic reviews as a starting point. This was supplemented by systematic searches of bibliographic databases to identify and retrieve more recently published trials that had not yet been summarised in systematic reviews or guidelines, or identify where evidence gaps existed. Full details are available from the published paper.<sup>19</sup>

Currently available treatments for patients with musculoskeletal pain consulting in primary care that were considered include self-management advice and education, exercise therapy, manual therapy, pharmacological interventions (oral and topical analgesics and local joint injections), aids and devices, and other treatments [ultrasound, transcutaneous electrical nerve stimulation (TENS), laser, acupuncture and ice/hot packs]. Referral options for psychosocial interventions (cognitive-behavioural therapy and pain-coping skills) and surgery were also included. Comparison groups could include usual care, no intervention or other active interventions. Subsequently, data on study populations, interventions, and the outcomes of the intervention on patients' pain and function were extracted. Secondary outcomes such as psychological well-being/depression, catastrophising, quality of life, work-related outcomes (e.g. days off work and return to work), and cost of treatment were also extracted. The methodological quality of included systematic reviews was assessed using A MeaSurement Tool to Assess systematic Reviews (AMSTAR)<sup>22</sup> and the overall strength of evidence on the effectiveness of each treatment was rated using a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.<sup>23</sup>



## Key findings

In addition to clinical guidelines, policy and care pathway documents, a total of 146 systematic reviews (71 Cochrane and 75 non-Cochrane) met the selection criteria and were included. Methodological quality was lower for non-Cochrane reviews, because these are susceptible to publication bias ( $\approx 80\%$  of studies); full details are available in the published paper.<sup>19</sup> Study quality was not always incorporated into the evidence synthesis nor appropriately used to formulate conclusions in over a third of the included studies.

For most musculoskeletal pain presentations, non-pharmacological treatments, especially exercise therapy and psychosocial interventions, were found to lead to medium to large improvements for pain and function in the short and long term. Corticosteroid injections lead to short-term benefit for patients with knee and shoulder pain. Non-steroidal anti-inflammatory drugs (NSAIDs) and opioids reduced pain in the short-term but effect sizes were modest and the potential for adverse effects needs careful consideration. With the exception of acupuncture, which was found to be beneficial for relieving pain in the short-term, the clinical effectiveness of other treatments (ultrasound, TENS, laser and ice/hot packs) either alone or in combination was not substantiated by strong evidence for pain and function. The effectiveness of surgery as a first-line treatment option was not established. Evidence of long-term effectiveness of surgery was also limited except where directly indicated by specific serious pathology such as end-stage degenerative knee joint disease, or where persistent pain and functional limitations were refractory to conservative treatments.

Current evidence was equivocal on the optimal dose and application of most treatments. There was little evidence specifically about characteristics that might identify those patients most likely to respond to different treatments (e.g. pain severity or duration, previous pain episodes, or function).

In summary, the evidence synthesis showed that primary care patients with musculoskeletal pain can be managed effectively with non-pharmacological treatments, such as self-management advice, exercise therapy and psychosocial interventions, with short-term benefits only from pharmacological treatments.

## Qualitative focus groups and interviews

### Methods

Four focus groups and six one-on-one telephone interviews were conducted with GPs ( $n = 23$ ), and three focus groups with patients ( $n = 20$ ). Data were analysed thematically, and identified themes mapped onto the Theoretical Domains Framework (TDF),<sup>24</sup> a behaviour change theory that synthesises 112 psychological constructs determining behaviour change into 14 domains in order to identify barriers to and facilitators of behaviour change. Full details of methods and results are available in our published paper by Saunders *et al.*<sup>20</sup>

### Key findings

In brief, four key themes were identified.

#### ***Theme one: acceptability of clinical decision-making guided by stratified care***

Several GPs were receptive to the principles of stratifying patients, and felt that using the prognostic tool could complement their usual approach to clinical decision-making. Most GPs felt that having matched treatment options recommended to them was acceptable, as long as they felt these made clinical sense. Patients also perceived stratified care as being acceptable; expressing positive views about receiving appropriate management as a result of stratification.

Some GPs, however, expressed concerns relating to stratified care not adding significantly to GPs' clinical decision-making and potentially leading to reduced clinical autonomy. Patients, too, stressed the importance of GPs' clinical judgement and experience in treatment decision-making, and were concerned that some GPs may rely solely on the stratification tool results.



***Theme two: impact on the therapeutic relationship***

Some GPs felt that stratified care could enhance the therapeutic relationship with the patient because they perceived that patients would respond positively to them investing more time asking questions about their musculoskeletal problem, a view also evident among patients. The contrasting view was also presented by GPs, however, that the prognostic tool could impede the GP's efforts to build therapeutic rapport. GPs also anticipated potential conflict if the recommended matched treatment options were not in line with patients' treatment preferences, which was again echoed in the patient data.

***Theme three: embedding a prognostic approach within a biomedical model***

Some GPs expressed concern that an overreliance on prognostic stratified care may result in GPs becoming less proficient in diagnosing musculoskeletal conditions, leading to the fear that serious underlying pathologies could be missed. The importance of diagnosis was also highlighted by some patients, who felt that a diagnostic scan was the most effective route to the resolution of their symptoms. However, some GPs placed less emphasis on diagnosis and saw added value in the stratified care approach allowing them to give patients prognostic information in the face of diagnostic uncertainty.

***Theme four: practical issues in using stratified care***

Some GPs expressed concerns that completing a prognostic tool could detract from salient elements of the consultation and disrupt its flow. There was some scepticism from patients about whether or not stratified care would be used in practice, owing to the time-constraints of consultations. GPs also highlighted that matched treatment options must correspond to locally available services. However, some GPs identified past experience of using similar tools for other conditions as an enabler to the delivery of stratified care.

***Summary***

When looking across the themes, the theoretical domains of knowledge, skills, professional role and identity, environmental context and resources, and goals were identified as salient to GPs' and patients' perceptions of stratified care. It was found that for GPs and patients to see stratified care as useful, it must be perceived to add to existing clinical knowledge and skills, while not undermining GPs' and patients' identities and roles or their perceived goals of the consultation; particularly, not undermining GPs' clinical autonomy or disrupting therapeutic rapport. The need for the tool and matched treatment options to be integrated into the environmental context of consultations with minimal disruption was highlighted.

***Consensus on matched treatment options******Methods***

Three multidisciplinary consensus group meetings were conducted with clinicians between April and May 2015. In total, 20 participants attended at least one meeting (group 1,  $n = 18$ ; group 2,  $n = 16$ ; group 3,  $n = 12$ ). Nominal Group Technique (NGT)<sup>25</sup> was used, a systematic approach to building consensus using structured in-person meetings of stakeholders that follows a distinct set of stages. NGT participants were provided with summaries of best evidence about treatment effectiveness from the earlier evidence synthesis study in this work package, presented as novel 'evidence flowers'.<sup>26</sup> These included summary tables of evidence about treatment effectiveness for each of the five pain presentations (back, neck, shoulder, knee and multisite pain) and each patient risk subgroup (low, medium and high risk). Participants could suggest additional treatment options that they felt were appropriate but missing from the evidence synthesis. Participants were also provided with anonymised case vignettes drawn from KAPS cohort participants in work package 1. These vignettes included key characteristics of patients in each risk subgroup in order to inform discussions and consensus decision-making about matched treatment options.

For each potential treatment option, participants anonymously rated, on a Likert scale, their appropriateness for patients with each musculoskeletal pain presentation and in each risk subgroup. Mean scores were calculated, then treatments were organised in rank order and further discussed, before participants again rated the appropriateness of each treatment option. Treatments were included in the final list if they achieved reasonable consensus (a mean rating of > 3.5 out of 7 on a Likert scale). Full details of methods and results are available in the published paper.<sup>21</sup>

### Key findings

In total, 17 treatment options were recommended: four for patients at low risk, 10 for patients at medium risk and 15 for patients at high risk. As the risk of poor outcome increased, recommended treatment options increased in both number and intensity. For all five pain presentations, 'education and advice' and 'simple oral and topical pain medications' were recommended for all risk subgroups. For patients at low risk, across all five pain presentations 'review by primary care practitioner if not improving after 6 weeks' was also recommended. Treatment options for those at medium risk differed slightly across pain presentations, but all included: 'consider referral to physiotherapy' and 'consider referral to musculoskeletal interface clinic'. Treatment options for patients at high risk also varied by pain presentation. Some of the same options were included as for patients at medium risk, and additional options included 'opioid medication' and 'consider referral to expert patient programme' (across all pain presentations), and 'consider referral for surgical opinion' (for back, neck, shoulder and knee pain). 'Consider referral to rheumatology' was agreed for patients at medium and high risk with multisite pain.<sup>21</sup> These recommended matched treatments were summarised in table format ready to be incorporated into the electronic template and training/support package in the final stage of this work package.

### *Development and specification of an electronic template and training/support package*

#### Methods

Using the results of the previous three studies in this work package, a training and support package to support the delivery of stratified care by GPs was developed and specified. This drew on previous evidence showing that clinical decision support systems are most effective when combined with education for the professionals using them and that their perceived usefulness is a key factor driving engagement and acceptance by clinicians.<sup>27</sup>

One of the barriers to the use of stratified care identified in the qualitative focus groups and interviews was the time taken to include it within short primary care consultations. To make it as easy as possible for GPs to deliver stratified care, an electronic platform or electronic template was developed within the EMIS Web (EMIS Health, Leeds, UK) clinical electronic health record (EHR). This included both components of the stratified care intervention (the Keele STarT MSK Tool and the recommended matched treatments). EMIS allows bespoke protocols and data entry templates to be designed then implemented in target general practices. A version of the tool for use during face-to-face consultations was developed and embedded within the EMIS system, which triggered automatically on entering a relevant musculoskeletal pain diagnosis or symptom into the patient's EHR, and asked the GP to complete the Keele STarT MSK Tool. This led to the automatic calculation of the patient's risk score, classification into one of the three risk subgroups (low, medium or high risk of poor outcome), and the recommended matched treatment options. In addition, to support the delivery of 'education and advice' for all patient risk subgroups, integrated self-management information resources were embedded into the electronic template which could be easily printed to be shared with the patient.

The electronic template was developed to meet both the needs of the research and the requirements of the user, that is to complement the consultation, be easy to use in the time-pressured environment of brief consultations and to record key clinical information enabling assessment of intervention fidelity

for the trials in later work packages of the programme. GPs were involved in the development of the template from the initial stages through to user testing. For user testing GPs were sampled to include new, inexperienced, experienced and GPs, and those unfamiliar with EMIS Web, and invited to evaluate a draft version of the template. GPs provided feedback on ease of use, time taken, layout, format and text descriptors, leading to refinement of the template for use within face-to-face consultations in later work packages. Further details showing screenshots of the EMIS template are given in [Appendix 2](#). Delivering stratified care within a musculoskeletal pain consultation required the GP to engage with the EHR and trigger the stratified care template, to use the Keele STarT MSK Tool and to share the recommended matched treatment options with the patient and agree on treatment(s). Informed by the findings of the qualitative focus groups and interviews earlier in this work package, we developed a training and support package for GPs, drawing on well-recognised adult learning principles.<sup>28-30</sup> The training and support package aimed to address GPs' beliefs about the validity, value and feasibility of the stratified care approach and ensure they had the skills required to deliver this in practice. This included discussion among GPs about how they consult and make treatment decisions, introducing the principles of stratified care and how it differs from usual care, and an opportunity to try using the stratified care template and reflect on its use. Given that performance monitoring and feedback are important elements for encouraging behaviour change, the training and support package also included a plan for data collection and feedback at the individual GP, practice and trial arm level during GP practice review visits in later work packages.

### Key findings

A bespoke electronic template was designed, specified, user-tested and amended, ready to support GPs to deliver stratified care (i.e. to use both the Keele STarT MSK Tool and recommended matched treatment options). In addition, a training and support package for use with GPs participating in the trials in later work packages was developed. Full details of the content of the GP training package are given in [Appendix 2, Table 4](#). It was initially developed to be delivered in two approximately 1-hour sessions with GPs in those practices randomly allocated to the stratified care arm of the feasibility and pilot trial in work package 3, and then further refined for use in the main trial in work package 4.

### Alterations to initial plans and work package limitations

We conducted more focus groups with patients than planned (four rather than two) to ensure that larger numbers of patients' views were included. Based on pragmatic considerations, only one physiotherapist focus group was conducted (rather than the two planned). In the original plans for the consensus group study, 70% or more was the intended cut-off point to recommend treatment options, but this was reduced to 50% or more given the challenge in gaining higher levels of consensus with our highly multidisciplinary group of clinicians. We may have reached higher levels of consensus about recommended matched treatments, or a smaller set of recommended matched treatment options, had we included a more homogenous group of clinicians.

In the evidence synthesis, despite an extensive search, we found a paucity of evidence about treatments for multisite musculoskeletal pain. The control interventions were often not well described, and it is possible this led to overestimation of the effectiveness of some treatments in some included studies. In the qualitative study, patients' and clinicians' views were sought prior to finalisation of the stratification tool and recommended matched treatments. Discussions were deliberately based on the general principles of prognostic stratified care so that the views of patients and clinicians could inform the specific stratified care approach. Experiences of clinicians delivering, and patients receiving, the stratified care intervention in this programme were sought in work packages 3 and 4.

## **Interrelationship with other parts of the programme**

The qualitative focus group and interview study and the consensus group study both drew on participants from the KAPS cohort in work package 1. The agreed recommended matched treatment options from this work package informed the cut-off points of the Keele STarT MSK Tool in work package 1 and along with the electronic template and GP training/support package, were tested for feasibility of delivery in work package 3 (feasibility and pilot trial).

# Work package 3

## Research aims and objectives

The aims of this work package were to test the feasibility of a future main cluster RCT to compare stratified care with usual primary care, and to test the feasibility of delivery of the stratified care approach.

The four specific objectives were as follows:

1. estimate participant recruitment and follow-up rates for the main cluster RCT
2. examine evidence of selection bias between trial arms and between participants and non-participants
3. assess GP fidelity to the stratified care intervention (i.e. use of the stratification tool and matched treatment options)
4. conduct secondary descriptive analyses of GP decision-making and patient outcomes.

## Methods

Full details of methods and results have been published in two papers: Hill *et al.*<sup>18</sup> and Saunders *et al.*<sup>31</sup> The design was a pragmatic pilot, parallel two-arm (stratified vs. non-stratified care), cluster RCT in eight UK GP practices (four allocated to the intervention and four as control). Practices were randomised with stratification by practice size, and the trial statistician and outcome data collectors were blind to allocation. Participants were adult consulters with one of the five most common musculoskeletal pain presentations (back, neck, shoulder, knee or multisite pain) and without indicators of serious pathologies, urgent medical needs or vulnerabilities. Potentially eligible patient records were electronically tagged following consultation at participating practices and individual patients were sent postal invitations to participate in data collection. The target sample was 500 participants.

Delivery of the stratified care intervention by GPs was supported by the bespoke EHR template which included the Keele STaRT MSK Tool validated in work package 1 to stratify patients into low, medium or high risk of poor outcome and the 17 recommended matched treatment options agreed in work package 2. Patients at low risk were matched to options that supported self-management, including over-the-counter (OTC) medication, and unnecessary investigations or referrals were discouraged. For those at medium risk, in addition to the low-risk treatment options, recommendations included referral to conservative non-pharmacological treatments (e.g. physiotherapist-led exercise therapy) and workplace assessment and advice. For those at high risk, in addition to the low- and medium-risk options, recommendations included referral for corticosteroid injections, specialist clinical services (including rheumatology, orthopaedics and pain clinics) and opioid medications. Full details of the matched treatment options are described in [Appendix 3, Box 2](#), and also in figure 2 of our pilot trial publication by Hill *et al.*<sup>18</sup>

General practitioners in the four practices randomised to deliver stratified care were invited to participate in the training/support package developed in work package 2, facilitated by an experienced GP trainer and the trial lead. The sessions lasted a total of 3–4 hours at each practice and covered the rationale for stratified care, how it differs from usual care, and familiarisation with the EHR template and its fit within the flow of a musculoskeletal consultation. The sessions also provided an opportunity for discussion and questions or concerns to be addressed. GPs also received a training update halfway through their practice's recruitment period at which feedback data were shared about individual GP

intervention fidelity, with peer-to-peer comparisons and discussion. Further details about the training and GP peer-to-peer comparisons are available in [Appendix 3, Tables 5 and 6](#).

A set of pre-defined feasibility success criteria were used to evaluate recruitment, follow-up rates, selection bias and GP intervention fidelity. To determine whether or not these success criteria were met, four sources of data were used:

1. the general practices' EHR participant identification screen, which captured point-of-consultation data in all 8 general practices, including patients' pain intensity and location (back, neck, shoulder, knee or multisite pain)
2. an initial and 6-month postal questionnaire which collected outcomes such as physical function, risk subgroup, overall musculoskeletal health, fear avoidance beliefs, patient-perceived reassurance (from their GP), health-related quality of life, satisfaction with care received, provision of written educational material from their GP, global rating of change in the musculoskeletal problem, employment characteristics, work absence and productivity and patient demographic descriptors (see our published paper by Hill *et al.*<sup>18</sup> for full details, and also see [Appendix 3, Table 7](#), for a summary of the participant self-reported measures)
3. monthly follow-up using either short message service (SMS) texts or brief paper questionnaires to capture pain intensity, musculoskeletal pain-related distress and self-efficacy
4. an anonymised general practice EHR audit to collect data to describe GP decision-making including prescriptions, referrals, imaging requests, sick certifications and repeat GP visits over a 6-month period following the index consultation when the stratified care template was first fired.

## Nested qualitative study

Stimulated recall interviews were conducted with patients and GPs in the stratified care arm of the feasibility and pilot trial ( $n = 10$  patients,  $n = 10$  GPs), prompted by consultation recordings, in order to explore the feasibility of delivering stratified care in consultations. Data were analysed thematically and mapped onto the capability, opportunity, motivation and behaviour (COM-B) behaviour change model, exploring the capability, opportunity and motivation GPs and patients had to engage with stratified care (see our published paper by Saunders *et al.*<sup>30</sup> for full details of the methods and results of the nested qualitative study).

## Key findings

Participants were recruited between October 2016 and May 2017 from eight general practices, during which GPs screened 3063 patients (stratified care  $n = 1591$ , usual care  $n = 1472$ ) and completed the bespoke EHR template with 1237 eligible patients (stratified care  $n = 513$ , usual care  $n = 724$ ). A total of 524 participants (42% of those who received the bespoke EHR template) consented to providing questionnaire outcome data (stratified care  $n = 231$ , usual care  $n = 293$ ) over 6 months' follow-up. Follow-up questionnaires and EHR audit data collection were completed by December 2017.

### **Objective 1: estimating participant recruitment and follow-up rates**

Recruitment took 28 weeks (target 12 weeks) and 91% of participants provided follow-up data (target > 75%). Anonymised EHR data were available for 1281 patients (529 from stratified care practices and 752 from usual care practices). Although the high follow-up rate exceeded the target, the slow recruitment rate was a key concern, suggesting that the main cluster RCT would struggle to recruit the numbers originally intended (target  $n = 3600$ ) within the available timeline. This led to the decision to revise the sample size for the main trial (for full details see the flowchart in our published paper by Hill *et al.*<sup>18</sup>).



### **Objective 2: evidence of selection bias**

Most participant characteristics (e.g. sex) were similar between the two arms of the pilot trial, although there were a few minor differences between participants in the trial compared with non-participants (e.g. participants were slightly older and from more deprived areas). Baseline values of the primary outcome measure intended for the main trial (pain intensity) was 6.22 [standard deviation (SD) 2.17,  $n = 230$ ] in the stratified care arm compared with 6.21 (SD 2.32,  $n = 293$ ) in the usual care arm, and the proportions of patients at low, medium and high risk of poor outcome were 31%, 55% and 14%, respectively, in the stratified care arm compared with 33%, 54% and 13%, respectively, in the usual care arm (for full details see table 1 in our published paper by Hill *et al.*).<sup>18</sup> Overall, the data suggested little evidence of selection bias and, therefore, no changes were required to identification or recruitment procedures for the main trial.

### **Objective 3: assessing general practitioner fidelity to the stratified care intervention**

The fidelity of participating GPs to both components of the stratified care intervention was assessed (i.e. use of the stratification tool and matching patients to recommended treatment options). The pre-specified target for use of the stratification tool was for 50% of eligible patients, but the bespoke EHR data showed that GPs actually used the tool for 40% of eligible patients. Key reasons for the lower level of fidelity to the use of the tool were identified in the nested qualitative study (see *Nested qualitative study key findings*). However, in the consultations in which GPs used the risk stratification tool, their fidelity to matching patients to the recommended treatment options was high, with 81% of patients at low risk, 89% of those at medium risk and 87% of those at high risk being correctly matched to a recommended treatment, as recorded using the bespoke EHR template.

### **Objective 4: descriptive analyses of general practitioner decision-making and patient self-reported outcomes**

Based on the anonymised EHR data, key descriptive differences in GP decision-making were identified when comparing stratified care with usual care practices. For example, 20% fewer patients were given a prescription for opioid medication (opioid medications were listed as a matched treatment option only for patients at high risk of poor outcome in the pilot trial) and 53% fewer patients were recorded as having musculoskeletal pain-related imaging in stratified care practices. In addition, referral to physiotherapy for patients at medium or high risk of poor outcome was recorded as occurring more frequently in stratified care than in usual care practices. By contrast, the numbers of corticosteroid injections, sick certifications and repeat musculoskeletal pain-related GP consultations over 6 months' follow-up were recorded as similar in stratified care and usual care practices.

Based on the participant self-reported data in the postal questionnaires, patients received more written self-management information from GPs in stratified care practices than usual care practices (71% compared with 17%). Descriptive data on patients' follow-up outcomes demonstrated that participant's mean 6-month pain intensity was 3.93 (SD 2.98) in stratified care practices and 4.18 (SD 2.88) in usual care practices, with a change in 6-month pain intensity from baseline of -2.6 (SD 3.1,  $n = 207$ ) in stratified care practices compared with -1.9 (SD 3.1,  $n = 266$ ) in usual care practices. Most other outcomes were similar at 6 months' follow-up (e.g. patient satisfaction with GP care) although there was less musculoskeletal pain-related time off-work in participants in stratified care practices (17.4%) than usual care practices (25.4%) (full details are available in table 4 of our published paper by Hill *et al.*).<sup>18</sup> We did not statistically compare participant outcomes between the two arms in the pilot trial.

### **Nested qualitative study key findings**

Qualitative interviews with GPs revealed that the main reasons for lower than anticipated fidelity in using the Keele STarT MSK Tool with eligible patients were GPs perceiving that using the tool increased their consultation workload; GPs preferring to only use the bespoke EHR template when musculoskeletal pain was the primary reason for the consultation; GPs being reluctant to use the tool when their clinics were running late or were very busy; and that often patients had left the consultation room before GPs used their EHR system to record their consultations. It was also noted that some

GPs rarely coded musculoskeletal pain consultations in the EMIS system and that others tended to use 'synonym' codes, which are a set of diagnostic codes that were not able to activate the bespoke EHR template, or caused it to activate for some non-musculoskeletal pain problems (e.g. chest pain). It was agreed that for the main trial the GP training package needed to include ways to mitigate these challenges.

Patients reported positive views, reporting that that stratified care enabled a more 'structured' consultation and felt that the tool items were useful in making GPs aware of patients' worries and concerns. However, the closed nature of the tool's items was seen as a barrier to opening up discussion during consultations. Both patients and GPs identified 'cumbersome' items that made it more difficult to use (i.e. 'capability' within the COM-B theoretical model). Their feedback suggested that four of the items needed modification to be less 'clunky and awkward' to ask patients. For example, item 4 asks: 'Do you have any other important health problems?' which was felt to confuse patients when asked by their own family doctor whom the patient expected to know their health problems well. Most GPs reported that the matched treatment options aided their clinical decision-making (i.e. 'motivation'), but identified some options that were not commonly available to them (e.g. occupational health/workplace assessment referral or supported self-management), not needed (e.g. review if not improving after 6 weeks) or that they felt were appropriate but missing from the recommendations (e.g. consider imaging). They also expressed concerns about potentially overloading physiotherapy services with referrals and sought reassurance that linked physiotherapy services had sufficient capacity. For full details of the nested qualitative study results, see our published paper by Saunders *et al.*<sup>31</sup>

## Alterations to initial plans and work package limitations

The learning from this work package led to important changes to the original plans, all of which were agreed with the PSC and funder, prior to the main Keele STarT MSK cluster trial in work package 4, which were as follows:

- This feasibility and pilot trial changed from the intended internal pilot phase to an external pilot trial, given it took twice as long as expected (28 weeks rather than 12) to recruit participants and only 2 of the 4 pre-specified success criteria were met.
- The main trial sample size was reduced (from  $n = 3600$  to  $n = 1200$ ) by focusing on the overall between-arm comparison rather than also powering the trial to detect between-arm differences at the level of patient risk subgroups (i.e. low, medium and high risk).
- The items within the Keele STarT MSK Tool were amended following further statistical analysis using the point-of-consultation data from the pilot trial to identify items that improved the tool's predictive validity, and a clinician-completed version (interview style rather than self-report style) was developed. A license to obtain both the self-report and clinician-completed versions of the tool is available on request at [www.keele.ac.uk/startmsk](http://www.keele.ac.uk/startmsk). The clinician-completed version of the tool is also available in [Appendix 3](#), [Figure 2](#).
- The number of recommended matched treatment options was reduced slightly from 17 in the pilot to 14 in the main trial and some options were recommended for additional patient subgroups (e.g. opioid medications). Following meetings with linked physiotherapy services, participating GPs were also reassured that these NHS physiotherapy services were informed about the trial and had capacity to receive referrals.

There were several research limitations that need to be highlighted. First, there is an inherent potential risk of bias from those involved in developing a new clinical tool being the researchers who test it in practice. In addition, although the qualitative research revealed useful insights about why GPs found the tool difficult to use and explored the perceptions of patients regarding the services and treatment options available to them, more information about the reasons for both of these elements would



have been useful. A further limitation is that there may have been a difference in the suitability of the approach according to the length of condition duration (i.e. acute vs. chronic pain) that we have not yet been able to fully explore.

## **Interrelationship with other parts of the programme**

This work package evaluated the feasibility of using the Keele STarT MSK Tool and matched treatment options that were agreed in work packages 1 and 2 and comprised the external feasibility and pilot trial that informed the main trial in work package 4. In addition, the data from this pilot trial were used for the external validation of the Keele STarT MSK Tool at the point of consultation.



# Work package 4

## Research aims and objectives

The aim of this work package was to determine whether or not stratified care (i.e. the use of the Keele STarT MSK Tool to subgroup patients and matching subgroups to recommended treatment options) leads to superior clinical effectiveness and cost-effectiveness compared with usual primary care in patients with one of the five most common musculoskeletal pain presentations (back, neck, shoulder, knee or multisite pain).

The four specific objectives were as follows:

1. determine the comparative clinical effectiveness of stratified care compared with usual primary care for patient outcomes
2. investigate GP fidelity to delivery of stratified care and the impact on clinical decision-making
3. undertake an economic evaluation of stratified care versus usual care
4. conduct a nested qualitative study to understand how stratified care was perceived and operationalised by clinicians and patients.

## Methods

Full details of the methods are presented in our publication by Hill *et al.*<sup>32</sup> The main trial was a pragmatic, parallel two-arm (stratified vs. non-stratified care), cluster RCT with a health economic analysis and mixed-methods process evaluation. The setting was UK primary care, comprising 24 general practices randomised (stratified by practice size) in a 1:1 ratio (12 per arm) and with the blinding of the trial statistician and outcome data collectors. Randomisation units were general practices and units of observation were adults consulting with one of the five most common musculoskeletal pain presentations without indicators of serious pathologies, urgent medical needs or vulnerabilities. Potentially eligible patient records were electronically tagged at participating practices and patients were sent postal invitations to participate in data collection. The target sample was 1200 participants with ~ 2500 musculoskeletal consultations available for anonymised medical record data comparisons.

Delivery of the stratified care intervention by GPs was supported by the bespoke EHR template, which included the clinician-completed version of the Keele STarT MSK Tool and the reduced set of recommended treatment options for patients at low, medium or high risk of poor outcome. Full details of the per-protocol matched treatment options for each patient subgroup are provided in [Appendix 4, Box 3](#), and in figure 3 of our main trial protocol paper by Hill *et al.*,<sup>32</sup> and were used to support secondary analysis based on treatment per-protocol in the main trial. Overall, the intervention sought to influence GP clinical behaviours in order to reduce pharmacological treatment, referral to imaging, and secondary care referrals, and increase non-pharmacological treatments such as self-management advice and referral to physiotherapy. In addition, when GPs made a referral to physiotherapy they were asked to provide additional information about the patient's tool scoring and risk subgroup. The GP training and support package was similar to that described for work package 3, except it was shortened to 2 hours. GPs received a 1-hour training-update after the first month of recruitment to share and discuss the first of their monthly feedback reports showing individual GP fidelity to the stratified care intervention, and engage in peer-to-peer comparisons.

## Data collection processes

### ***Objective 1: determine the comparative clinical effectiveness of stratified care compared with usual primary care for patient outcomes***

Patient-reported data were collected via postal questionnaires and text messages: an initial questionnaire (posted shortly after the index GP consultation), a monthly SMS text or brief questionnaire with 3 items (pain intensity, pain-related distress and self-efficacy) and a 6-month follow-up questionnaire. The primary outcome was mean overall pain intensity on a 0 to 10-point NRS, measured monthly over 6 months (individual monthly scores were secondary end points). Secondary outcomes included physical function, quality of life, patient satisfaction, musculoskeletal risk subgroup, and overall musculoskeletal health status. All items from both versions of the tool were collected in the main trial to provide additional data for external validity testing. See our published protocol paper by Hill *et al.*<sup>32</sup> for full questionnaire details.

### ***Objective 2: investigate general practitioner fidelity to delivery of stratified care and the impact on clinical decision-making***

An EHR data audit was conducted over a 6-month period in each participating general practice to assess (1) trial recruitment, (2) selection bias, (3) GP fidelity to the stratified care intervention and (4) the impact of the intervention on clinical decision-making. See our published protocol paper by Hill *et al.*<sup>32</sup> for full details of the trial EHR template and the variables collected in the EHR audit.

### ***Objective 3: undertake an economic evaluation of stratified care versus usual care***

The primary economic evaluation was performed from an NHS and personal social services (PSS) perspective, with secondary analysis from health-care and societal perspectives. Resource use data, productivity loss and changes in quality of life (EQ-5D-5L) required for the economic evaluation were collected from participants within the 6-month postal follow-up questionnaire. Unit costs (in 2019 Great British pounds) were obtained and used in accordance with standard sources and attached to resource use items.<sup>33-35</sup> Utility data were generated using EQ-5D-5L participant responses from baseline and 6-month follow-up questionnaires.

### ***Objective 4: conduct a nested qualitative study to understand how stratified care was perceived and operationalised by clinicians and patients***

Data collection for the nested qualitative study used semi-structured interviews for patients ( $n = 27$ ), and focus groups and telephone interviews for clinicians ( $n = 20$ ; 13 GPs, one first contact physiotherapist and six community physiotherapists) in the stratified care arm.

### ***Sample size and analysis***

The target sample size of 1200 participants provided 90% power to test the superiority of stratified care compared with usual care, based on 5% alpha (two-tailed) and a small 'effect size' [standardised mean difference (SMD)] of 0.2 points<sup>36</sup> in the primary outcome (NRS pain intensity), taking account of clustering within practices (intracluster correlation 0.1) and coefficient of variation of 0.65,<sup>37</sup> baseline-outcome and repeated measures correlations of 0.5 and 0.7, respectively,<sup>38</sup> and 20% loss to follow-up. The main analysis was by intention to treat (i.e. analysing participants according to randomisation). All statistical testing was at the two-sided 5% significance level. Trial findings are reported in accordance with the cluster trial CONSORT (Consolidated Standards of Reporting Trials) statement.<sup>39</sup> Further details on the minimal clinically important change (MCIC) used for the primary outcome (a change of > 1 point on the 0–10 Numerical Rating Scale), models and adjustments used, sensitivity analyses, exploratory subgroup analyses, secondary outcome analyses and methods used to explore selection bias are provided in the trial protocol by Hill *et al.*<sup>32</sup>

### ***Health economics evaluation***

Participant responses from EQ-5D-5L utility data were used to estimate quality-adjusted life-years (QALYs) for every participant, using the area under the utility curve approach, assuming linear

interpolation between the utility measurements.<sup>40</sup> A multilevel modelling statistical approach, taking into consideration clustering in cost and effect data and multiple imputation of missing data, was adopted to estimate the incremental cost-per-QALY gained for stratified care compared with usual care.<sup>41</sup> Further methodological details, including sensitivity analyses relating to the uncertainty of the economic evaluation outcomes and prespecified subgroup analyses, are provided in [Appendix 5](#). Trial findings are reported in accordance with the Consolidated Health Economic Evaluation Reporting Standards (CHEERS).<sup>42</sup>

### ***Nested qualitative study***

Data were analysed thematically, and identified themes mapped onto the TDF (as outlined in *Work package 2, Qualitative focus groups and interviews, Methods*) and Normalisation Process Theory (NPT),<sup>43</sup> a framework for understanding why some health-care interventions are accepted and more successfully embedded in routine clinical practice than others through exploring four components: coherence (whether or not an intervention ‘makes sense’ within existing ways of working); cognitive participation (whether or not individuals are prepared to invest time and energy to engage with the intervention); collective action (work done to enable the intervention to be adopted); and reflexive monitoring (individuals’ appraisal of its benefits and costs). Identified themes were mapped onto the TDF and NPT frameworks.

## **Key findings**

### ***Objective 1: determine the comparative clinical effectiveness of stratified care compared with usual primary care for patient outcomes***

Twenty-four general practices [12 per arm; total adult practice size of 104 GPs and 185,088 patients (96,397 allocated to stratified care, 88,691 receiving usual care)] participated from the West Midlands region of England. A total of 1211 (49%) patients consented to data collection (534 allocated to stratified care and 677 receiving usual care) and responded to the initial questionnaire within 30 days. Mean age was 60 years (range 18–95 years) and 58.9% of patients were female. Mean pain intensity on a 0–10-point NRS scale at the point-of-consultation was 6.73 points (6.77 points in the stratified care arm and 6.70 points in the usual care arm). Over 6 months of follow-up, 1178 (97%) participants provided at least one pain intensity response [515 (96%) in the stratified care arm and 663 (98%) in the usual care arm] and 80.9% responded to the follow-up questionnaire at 6 months (77.9% in the stratified care arm and 83.3% in the usual care arm). Full details of patient recruitment and follow-up through the trial are described in [Appendix 6, Figure 4](#).

There were few differences between the characteristics of patients who agreed to participate in data collection and those that did not (see [Appendix 6, Table 8](#)). The population characteristics of general practices randomised to each arm were similar (see [Appendix 6, Table 9](#)). Most characteristics of participants in each arm of the trial were similar, including the proportions of patients in each risk subgroup, which for stratified care and usual care, respectively, were 19.5% and 20.1% for low risk, 47.3% and 45.6% for medium risk, and 33.2% and 34.3% for high risk. Although there were statistically significant differences in age, employment status and pain site (see [Appendix 6, Table 18](#)), overall, these differences were modest and did not amount to a concerning signal of selection bias.

In the primary analysis there were no statistically significant differences in pain intensity over months 1–6 between the two arms, with mean values of 4.4 points (SD 2.3 points) for stratified care and 4.6 points (SD 2.4 points) for usual care (see [Appendix 6, Figure 5](#) and [Table 19](#)). The adjusted mean difference was –0.16 points [95% confidence interval (CI) –0.65 to 0.34 points;  $p = 0.535$ ], translating to a SMD (effect size) of –0.08 (95% CI –0.33 to 0.17) (see [Appendix 6, Table 19](#)). Mean differences in pain intensity were consistently greater in the last 3 months than the first 3 months; however, the average mean difference between the stratified care arm and usual care arm over months 4–6 was not significant

(0.33 points, 95% CI -0.84 to 0.19 points;  $p = 0.211$ ). Most sensitivity analyses showed no statistically significant between-arm differences, despite showing consistent slightly favourable results for stratified care (see [Appendix 6, Figures 10–17](#)). Analysis of the MCIC of a > 1-point difference in pain outcome gave an overall odds ratio (OR) of 1.26 (95% CI 0.89 to 1.78;  $p = 0.188$ ); however, there was a significant OR (1.54, 95% CI 1.09 to 2.17;  $p = 0.013$ ) for the 4–6 months comparison (see [Appendix 6, Table 20](#)).

Subgroup analyses showed some between-arm mean differences with a greater difference (although statistically non-significant) in patients at high risk than those at low risk, and in those with shoulder and knee pain than those with neck and back pain (see [Appendix 6, Table 21–23](#)). There were no statistically significant differences in secondary clinical outcomes at 6 months except for a significantly greater improvement in shoulder pain and function and a higher satisfaction with care in the stratified care arm than the usual care arm (see [Appendix 6, Table 24](#)).

### **Objective 2: investigate general practitioner fidelity to delivery of stratified care and the impact on clinical decision-making**

The trial EHR template activated in 11,412 patients after April 2019 (with 197 patients having the tool used during GP consultations between May 2018 and June 2019) and was completed in 2494 potentially eligible patients [1056 (18%) from stratified care practices and 1438 (27%) from usual care practices]. GP fidelity to the first component of the stratified care intervention was low, with the clinician version of the Keele STarT MSK Tool being completed in only 29.76% (1056/4742) of eligible consultations in stratified care practices. However, in those patients for whom the tool was completed, appropriate recommended matched treatment options were selected (i.e. recorded on the bespoke EHR template) for over three-quarters of patients (77.2% for patients at low risk, 77.8% for patients at medium risk and 76.7% for patients at high risk). Full details are presented in [Appendix 6, Tables 25–28](#). It should be noted that selection of matched treatments as indicated on the trial EHR template tended to be favourably reported when compared with actual treatment behaviour identified from the anonymised EHR audit.

The anonymised EHR audit data were available for 2494 patients across all practices and demonstrated several important impacts from stratified care on GP treatment decision-making (see [Appendix 6, Table 29](#)). Compared with usual care, there were significantly fewer prescriptions for muscle relaxant medications [incident rate ratio (IRR) 0.48, 95% CI 0.25 to 0.92], and significantly more patients provided with written self-management information (58% vs. 26%) and referrals to physiotherapy (OR 12.7, 95% CI 5.47 to 29.7) in the stratified care arm. For patients at low risk of poor outcome, GPs requested less musculoskeletal imaging (IRR 0.11, 95% CI 0.01 to 0.79) and fewer sick certifications (IRR 0.12, 95% CI 0.01 to 0.97), yet provided more prescriptions for simple analgesic medication (IRR 2.80, 95% CI 1.33 to 5.88) and more referrals to physiotherapy (OR 5.75, 95% CI 2.1 to 15.8) in the stratified care arm than the usual care arm. There were also fewer repeat GP consultations over 6 months in the stratified care arm (IRR 0.60, 95% CI 0.36 to 0.98). For patients at medium risk, more were offered opioid medication (IRR 8.86, 95% CI 1.74 to 45.1) and physiotherapy referral (OR 17.2, 95% CI 7.08 to 41.8) in the stratified care arm than the usual care arm. Significantly more patients at high risk of poor outcome were referred to physiotherapy in the stratified care arm (OR 29.7, 95% CI 9.9 to 89.2).

### **Objective 3: undertake an economic evaluation of stratified care compared with usual care**

The full results of the economic evaluation are planned to be published separately. In summary, the costs of care were very similar between the two arms of the trial: the mean cost of stratified care was £356.36 (SD £864.01) compared with £343.44 (SD £942.92) for usual care. The adjusted incremental cost of stratified care compared with usual care over the 6 months was £6.85 (95% CI -£107.82 to £121.54), with incremental QALYs of 0.0041 (95% CI -0.0013 to 0.0094), representing a net QALY gain. Stratified care was associated with a cost-per-QALY gain of £1670. At a willingness-to-pay threshold of £20,000 per QALY ( $\lambda$ ), the incremental net monetary benefit was £132 and the probability of stratified care being cost-effective was approximately 73% (see [Appendix 3, Figure 2](#)). Sensitivity analyses from alternative perspectives showed that stratified care was associated with minimal cost savings. Stratified

care was shown to be potentially cost-effective compared with usual care using commonly applied willingness-to-pay threshold values of £20,000 to £30,000 per QALY gained from NHS and health-care perspectives. However, the favourable incremental cost-effectiveness ratio is a result of very small differences in costs and QALY outcomes, suggesting caution in the interpretation of this result.

***Objective 4: conduct a nested qualitative study to understand how stratified care was perceived and operationalised by clinicians and patients***

Three main themes emerged from the qualitative study.

**Theme 1: role of stratified care in informing clinical decision-making**

During the interviews several GPs in stratified care practices reported a shift in their usual clinical behaviours towards giving greater attention to psychosocial issues and taking a more functional approach towards patients with musculoskeletal pain, particularly for shoulder and knee pain patients with whom they felt they had previously tended to adopt a more traditional biomechanical/diagnostic approach. Risk stratification was felt to be useful for the early identification of patients who may develop chronic pain problems and later potentially become long-term opioid users. GPs felt that stratified care helped them to overcome some of the barriers to adopting a functional approach, such as patient expectations of being referred for imaging, by helping to facilitate the negotiation with patients about treatment options. By contrast, most patients reported being unaware that a stratification tool was being used in their GP consultation for the purpose of treatment-matching but, when shown the tool, felt that the items added value (e.g. they felt that the questions about mood would facilitate a holistic approach). For some GPs stratified care was felt to have less influence on their behaviour, particularly those that reported completing the tool after they had already decided on a patient's management plan. A few GPs also perceived less value from stratified care because they interpreted the risk subgroup status as referring to longer-term risk of chronicity over several years, rather than being linked to their current treatment decision-making.

**Theme 2: perceived influence of stratified care on patient management**

Patients generally reported satisfaction with their clinical management and reported surprise regarding some options that they did not associate with usual GP care (e.g. social prescribing of lifestyle interventions). Patients who had experienced a reduction in pain attributed this partly to increased confidence in, and knowledge about, self-management based on the advice given by the GP or physiotherapist. Those referred to physiotherapy were positive about the time from referral to treatment. Interestingly, some patients and GPs perceived that patients would receive physiotherapy more quickly as part of the trial, despite this not being the case (waiting times were unaffected by the trial processes). Physiotherapists reported patients from stratified care practices being more motivated to engage with physiotherapy than those from usual care practices.

**Theme 3: implications of stratified care for interdisciplinary working**

Physiotherapists reported finding the additional Keele STarT MSK Tool information about referred patients to be useful in alerting them to areas they might need to explore, particularly with patients at high risk of poor outcome. GPs and physiotherapists saw added value in closer interprofessional communication for patients at high risk, a view echoed by patients, who felt that improvements in health-care professional communication about their care were needed.

**Summary**

When looking across the themes, the theoretical domains of skills, professional role and identity, goals, intentions and decision-processes (decision-making) were identified as salient to patients' and clinicians' experiences. GPs' behaviour change appeared to involve shifting GPs' perceived goals and intentions regarding the consultation towards a functional, more biopsychosocial-informed approach. This indicated a change in their professional role and identity, with less emphasis on the GP's role as a diagnostician. Stratified care also supplemented GPs' skills to help them identify patients at risk of poor outcome and informed their decision-making processes by aiding treatment negotiation with patients.



Stratified care had a strong level of 'coherence' (i.e. made sense) within general practice that encouraged 'cognitive participation', that is a willingness to put in time/energy to engage with it. However, some barriers to using the Keele STarT MSK tool remained that were also identified in the feasibility and pilot trial, such as GPs preferring to use it only when musculoskeletal pain was the primary reason for the consultation and also being less inclined to use it if there was an existing management plan in place for the patient. GPs reported that stratified care had a strong level of 'coherence' (i.e. made sense) within their practice, which encouraged their 'cognitive participation'. In terms of 'reflexive monitoring' (i.e. assessing the cost or benefits), GPs, physiotherapists and patients expressed positive views about stratified care and suggested a willingness to undertake 'collective action' to adopt stratified care in the future.

## Alterations to initial plans and work package limitations

There were no further alterations to the plans for the main cluster RCT during work package 4 (see *Work package 3, Alterations to initial plans and work package limitations* for the changes made to the main trial following the pilot RCT). It was not possible to limit the automatic firing of the stratified care EHR template within the EMIS system as much as would be desirable in the consultation. This meant it fired when the patient's musculoskeletal treatment plan was not necessarily being considered in a consultation (e.g. when recording medication-only changes), or where musculoskeletal pain was not the primary reason for the consultation (e.g. in cases where musculoskeletal pain was a comorbidity to a different main clinical issue). In addition, to reduce GP burden, the stratified care EHR template only fired once per patient and not during their subsequent consultations. This meant that the opportunity to change GPs' clinical decision-making for an individual patient according to stratified care was limited to the one consultation in which the stratified care template 'fired'.

There are several limitations that are worth highlighting, particularly as the trial findings contrast with those of our STarT Back stratified care trial<sup>6</sup> conducted among patients with low back pain. Possible explanations for this include the low GP fidelity to using the risk tool (29.76%) and the potential lack of effectiveness of the matched treatments options used. We saw no evidence that the low levels of GP tool use were due to difficulties in accessing or using the medical record interface (probably because our intervention template was embedded into their existing record system). We think it more likely that the low levels of use were due to the timing of template trigger, which often occurred after patients had left the room, and also due to the current time-pressured context of UK primary care. This was evidenced by GPs stating they 'do not have time', or 'patient was not present' in 49.80% of potentially suitable musculoskeletal consultations. Following the pilot trial findings (where GP tool fidelity was 40% of eligible consultations) we revised our GP fidelity expectations to 25% of musculoskeletal-coded consultations, primarily because GPs reported the intervention was only appropriate where musculoskeletal pain was the primary problem for the visit, which may be as low as only 50% of all GP musculoskeletal-coded consultations.

In relation to the specific limitations of the matched treatment options, it should be noted that in this trial we did not train physiotherapists in biopsychosocial approaches or optimise the musculoskeletal clinical pathways to deliver improved risk-matched treatments. In our previous STarT Back trial<sup>6</sup> we included a 3-day physiotherapy training programme for physiotherapists delivering the medium risk treatment and a 6-day training programme (and ongoing regular mentoring) on psychologically-informed physiotherapy skills for those treating patients at high risk of poor outcome. In the STarT Back trial,<sup>6</sup> mean change in Roland Morris Disability Questionnaire (RMDQ) score in the intervention arm at 4 months was 4.7 versus 3.0 points in the control arm. In the Keele STarT MSK trial,<sup>6</sup> back pain function improved by much less at 6 months (RMDQ score: intervention = 3.5 points versus control = 3.1 points). To date, the authors are not aware of any other stratified care trials for low back pain providing as intense a training and mentoring programme as the STarT Back Trial.<sup>6</sup> Another consideration is that other successful primary care risk-prediction tools (such as QRISK, a risk prediction algorithm to help enable



doctors to identify patients at most risk of heart disease and stroke,<sup>44</sup> for estimating cardiovascular risk) have the benefit of highly effective pharmaceutical treatment for those identified at increased risk. Therefore, it seems that a key challenge for future trials of musculoskeletal risk-based stratified care will be to provide more effective treatments for those at increased risk of poor outcome.

## Conclusions

The results of this cluster RCT demonstrated that the model of stratified care tested in this programme and delivered by GPs for patients consulting with one of the five most common musculoskeletal pain presentations did not lead to superior clinical outcomes compared with usual non-stratified care. However, data on clinical decision-making showed that the stratified care EHR template and clinician training led to positive changes in GP clinical decision-making, with fewer prescriptions for muscle relaxants, better provision of written self-management information and many more referrals to physiotherapy. The health economic analysis showed that although the costs of stratified care were very similar to usual care, there were small benefits from the stratified care intervention. Although the findings suggest that risk-based stratified care is potentially a cost-effective use of health-care resources when applying conventional rules of cost-effectiveness, the minimal differences suggest caution in the interpretation of this result. Qualitative data from GPs, physiotherapists and patients involved in the main trial about their experiences of stratified care were generally positive.

## Interrelationship with other parts of the programme

This work package followed work packages 1–3 and tested the clinical effectiveness and cost-effectiveness of the prognostic stratified care approach for patients consulting in primary care with one of the five most common musculoskeletal pain presentations in primary care.



# Patient and public involvement and engagement

Our research team involved PPIE members from the initial development phases of the programme through to the planning and conduct of each work package, and in the interpretation of results. Our PPIE co-applicant (Mr John Murphy) helped develop the plans for the programme prior to funding. Following funding approval John Murphy provided PPIE on the PSC, advising on all aspects of the programme, and participated in other PPIE meetings, helped develop participant information and documentation, and advised the lead for the qualitative studies on key questions to explore in patient interviews. Further PPIE was provided through a second patient representative (Mrs Angie Emery) joining the PSC. Together, these two PPIE representatives were involved in discussions of emerging results throughout the programme, providing interpretations from the patient perspective and contributing to discussions about key changes to the programme over time. One of these representatives moved home during the programme, but geographical differences were overcome through the use of teleconferencing, which worked well. Additional programme-specific PPIE was supported through the Keele University Research User Group (RUG), the members of which have or care for people with a range of musculoskeletal problems. Activities in each work package are summarised in the following sections.

## Work package 1

During development of the plans for this work package, a PPIE meeting was held to discuss and feedback on patient-facing materials (e.g. patient information leaflets, consent forms and questionnaires). PPIE members provided feedback on the medical record review procedures as well as the novel patient reminder systems (via text and e-mail reminders). Members reviewed the draft stratification tool (the Keele STarT MSK Tool) for face validity, and gave important feedback on the potential replacement candidate items for the revised tool. Once data collection and analysis were complete, a PPIE meeting (attended by 9 patients/carers) discussed the findings and suggested how best to share results with patients and general practices. This resulted in results being displayed on posters in the practices rather than distributed through leaflets. PPIE members suggested simplification of the content shared with participants, with an emphasis on thanking them for their involvement rather than sharing complex diagrams of the results. These suggestions were implemented.

## Work package 2

The evidence about the effectiveness of available treatments for the five most common musculoskeletal pain presentations was summarised and presented to stakeholders, including our programme PPIE members. A PPIE member was invited to the meeting of the wider stakeholder group for work package 2 to assist the team with decisions about which existing patient-facing websites and patient-written self-management information leaflets GPs should use as part of the suite of matched treatment options.

## Work package 3

Our PPIE co-applicant (Mr John Murphy) attended a protocol development day for work packages 3 (feasibility and pilot trial) and 4 (main trial) to advise on the protocol from a patient's perspective. Subsequently, a PPIE day (attended by 12 PPIE members with musculoskeletal pain conditions) was held to review the protocol and assist with development of documentation for the trials. Those who attended

were later involved in follow-up activities via post and e-mail to review further amended documents, including documents supporting the qualitative research in the trial and the postal questionnaires that were sent to patients.

Once the feasibility and pilot trial was under way, nine patients/carers took part in a further PPIE meeting to discuss progress and the emerging plans for the main trial, including the patient-facing documentation. PPIE members made suggestions on ways to improve the uptake of patient interviews for the qualitative component. At the end of the pilot trial, seven PPIE members met to discuss plans for the main trial, provided valuable feedback on the pilot trial results and helped finalise the wording of the clinician-completed version of the Keele STarT MSK Tool, pointing out that asking about the last 2 weeks for patients with a long-term condition may not seem relevant and suggesting that the GP needed to provide context to help the patient understand why they were asking the questions on the tool. They also suggested re-ordering the questions to improve flow and understanding. Amendments were made as a result and the decision was made to use the clinician version of the tool at the point-of-consultation in the main trial.

The qualitative researchers met with four PPIE members to discuss emerging themes from the qualitative research in the pilot trial, as well as data extracts from patient interviews. PPIE members provided their interpretation of the key issues in the data. This informed analysis and fed into the resulting paper from the pilot trial.

## Work package 4

The study PPIE members met to review the near-final versions of the main trial patient-facing documentation (e.g. the invitation letter, information leaflet, questionnaire and consent form) prior to submission for ethics approval. They suggested a number of improvements to the documents that were implemented, including improved consistency of the wording used, use of bullet points rather than lengthy sentences, use of an A4 booklet for the information leaflet and more use of colour. Plans for the main trial were also adapted based on discussions held during the meeting. The patient prize draw used in the pilot trial to incentivise patients to return their questionnaires was not taken forward into the main trial plans. PPIE members were very supportive of the plan to ask GPs to seek verbal consent from patients in the consultation to share their contact details with the research team in order to send out information about the study.

A further PPIE meeting was held to discuss the main trial with the PPIE group and to share available quantitative and qualitative findings in order to seek PPIE members' views on the interpretation of the data. PPIE members contributed by reviewing some of the qualitative research data from the main trial, reading through direct quotes from trial participants in pairs and noting their thoughts before joining a wider group discussion about the interpretations of the findings. PPIE members helped to influence the plans for patient-focused dissemination of the results from the main trial. Beyond this programme, our PPIE members continue to be involved in other research studies (e.g. as steering group members) and have contributed to other bids and funded research led by Keele University, sustaining and further developing PPIE partnership working and supported by the broader RUG at Keele.

## Overall programme conclusions

We set out to refine and validate a new stratification tool with which to identify the risk of poor outcome in patients consulting in primary care with one of the five most common musculoskeletal pain presentations (back, neck, shoulder, knee and multisite pain); agree which available treatments in the NHS primary care context should be recommended as matched treatment options for patients at low, medium and high risk of poor outcome; test the feasibility of delivering this prognostic model of stratified care and the feasibility of conducting a randomised trial; and finally conduct a large cluster randomised trial to compare stratified care with usual primary care on patient outcomes, cost-effectiveness and clinical decision-making. We involved a wide range of health-care professionals who care for patients with musculoskeletal pain and many patients with musculoskeletal pain problems in four work packages to finalise and then test the stratified care approach. Their responses, views and experiences shaped the content of, and revisions to, the stratification tool [the Keele STaRT MSK Tool with its two versions: one to be completed by patients (the self-report version) and one to be asked by clinicians during the consultation (the clinician-completed version)] and the recommended matched treatments for patients in each risk subgroup.

We demonstrated that the stratification tool works well to identify patients in different risk subgroups who have different characteristics, different health-care use and associated costs, and different prognoses. The Keele STaRT MSK Tool is a valid tool with which to identify patients with different levels of risk of persistent pain, and therefore provides additional systematic information about an individual patient's prognosis that can help clinicians to direct patients to the most appropriate treatments. The approach of using one tool for this wide range of patients has the major benefit of simplicity for clinical practice, removing the complexity that would result from multiple, pain site-specific stratification tools.

The matched treatment options recommended for patients at low, medium and high risk of persistent pain were underpinned by an evidence synthesis including previous clinical guidelines, systematic reviews and recent randomised trials, combined with the consensus of a large and multidisciplinary group of health-care professionals. The planned delivery of stratified care was shaped by qualitative focus groups and interviews with GPs and patients to try to reduce the burden of additional time during consultation and to provide GPs with practical support through a training and support package that included the provision of a stratified care EHR template that fired during consultations.

Our feasibility and pilot randomised trial with 8 general practices and 524 patients showed that although we could recruit and retain patients, we needed twice as long as anticipated to recruit the target number. GP fidelity to matching subgroups of patients to recommended treatments was high, but they found it challenging to use the tool in many of the musculoskeletal pain consultations for which it was intended. Following amendments to the tool, treatment options and EHR template, our main cluster randomised trial was conducted with 24 general practices, achieving the target sample size of 1200 patients and high follow-up rates over 6 months.

The main trial results showed no significant differences, overall, in the primary patient outcome of pain intensity between stratified care and usual care, despite showing consistent slightly favourable results for stratified care and some subgroup analyses showing between-arm mean differences in patients at high risk, and in those with shoulder pain. The health economic evaluation showed that although the costs of care were very similar in the two arms of the trial, stratified care was associated with a small improvement in quality of life compared with usual care, resulting in a 73% probability of stratified care being cost-effective for the NHS. Anonymised EHR audit data for 2494 patients demonstrated significant differences in some aspects of GP clinical decision-making about treatments. Compared with usual care, there were significantly fewer prescriptions for muscle relaxant medications, and significantly more patients provided with written self-management information and referred to physiotherapy in the

stratified care arm. There were also several other significant differences in decision-making for patients in each of the risk subgroups.

The results of the main trial show that some improvements in clinical decision-making about the care of patients with common musculoskeletal pain can be achieved using an EHR template within the consultation that helps identify patients' risk of persistent pain and recommends matched treatment options for the GP and patient to consider. Four explanations might explain the lack of significant improvements at the level of patients' clinical outcomes.

First, in order to minimise the burden on GPs, the stratified care EHR template fired only once per patient in the trial, despite participants having, on average, 4 or 5 consultations over the period of 6 months. This was likely to have limited the stratified care intervention's ability to change GP behaviour to the one short consultation in which the template fired. The EHR audit data confirmed that where differences in clinical decision-making were identified, these were concentrated in the 7-day period from the index consultation, with few differences over the following 6 months.

Second, despite the changes made to the stratification tool and matched treatments as a result of the pilot trial, GPs continued to find it challenging to use the tool in many of their musculoskeletal pain consultations (use of the tool reduced from 40% in the pilot trial to 29.76% in the main trial). There was also a reduction in GP fidelity to offering matched treatments based on the patients' risk subgroup (down from over 80% in the pilot trial to just over 76% in the main trial). In the current pressurised context of primary care, GPs found it challenging to deliver this model of prognostic stratified care. There were also likely cumulative losses in fidelity, since what GPs in stratified care practices told us they intended to do regarding recommended treatments on the bespoke EHR template tended to overstate treatment fidelity when compared with the anonymised EHR data audit, meaning that it is possible that GPs ticked the recommended treatment options on the bespoke template but subsequently did not provide all of the treatment they had ticked on the electronic form.

Third, only treatment or referral options that were already available to GPs were included in this trial and there was no attempt to optimise or improve the effectiveness of those treatments. In particular, since referral to physiotherapy was a key recommended treatment for which we observed significant between-arm differences in favour of stratified care, the fact that we did not attempt to optimise the content of the physiotherapy treatments offered may help explain the trial findings.

Finally, our stratified care model, a tool with 10 items, may have been too complex, resulting in three patient subgroups for which a total of 14 treatment options were recommended for consideration. Although the recommended treatment options were underpinned by best evidence and the expert consensus of many clinicians, some were rarely used and a simpler model of stratified care may have been more successful.

These last two potential explanations (the need to optimise the content of treatments and the need to simplify the model of stratified care) may also explain the differences in results of this trial from our previous STaRT Back trial<sup>6</sup> in which we evidenced the clinical effectiveness and cost-effectiveness of a simpler model of stratified care for back pain with only three matched treatments. In addition, these three treatments were optimised and physiotherapists were trained in their delivery.

The findings of this trial contrast with our previous successful stratified care trial in low back pain in the UK.<sup>6</sup> We did not see evidence that low GP fidelity was due to difficulties in accessing/using the GP computer-based stratified care interface, possibly because the risk template was embedded into the GP record system. However, we therefore suspect low GP fidelity was due to the template triggering on entry of a diagnostic code, which often occurred after the patient had left the room. We also note that GPs felt the stratification tool added time to the consultation. These points are evidenced by GPs stating they 'do not have time', or 'patient was not present' in 49.80% of potentially

suitable musculoskeletal consultations. We would also note that GPs reported the intervention was only appropriate where musculoskeletal pain was the primary reason for the visit and they felt the tool often fired when musculoskeletal pain was presenting as a comorbid condition. The observed increase in the prescribing of short-term strong opioids among GPs in the intervention arm (20.3% in the intervention arm vs. 1.0% in the control arm) was a surprise and was the opposite of what we observed in our pilot trial, where opioid prescribing reduced. These differences are likely to relate to a change to our risk-matched treatment options; in the pilot trial opioids were only recommended for high risk patients, whereas in the main trial weak opioids were a recommended treatment option for medium risk patients as well. It should be noted that the small differences observed between arms started at around 3 months of follow-up, which is when we understand patients began receiving NHS physiotherapy owing to a 6–8-week waiting list at the time of the trial. In this trial we did not upskill physiotherapists or optimise clinical pathways to deliver risk-matched treatments. By contrast, the STarT Back trial<sup>6</sup> provided physiotherapists with a 3-day training programme for treating medium-risk patients and a 6-day training programme (and ongoing regular mentoring) on psychologically informed physiotherapy for high-risk patients. In the STarT Back trial, mean change in RMDQ score at 4 months was 4.7 points in the intervention arm versus 3.0 points in the control arm. In this trial, back pain function improved by far less at 6 months (RMDQ score of 3.5 points in the intervention arm vs. 3.1 points in the control arm). To our knowledge, no other stratified care trials for low back pain have provided as intense a training and mentoring programme as the STarT Back trial. Key challenges for future trials of musculoskeletal risk-based stratified care are, first, to find more feasible methods for stratifying patients in short consultations and, second, to provide more effective treatments for those at increased risk through appropriate workforce training and upskilling.

Particular programme strengths included a systematic approach to the refinement, validation and specification of our stratified care intervention; the large sample sizes of patients with musculoskeletal pain consulting in primary care in the cohort study and randomised trials; the development and use of a bespoke EHR template to support GPs to deliver stratified care; the additional use of anonymised routinely collected EHR data; and the participation of clinicians and patients throughout the programme.

## Implications for practice

- The Keele STarT MSK Tool is a valid tool with which to identify patients with different levels of risk of persistent pain, and therefore provides additional systematic information about an individual patient's prognosis that can help clinicians direct patients to the most appropriate treatments. The approach of using one tool for this wide range of patients has the major benefit of simplicity for clinical practice, removing the complexity that would result from multiple, pain site-specific stratification tools.
- Given the high prevalence of musculoskeletal pain and the variability in clinical decision-making in primary care, the positive impacts on some aspects of clinical decision-making we observed together with the finding that this new way of working does not negatively affect patients' outcomes and is likely to be cost neutral suggests that this model of stratified care may bring benefits for the NHS.
- Supporting GPs with bespoke computerised EHR templates that fire within the consultation with a patient and support clinical decision-making can help change specific behaviours in ways that reduce low-value care. Specifically, we showed that our EHR template increased the provision of written self-management information and referral for non-pharmacological treatments for patients with musculoskeletal pain.
- Key challenges to overcome, however, include how to support GPs in busy clinical consultations to use stratification tools with and offer recommended treatments to as many appropriate patients as possible, and how to improve primary care treatments in ways that lead to better patient outcomes.



## Recommendations for future research

- There are challenges in designing and testing stratified care interventions that are sufficiently potent to bring about change in the face of challenging clinical contexts and likely losses to fidelity. We recommend conducting careful feasibility work prior to testing stratified care interventions in definitive pragmatic trials.
- Our stratified care intervention had two components: the use of a stratification tool and then the matching of patient subgroups to recommended treatment options. The first of these worked well to identify patient subgroups and the tool has already been shared with over 1000 tool license requestees, leading to other research. We found that GPs struggled to incorporate use of the tool in their short consultations with patients, and losses in fidelity to the matched treatment options are key potential explanations for the trial results. Future research to test ways to better support clinicians to use stratification tools (including qualitative research) and to better match patient subgroups to appropriate treatments is needed.
- To help address GPs' reluctance to use stratification tools. In particular, testing the use of an EHR template that fires at every patient consultation and/or testing a simpler model of stratified care that may be easier to deliver would be valuable.
- The stratification tool relied on self-reported data collected by GPs from patients within short consultations, since routinely collected primary care data does not contain many variables that are known to predict outcomes specifically in musculoskeletal pain. Future research that identifies ways to ensure such prognostic factors are routinely collected and can be used to provide clinicians with prognostic information in ways that do not add time to short consultations would be helpful. This approach would also overcome difficulties for patients who may have low health literacy levels or learning needs.
- A key next step in research is to use prognostic information not only to stratify patients into subgroups as part of stratified health care, but to provide personalised outcome predictions as part of personalised health care. This could include predicting an individual's likely future pain, function scores or their probability of being absent from work to achieve greater treatment tailoring. The European Union-funded Back-UP research programme (<https://backup-project.eu/>) developed personalised prognostic models for back and neck pain patients using the data sets from this Keele STarT MSK programme.
- Clearly it would be useful for the prognostic tool to be available in more languages than English. We therefore recommend researchers help to culturally translate the tool and validate it in different languages. To date, there have been several translations of the STarT MSK Tool, including Hebrew, German, Norwegian, Persian, and Dutch versions. However, there are around a further 15 language versions in development. This information will be available in due course online (<https://www.keele.ac.uk/startmsk/>).
- We recommend that further research is conducted to better understand how the new prognostic tool relates to the secondary care context. There may also be possibilities to use the tool and evaluate its role in initiatives focused on reducing current long NHS waiting lists (e.g. in supporting decisions about which patients might be appropriate for sign-posting to self-management resources).



# Acknowledgements

The authors would like to thank the Keele STarT MSK Committee PSC for overseeing the research; the staff from Keele Clinical Trials Unit, in particular Steff Garvin, Andrea Cherrington, Alicia Bratt and Nicola Halliday; the PPIE group for their support and input, in particular John Murphy and Angie Emery; the wider Keele STarT MSK programme research team (Tom Sanders, Ebenezer Afolabi, Ying Chen, Carolyn Chew-Graham, Hollie Birkinshaw, Elaine Hay, Danielle, van der Windt, Christian Mallen, Vincent Cooper, Chris Main, Hazel Mackay and Rhian Hughes); and all the patients and clinicians for their participation. The Keele STarT MSK research team acknowledges the support of Centre of Excellence grant funding from Versus Arthritis in the development of this research programme (grant number 20202) and the National Institute for Health and Care Research (NIHR) Clinical Research Network. Nadine E Foster is a NIHR Senior Investigator and was supported through a NIHR Research Professorship (NIHR-RP-011-015).

## Non-author contributions

Ying Chen (Associate Professor in Biostatistics & Epidemiology) served as the junior statistician for work package 3. Steff Garvin (Trial manager) served as the trial manager for work packages 3 and 4 and supported their operational management for the Keele Clinical Trials Unit. Vince Cooper (Senior Lecturer in General Practice) served to support work packages 2, 3 and 4 by providing specialist GP clinical knowledge to help design the intervention, recruit practices, design and deliver the GP training, interpret the pilot and main trial findings and support the publications for these work packages. Christian Mallen (Professor of General Practice) served as a co-applicant for the programme, and supported the interpretation of the main trial results and its publication. Danielle van der Windt (Professor of Primary Care Epidemiology) served as a co-applicant for the programme, and supported the interpretation of the main trial results and its publication. Elaine Hay (Professor of Community Rheumatology) served as a co-applicant for the programme, and supported the interpretation of the main trial results and its publication.

## Contribution of authors

**Nadine E Foster** (<https://orcid.org/0000-0003-4429-9756>) (Professor of Musculoskeletal Health) served as overall chief investigator for the whole programme of research, led the funding application, provided senior support on the detailed protocol development and delivery of each work package, chaired the Programme Management Group, and was the senior author on the pilot and main randomised trials.

**Kate M Dunn** (<https://orcid.org/0000-0002-6202-2606>) (Professor of Epidemiology) served as the principal investigator for work package 1 and led the conception and writing of this work package for funding; led the design and development of the STarT MSK tool and its validation; and led or contributed to the publications from this and other work packages.

**Joanne Protheroe** (<https://orcid.org/0000-0002-6202-2606>) (Professor of General Practice) served as the principal investigator for work package 2; led the conception and writing of this work package for funding; led the design, development of the matched treatment options and intervention format and content of GP support packages; and led or contributed to publications from this and other work packages.

**Jonathan C Hill** (<https://orcid.org/0000-0001-6246-1409>) (Professor of Physiotherapy) served as the principal investigator for work packages 3 and 4; led the conception and writing of these work packages for funding; led the design, delivery, analysis and write-up of these work packages; and led first draft

and had overall responsibility for submitting this overall report for publication. He is the corresponding author for the report.

**Martyn Lewis** (<https://orcid.org/0000-0001-5290-7833>) (Reader in Medical Statistics) served as the senior statistician for the trial. He had full access to the data and took responsibility for data integrity and accuracy of data analysis. He participated in the design, conduct, statistical analysis and publication writing of all work packages.

**Ben Saunders** (<https://orcid.org/0000-0002-0856-1596>) (Lecturer in Applied Health Research) served as qualitative research associate in work packages 2 and 3, and social science lead in work package 4; and led the design, data collection, analysis and publication of the qualitative research informing the intervention development and the nested qualitative studies within the feasibility trial and main trial.

**Sue Jowett** (<https://orcid.org/0000-0001-8936-3745>) (Professor of Health Economics) served as the lead health economist, leading on the design, data collection, analysis and publication of the health economics presented in work packages 1 and 4.

**Susie Hennings** (<https://orcid.org/0000-0002-8160-6658>) (Senior Trials Manager for the Keele Clinical Trials Unit) supported the operational management of the programme of research for the final two work packages.

**Paul Campbell** (<https://orcid.org/0000-0001-9148-882X>) (Senior Research Associate) served as research fellow for work package 1 of the programme of research; contributed to the development of the Keele STarT MSK Tool; led the work package 1 protocol publication, co-ordinating work package 1 cohort study, supporting data analysis; and was co-author on the Keele STarT MSK Tool and health economic publications.

**Kieran Bromley** (<https://orcid.org/0000-0002-4129-2519>) (Research Associate) was the junior statistician who supported the analyses of the main trial in work package 4.

**Bernadette Bartlam** (<https://orcid.org/0000-0002-8557-6222>) (Senior Social Scientist) served as senior qualitative lead for work package 2, where she led the design, data collection, analysis and publication of the qualitative aspects of work package 2. She also supported Ben Saunders in the qualitative aspects of work packages 3 and 4 in relation to the design, data collection and publications.

**Opeyemi Babatunde** (<https://orcid.org/0000-0002-5064-6446>) (Lecturer: Evidence Synthesis and Applied Health) served as the research associate (systematic review) in work package 2, leading on the design, data collection, analysis and publication of the evidence synthesis informing the intervention development.

**Simon Wathall** (<https://orcid.org/0000-0002-7107-5785>) (Health Informatics Specialist) served by building the intervention and recruitment computer templates used in work packages 3 and 4, and developing the data extraction searches used to extract patient data from GP electronic health records across the programme work packages.

**Raymond Oppong** (<https://orcid.org/0000-0002-0815-4616>) (Associate Professor in Health Economics) served as a health economist and was involved in design, data collection, analysis and publication of health economics aspects in work package 1.

**Jesse Kigozi** (<https://orcid.org/0000-0001-7608-4923>) (Assistant Professor in Health Economics) served in a health economics analysis and was involved in design, data collection, analysis and publication of health economics work presented in work package 4.

**Adrian Chudyk** (<https://orcid.org/0000-0002-2990-9651>) (Clinical Lecturer in General Practice) served to support work packages 3 and 4 by providing specialist GP clinical knowledge to help design the intervention, recruit and train general practices, interpret the pilot and main trial findings and the publications for these work packages.

## Publications

Campbell P, Hill JC, Protheroe J, Afolabi EK, Lewis M, Beardmore R, *et al.* Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK Tool for subgrouping musculoskeletal patients in primary care. *J Pain Res* 2016;**9**:807–18. <https://doi.org/10.2147/JPR.S116614>

Dunn KM, Campbell P, Lewis M, Hill JC, van der Windt DA, Afolabi E, *et al.* Refinement and validation of a tool for stratifying patients with musculoskeletal pain. *Eur J Pain* 2021;**25**:2081–93. <https://doi.org/10.1002/ejp.1821>

Oppong R, Lewis M, Campbell P, Dunn KM, Foster NE, Hill JC, Jowett S. Comparison of healthcare utilisation, costs and health-related quality of life across the subgroups defined by the Keele STarT MSK Tool. *Rheumatol* 2022;**3**:keac560.

Campbell P, Lewis M, Chen Y, Lacey RJ, Rowlands G, Protheroe J. Can patients with low health literacy be identified from routine primary care health records? A cross-sectional and prospective analysis. *BMC Fam Pract* 2019;**20**:101. <https://doi.org/10.1186/s12875-019-0994-8>

Saunders B, Bartlam B, Foster NE, Hill JC, Cooper V, Protheroe J. General practitioners' and patients' perceptions towards stratified care: a theory informed investigation. *BMC Fam Pract* 2016;**17**:125. <https://doi.org/10.1186/s12875-016-0511-2>

Babatunde OO, Jordan JL, Van der Windt DA, Hill JC, Foster NE, Protheroe J. Effective treatment options for musculoskeletal pain in primary care: a systematic overview of current evidence. *PLoS One* 2017;**12**:e0178621. <https://doi.org/10.1371/journal.pone.0178621>

Protheroe J, Saunders B, Bartlam B, Dunn KM, Cooper V, Campbell P, *et al.* Matching treatment options for risk sub-groups in musculoskeletal pain: a consensus groups study. *BMC Musculoskelet Disord* 2019;**20**:271. <https://doi.org/10.1186/s12891-019-2587-z>

Hill JC, Garvin S, Chen Y, Cooper V, Wathall S, Saunders B, *et al.* Stratified primary care versus non-stratified care for musculoskeletal pain: findings from the STarT MSK feasibility and pilot cluster randomised controlled trial. *BMC Fam Pract* 2020;**21**:30. <https://doi.org/10.1186/s12875-019-1074-9>

Saunders B, Hill JC, Foster NE, Cooper V, Protheroe J, Chudyk A, *et al.* Stratified primary care versus non-stratified care for musculoskeletal pain: qualitative findings from the STarT MSK feasibility and pilot cluster randomised controlled trial. *BMC Fam Pract* 2020;**21**:31. <https://doi.org/10.1186/s12875-020-1098-1>

Hill J, Garvin S, Chen Y, Cooper V, Wathall S, Bartlam B, *et al.* Computer-based stratified primary care for musculoskeletal consultations compared with usual care: study protocol for the STarT MSK cluster randomised controlled trial. *JMIR Res Protoc* 2020;**9**:e17939.

Hill JC, Garvin S, Bromley K, Saunders B, Kigozi J, Cooper V, *et al.* Risk-based stratified primary care for common musculoskeletal pain presentations: results of the STarT MSK cluster randomised controlled trial. *Lancet Rheumatol* 2022;4:E591–E602. [https://doi.org/10.1016/S2665-9913\(22\)00159-X](https://doi.org/10.1016/S2665-9913(22)00159-X)

### Data-sharing statement

Any requests for access to the anonymised data will follow our data-sharing procedure. Requests for anonymised data will be reviewed by our Data Custodian and Academic Proposals Committee. The full statement on data sharing is publicly available from [www.keele.ac.uk/health/fmhsresearchthemes/](http://www.keele.ac.uk/health/fmhsresearchthemes/). All information will be held securely and in strict confidence. Each person in this study will be given a study number so that data from the study will not have any identifiable information, such as names and addresses, and cannot be traced. On this basis, these anonymised data will be kept electronically and may be used in other research studies. Requests for access to the data should be addressed to the corresponding author or to the data custodian.

### Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

# References

1. March L, Smith EU, Hoy DG, Cross MJ, Sanchez-Riera L, Blyth F, *et al.* Burden of disability due to musculoskeletal (MSK) disorders. *Best Pract Res Clin Rheumatol* 2014;**28**:353–66. <https://doi.org/10.1016/j.berh.2014.08.002>
2. 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. <https://doi.org/10.1186/1471-2474-11-144>
3. Foster NE, Hill JC, O'Sullivan P, Hancock M. Stratified models of care. *Best Pract Res Clin Rheumatol* 2013;**27**:649–61. <https://doi.org/10.1016/j.berh.2013.10.005>
4. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, *et al.* Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;**346**:e5793. <https://doi.org/10.1136/bmj.e5793>
5. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, Hay EM. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008;**59**:632–41. <https://doi.org/10.1002/art.23563>
6. 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**:1560–71. [https://doi.org/10.1016/S0140-6736\(11\)60937-9](https://doi.org/10.1016/S0140-6736(11)60937-9)
7. 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**:102–11. <https://doi.org/10.1370/afm.1625>
8. Hill JC, Afolabi EK, Lewis M, Dunn KM, Roddy E, van der Windt DA, Foster NE. Does a modified STarT Back Tool predict outcome with a broader group of musculoskeletal patients than back pain? A secondary analysis of cohort data. *BMJ Open* 2016;**6**:e012445. <https://doi.org/10.1136/bmjopen-2016-012445>
9. Campbell P, Hill JC, Protheroe J, Afolabi EK, Lewis M, Beardmore R, *et al.* Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK Tool for subgrouping musculoskeletal patients in primary care. *J Pain Res* 2016;**9**:807–18. <https://doi.org/10.2147/JPR.S116614>
10. Salisbury C, Montgomery AA, Hollinghurst S, Hopper C, Bishop A, Franchini A, *et al.* Effectiveness of PhysioDirect telephone assessment and advice services for patients with musculoskeletal problems: pragmatic randomised controlled trial. *BMJ* 2013;**346**:f43. <https://doi.org/10.1136/bmj.f43>
11. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain* 1992;**50**:133–49. [https://doi.org/10.1016/0304-3959\(92\)90154-4](https://doi.org/10.1016/0304-3959(92)90154-4)
12. Hayden SR, Brown MD. Likelihood ratio: a powerful tool for incorporating the results of a diagnostic test into clinical decision making. *Ann Emerg Med* 1999;**33**:575–80. [https://doi.org/10.1016/S0196-0644\(99\)70346-X](https://doi.org/10.1016/S0196-0644(99)70346-X)
13. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;**365**:1500–5. [https://doi.org/10.1016/S0140-6736\(05\)66422-7](https://doi.org/10.1016/S0140-6736(05)66422-7)
14. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:389–91. <https://doi.org/10.1001/jama.271.5.389>

15. Department of Health and Social Care. *NHS Reference Costs 2015 to 2016*. URL: [www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016](http://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016) (accessed 29 April 2022).
16. Curtis L, Burns A. *Unit Costs of Health and Social Care 2016*. Canterbury: PSSRU, University of Kent; 2016.
17. Joint Formulary Committee. *British National Formulary*. 72 ed. London: BMJ Group and Pharmaceutical Press; 2016.
18. Hill JC, Garvin S, Chen Y, Cooper V, Wathall S, Saunders B, et al. Stratified primary care versus non-stratified care for musculoskeletal pain: findings from the STarT MSK feasibility and pilot cluster randomized controlled trial. *BMC Fam Pract* 2020;**21**:30. <https://doi.org/10.1186/s12875-019-1074-9>
19. Babatunde OO, Jordan JL, Van der Windt DA, Hill JC, Foster NE, Protheroe J. Effective treatment options for musculoskeletal pain in primary care: a systematic overview of current evidence. *PLOS One* 2017;**12**:e0178621. <https://doi.org/10.1371/journal.pone.0178621>
20. Saunders B, Bartlam B, Foster NE, Hill JC, Cooper V, Protheroe J. General practitioners' and patients' perceptions towards stratified care: a theory informed investigation. *BMC Fam Pract* 2016;**17**:125. <https://doi.org/10.1186/s12875-016-0511-2>
21. Protheroe J, Saunders B, Bartlam B, Dunn KM, Cooper V, Campbell P, et al. Matching treatment options for risk sub-groups in musculoskeletal pain: a consensus groups study. *BMC Musculoskelet Disord* 2019;**20**:271. <https://doi.org/10.1186/s12891-019-2587-z>
22. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;**7**:10. <https://doi.org/10.1186/1471-2288-7-10>
23. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490. <https://doi.org/10.1136/bmj.328.7454.1490>
24. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A, 'Psychological Theory' Group. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* 2005;**14**:26–33. <https://doi.org/10.1136/qshc.2004.011155>
25. Delbecq AL, van de Ven AH, Gustafson DH. *Group Techniques for Problem Planning*. Glenview, IL: Scott-Foresman; 1975.
26. Babatunde OO, Tan V, Jordan JL, Dziedzic K, Chew-Graham CA, Jinks C, et al. Evidence flowers: an innovative, visual method of presenting 'best evidencesqu; summaries to health professional and lay audiences. *Res Synth Methods* 2018;**9**:273–84. <https://doi.org/10.1002/jrsm.1295>
27. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 2015;**350**:h1258. <https://doi.org/10.1136/bmj.h1258>
28. Kortteisto T, Raitanen J, Komulainen J, Kunnamo I, Mäkelä M, Rissanen P, Kaila M. Patient-specific computer-based decision support in primary healthcare – a randomized trial. *Implement Sci* 2014;**9**:15. <https://doi.org/10.1186/1748-5908-9-15>
29. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implement Sci* 2012;**7**:37. <https://doi.org/10.1186/1748-5908-7-37>
30. Brookfield S. *Understanding and Facilitating Adult Learning: A Comprehensive Analysis of Principles and Effective Practices*. Buckingham: Open University Press; 1986.



31. Saunders B, Hill JC, Foster NE, Cooper V, Protheroe J, Chudyk A, *et al.* Stratified primary care versus non-stratified care for musculoskeletal pain: qualitative findings from the STarT MSK feasibility and pilot cluster randomized controlled trial. *BMC Fam Pract* 2020;**21**:31. <https://doi.org/10.1186/s12875-020-1098-1>
32. Hill J, Garvin S, Chen Y, Cooper V, Wathall S, Bartlam B, *et al.* Computer-based stratified primary care for musculoskeletal consultations compared with usual care: study protocol for the STarT MSK cluster randomized controlled trial. *JMIR Res Protoc* 2020;**9**:e17939. <https://doi.org/10.2196/17939>
33. Curtis L, Burns A. *Unit Costs of Health and Social Care* 2019. Canterbury: PSSRU, University of Kent; 2019.
34. Department of Health and Social Care (DHSC). *National Schedule of Reference Costs: 2018–19*. London: DHSC; 2019.
35. Joint Formulary Committee. *British National Formulary*. 78 ed. London: BMJ Group and Pharmaceutical Press; 2019.
36. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. Hillsdale, NJ: Lawrence Earlbaum; 1988.
37. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006;**35**:1292–300. <https://doi.org/10.1093/ije/dyl129>
38. Vickers AJ. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Med Res Methodol* 2003;**3**:22. <https://doi.org/10.1186/1471-2288-3-22>
39. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;**345**:e5661. <https://doi.org/10.1136/bmj.e5661>
40. Morris S, Devlin N, Parkin D. *Economic Analysis in Health Care*. Chichester: John Wiley and Sons; 2007.
41. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008;**26**:733–44.
42. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS) – explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013;**16**:231–50. <https://doi.org/10.1016/j.jval.2013.02.002>
43. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, *et al.* Development of a theory of implementation and integration: Normalization Process Theory. *Implement Sci* 2009;**4**:29. <https://doi.org/10.1186/1748-5908-4-29>
44. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;**357**.





# Appendix 1 Tables and figures from work package 1

**TABLE 1** Baseline characteristics of participants in the cohort study

Characteristic	Cohort study (N = 1890)
Age (years), mean (SD)	58.3 (16.1)
Female, n (%)	1145 (60.6)
Index pain site, n (%)	
Knee	349 (18.5)
Neck	57 (3.0)
Back	408 (21.6)
Shoulder	103 (5.4)
Multisite	973 (51.5)
Live alone, n (%)	394 (21.0)
Employed, n (%)	747 (41.1)
Time off work in last 6 months, n (%)	318 (16.8)
Pain at consultation, mean (SD)	NA
Pain intensity, mean (SD) points	
Mean of least, average and current pain	5.3 (2.4)
Usual pain	6.2 (2.5)
Duration: time since no pain, n (%) (n = 29 missing)	
< 3 months	403 (21.7)
3–6 months	225 (12.1)
7–12 months	212 (11.4)
1–5 years	521 (27.6)
≥ 6 years	500 (26.5)
SF-36 component scales, mean score (SD) (n = 116 missing)	
Physical	36.2 (10.1)
Mental	43.6 (13.2)
PROMIS pain interference, mean (SD) points (n = 46 missing)	62.1 (8.1)
Pain self-efficacy, mean (SD) points (n = 31 missing)	37.2 (16.1)
Catastrophising, mean (SD) points (n = 13 missing)	9.7 (8.9)
Long-term medical conditions, n (%)	
Diabetes	217 (11.5)
Breathing problems/COPD/asthma	334 (17.7)
Heart problems or high blood pressure	579 (30.7)

continued

**TABLE 1** Baseline characteristics of participants in the cohort study (*continued*)

Characteristic	Cohort study (N = 1890)
Chronic fatigue, ME, fibromyalgia, widespread pain	84 (4.5)
Anxiety, depression, stress	446 (23.6)
Other	495 (26.2)
Health literacy problems, n (%)	
Never/rarely	1555 (82.3)
Sometimes/often/always	325 (17.3)
EQ-5D-5L score, mean (SD)	0.56 (0.27)

COPD, chronic obstructive pulmonary disease; ME, myalgic encephalomyelitis; NA, not applicable.

**TABLE 2** Characteristics and outcomes in the KAPS cohort population, overall and within subgroups defined by the Keele STarT MSK Tool

Characteristic/outcome	All	Keele STarT MSK Tool subgroup		
		High risk	Medium risk	Low risk
SF-36 PCS, mean (SD)				
Baseline	36.2 (10.1)	28.4 (7.3)	36.8 (8.1)	45.8 (8.0)
6 months	38.6 (11.4)	30.2 (8.8)	39.0 (10.0)	48.0 (8.2)
Poor outcome at 6 months, n (%)	581 (53.4)	287 (87.5)	257 (53.7)	43 (15.1)
Pain intensity, mean (SD) points				
Baseline	5.3 (2.4)	7.2 (1.6)	5.3 (1.7)	2.8 (1.6)
6 months	4.1 (2.8)	6.2 (2.3)	4.0 (2.4)	1.9 (1.9)
Poor outcome at 6 months, n (%)	482 (42.3)	263 (75.1)	200 (40.7)	33 (11.1)
PROMIS pain interference scale, mean (SD) points				
Baseline	62.1 (8.1)	68.8 (4.9)	61.9 (5.6)	53.6 (6.6)
6 months	59.1 (9.0)	65.9 (6.7)	58.4 (7.3)	51.3 (7.4)
EQ-5D-5L score, mean (SD)				
Baseline	0.56 (0.27)	0.33 (0.26)	0.62 (0.18)	0.78 (0.11)
6 months	0.62 (0.26)	0.42 (0.28)	0.66 (0.19)	0.81 (0.15)
Pain self efficacy questionnaire, mean (SD) points				
Baseline	37.2 (16.1)	24.3 (13.6)	39.3 (12.7)	51.6 (8.8)
6 months	39.9 (16.1)	27.0 (14.5)	42.1 (13.2)	52.3 (10.0)
SF-36 mental component score, mean (SD)				
Baseline	43.6 (13.2)	35.1 (12.3)	45.4 (11.7)	52.4 (9.1)
6 months	47.7 (11.9)	40.2 (13.0)	49.2 (10.4)	54.1 (7.5)
Pain catastrophising, mean (SD) points				
Baseline	9.7 (8.9)	16.3 (9.2)	8.3 (6.8)	3.4 (4.7)
6 months	7.8 (8.4)	13.8 (9.3)	6.9 (7.1)	2.4 (4.0)

**TABLE 2** Characteristics and outcomes in the KAPS cohort population, overall and within subgroups defined by the Keele STarT MSK Tool (continued)

Characteristic/outcome	All	Keele STarT MSK Tool subgroup		
		High risk	Medium risk	Low risk
Sleep problems, n (%)				
Baseline	1193 (63.1)	464 (82.1)	449 (63.7)	161 (38.4)
6 months	675 (54.3)	266 (76.4)	261 (53.5)	82 (28.0)
Global change: 'much improved' at 6 months, n (%)	353 (24.3)	38 (9.0)	118 (21.0)	167 (50.1)
All between-subgroup statistical tests for differences in summary measures were significant ( $p < 0.001$ ) through one-way analysis of variance with linear contrast and non-parametric Jonckheere–Terpstra tests (for numerical outcomes) and chi-square test-for-trend (for categorical outcomes).				

**TABLE 3** Health-care utilisation over 6 months (complete cases)

Health-care resource	Overall (N =1253)	Keele STarT MSK risk subgroup		
		Low (N = 298)	Medium (N = 491)	High (N = 350)
Primary care health-care utilisation, mean (SD)				
GP consultations				
Practice	1.44 (2.191)	0.66 (0.959)	1.40 (2.154)	2.22 (2.738)
Home	0.12 (0.862)	0.04 (0.249)	0.07 (0.363)	0.18 (0.821)
Nurse consultations				
Practice	0.19 (0.774)	0.08 (0.348)	0.18 (0.777)	0.29 (1.029)
Home	0.05 (0.578)	0.04 (0.695)	0.05 (0.662)	0.06 (0.400)
Other primary care consultations <sup>c</sup>				
Practice	1.06 (2.823)	0.53 (1.500)	1.28 (3.227)	1.27 (3.186)
Home	0.10 (0.951)	0.01 (0.141)	0.13 (1.295)	0.14 (0.841)
Secondary care health care-utilisation, mean (SD)				
Consultant <sup>d</sup>				
NHS	0.44 (1.349)	0.17 (0.505)	0.42 (1.196)	0.75 (1.982)
Private	0.21 (1.010)	0.11 (0.737)	0.26 (1.038)	0.26 (1.234)
Physiotherapist				
NHS	0.51 (1.667)	0.34 (1.191)	0.50 (1.542)	0.77 (2.238)
Private	0.33 (1.519)	0.20 (1.094)	0.41 (1.531)	0.39 (1.900)
Acupuncture				
NHS	0.08 (0.917)	0.03 (0.337)	0.10 (0.337)	0.11 (1.436)
Private	0.07 (0.596)	0.06 (0.562)	0.10 (0.706)	0.10 (0.553)
Osteopath				
NHS	0.01 (0.135)	0.01 (0.116)	0	0.02 (0.232)
Private	0.04 (0.536)	0.02 (0.173)	0.03 (0.241)	0.09 (0.953)
continued				

**TABLE 3** Health-care utilisation over 6 months (complete cases) (*continued*)

Health-care resource	Overall (N =1253)	Keele STarT MSK risk subgroup <sup>b</sup>		
		Low (N = 298)	Medium (N = 491)	High (N = 350)
Other secondary care health-care utilisation and OTC medication, n (%)				
Overnight stay in hospital	43 (3.4)	2 (0.7)	19 (3.9)	21 (6.0)
Treatments or investigations	345 (28)	43 (14.5)	144 (29.5)	129 (36.9)
OTC medication	608 (49)	121 (40.7)	256 (52.4)	180 (51.6)
Prescribed medication, n (%) <sup>a</sup>				
Total number of prescriptions, N	7039	536	2125	3659
Simple analgesic	865 (12.3)	62 (11.6)	309 (14.5)	406 (11.1)
Topical analgesic	624 (8.9)	81 (15.1)	180 (8.5)	285 (7.8)
Compound analgesic	1504 (21.4)	130 (24.3)	509 (23.9)	697 (19.1)
NSAID	1001 (14.2)	128 (23.9)	397 (18.7)	391 (10.7)
Skeletal muscle relaxant	245 (3.5)	12 (2.2)	50 (2.4)	173 (4.7)
Neuropathic pain medication	734 (10.4)	21 (3.9)	186 (8.8)	483 (13.2)
Opioid medication	1818 (25.8)	45 (8.4)	411 (19.3)	1177 (32.2)
Corticosteroid injection	77 (1.1)	20 (3.7)	31 (1.5)	19 (0.5)
Other treatments <sup>f</sup>	171 (2.43)	37 (6.9)	52 (2.5)	28 (0.8)
<p>a Overall analysis includes the 114 participants with complete data who were without a Keele STarT MSK subgroup classification.</p> <p>b Subgroup analyses exclude the 114 participants with complete data who were without a Keele STarT MSK risk classification.</p> <p>c Includes visits to physiotherapists in primary care.</p> <p>d Includes visits to rheumatologists, podiatrists, chiropractors and surgeons in secondary care.</p> <p>e Obtained from medical records, with numbers indicating the number of times each drug class was prescribed as opposed to the number of patients who used them.</p> <p>f Includes other miscellaneous treatments.</p>				

## Appendix 2 Tables and figures from work package 2

### BOX 1 Clinician support package developed in work package 2

#### Overall scope and plan

Total training time available to GPs is 4 hours, provisionally to be in two 2-hour sessions. Optionally, this can be supplemented by one 'catch-up' session with individual GPs at their request or in response to problems identified by the study team.

Two TAPS facilitators to attend each session, aiming at continuity of at least one for both sessions.

#### Training approach

Training is for individual practices and based on all GPs attending both sessions and working as a small group with the Keele GP facilitators. There are some knowledge and skills components to be covered and the entire sessions should be interactive and collaborative, exploring and building on the GPs' current practice. Particularly during the pilot phase, there will be lessons for the study team to learn and, possibly, some changes to be made to the intervention, so the facilitators will gather information for the team as well as delivering and documenting the training.

#### Key issues to address

- Tool complements normal clinical practice and does not replace it.
- It is a prognostic tool to aid management, not a diagnostic tool.
- A key step in integrating the tool into the consultation is the need to enter a provisional read code *during* the consultation to trigger the template.

#### Requirements for delivering the training

- Protected time for all GPs to attend.
- Co-ordination with practice manager.
- Training room suitable for small group learning.
- Computer, linked to clinical system, with display visible to the group.
- TAPS templates installed and tested.

#### Support materials

- Slide sets for sessions 1 and 2.
- Patient vignettes from TAPS.
- Laminated copy of the Keele STarT MSK tool and matched treatment options.
- A plan of proposed training sessions and a blank training record sheet to be completed after each session.

TAPS, Treatment for Aches and Pains Study.

TABLE 4 Outline schedule for GP training session

Timing	Topic	Detail	Methods and resources
10 minutes	Introductions	Personal introductions, roles, etc. Brief outline of the practice and its population Special interests of GPs	<i>Pre-trial background sheet completed by practice</i> Informal chat to get people warmed up
10 minutes	Brief outline of study its back-ground and scope	Origins of research in STarT Back Explain prognostic risk Clinical conditions and sites involved What we are investigating, in general terms	<i>Few slides – scant detail</i> Interactive presentation and brief Q/A
continued			

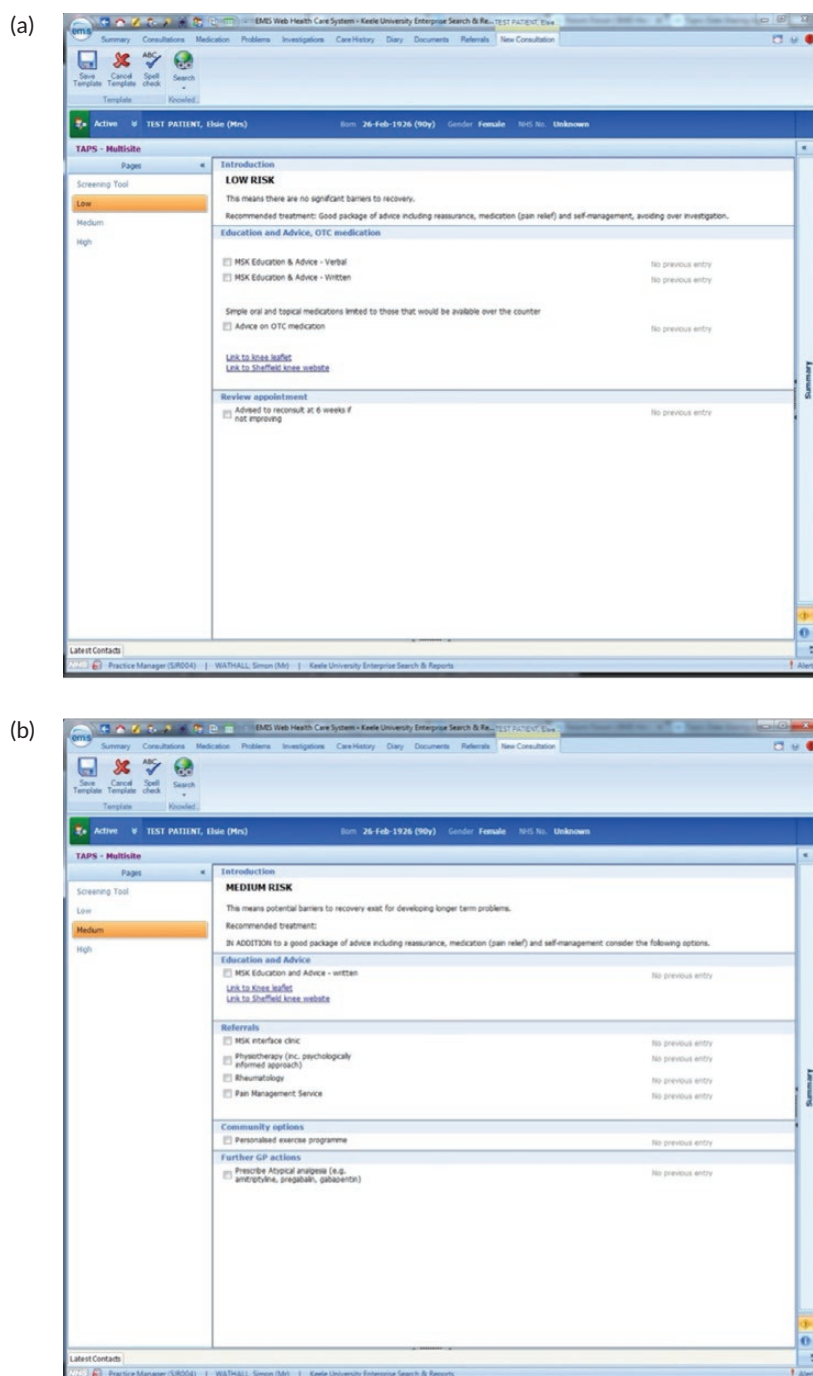
TABLE 4 Outline schedule for GP training session (continued)

Timing	Topic	Detail	Methods and resources
10 minutes	GPs' current management of these conditions	Diagnostic approaches: biomechanical/biopsychosocial – use shoulder pain as example Investigations routinely used: what and where Advice generally given to these patients Sickness certification Medication preferences and usage Physiotherapy, etc. availability and usage Referral options and patterns for different pain sites: musculoskeletal, surgical etc Significant constraints they experience Patients' expectations, e.g. imaging, certificates, referral	<i>Pre-trial background sheet</i> General discussion to gauge GPs' philosophy and general approaches: helps build relationship and aid to tailoring our approach to training Avoid detail on specific conditions within musculoskeletal <i>Flip chart to explore treatment/referral options for the practice</i>
20 minutes	GPs' usual consultation habits	Map out their usual consultation process/flow Is computer used during or after consultations? Read coded diagnosis entered at provisional stage or not Any existing use of templates and decision aids? Use of interactive tool plus printed advice, e.g. PILS	More informal discussion <i>A4 sheet with a few prompt statements for GPs</i> <i>Pads of paper for GPs' notes</i> <i>Sticky-note pads to capture notes and queries for later</i>
20 minutes	Stratified care approach	What is stratified care and how does it differ?  Why it may have advantages for patients and NHS Basis for prognostic stratification tool Expected proportion in each risk group The tool identifies potential treatment targets How this complements usual diagnostic clinical practice Matched treatment options and how we devised them No change in local pathways during the study: treatment options are pointers to be used with these pathways	Interactive presentation and Q/A <i>Slides:</i> <ul style="list-style-type: none"> <li>• <i>knowledge about stratified care</i></li> <li>• <i>establish credibility of tool and matched treatments</i></li> <li>• <i>emphasise 'risk' is of chronicity/complexity not pathology</i></li> <li>• <i>explain complementarity with diagnostic process</i></li> <li>• <i>no new pathways at this stage</i></li> </ul>
45 minutes	The Keele STarT MSK tool in practice	Overview of questionnaire and matched treatments Key GP behaviours the tool tries to nudge/change Providing the tool score to onward treating clinicians Trying out the tool: paper exercise – <ul style="list-style-type: none"> <li>• GPs work in pairs, each with a vignette</li> <li>• One asks questions and completes paper tool, other responds from vignette</li> <li>• Swap roles for second vignette</li> <li>• Compare scores and experience of using tool</li> </ul> Demonstration of integrated template by facilitator All GPs trying it out with support	Discussion around <i>slides</i> <i>Pyramid slide for overview</i> <i>Questionnaire and matched treatments</i> Giving patients score and recommended options Communicating score in referrals  <i>Paper copies of vignettes and risk tool</i> <i>Live EMIS system with template</i> <i>Demo of template use</i> All GPs trying out template, using vignettes, with no attempt at consultation elements <i>Vignettes needed: low risk knee-pain, medium-risk shoulder pain, High risk multisite pain with co-morbidity</i>
5 minutes	Suggested preparation for Session 2	Try template a few more times with dummy patients Look at treatment options and linked patient info	Replace this with a short break if running 2 sessions together: <i>would need refreshments</i>

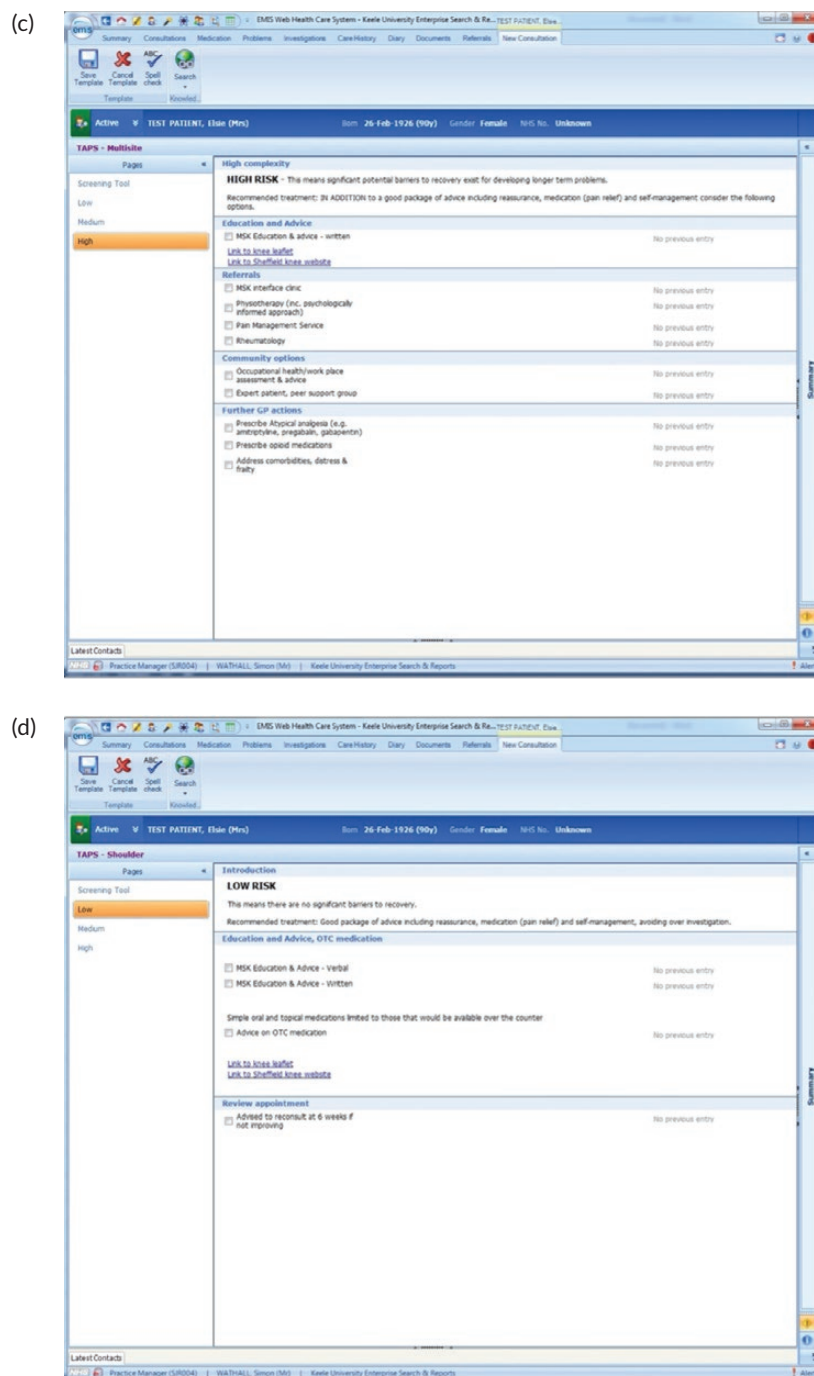
Q/A, question/answer.

**Note**

Resources used are indicated by italic text.

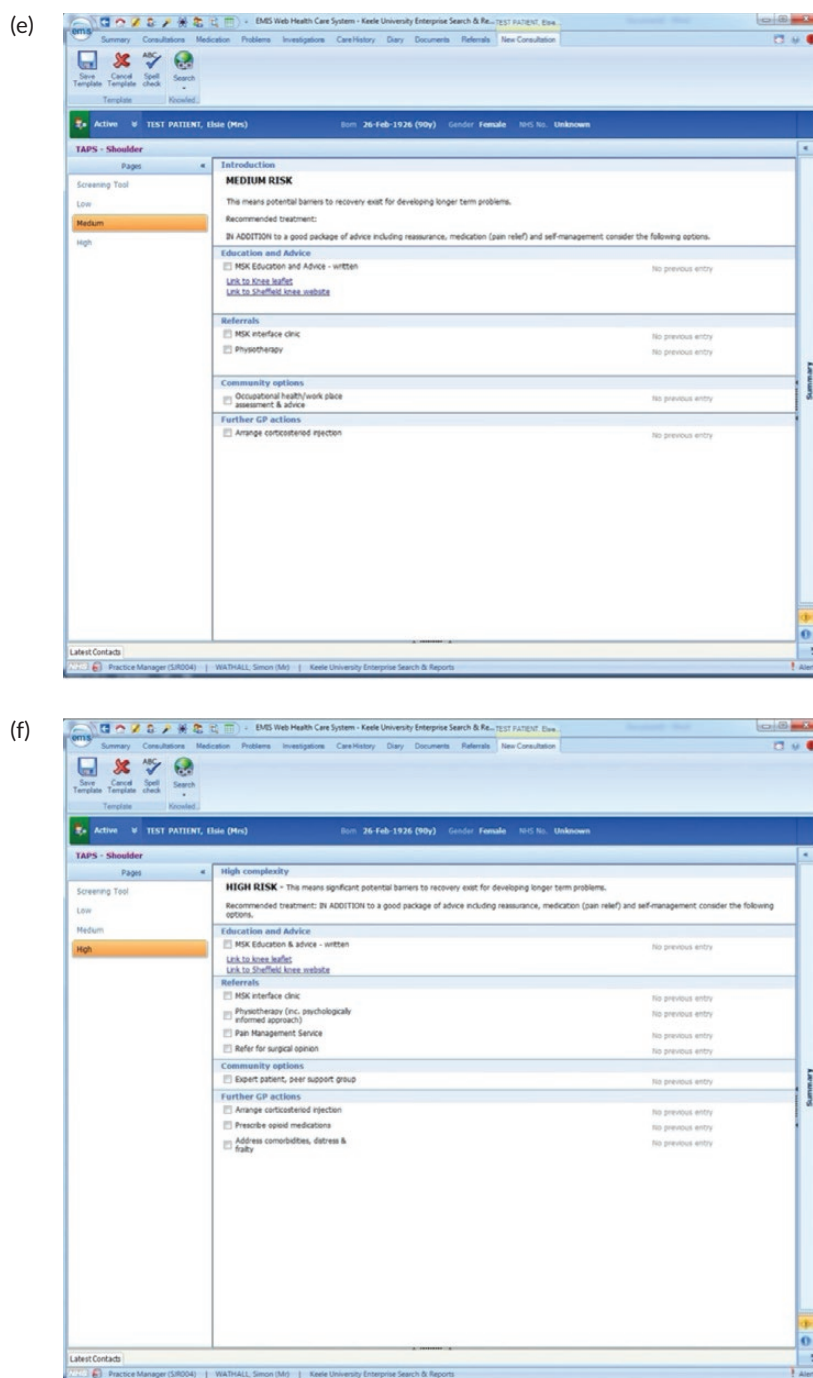


**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)

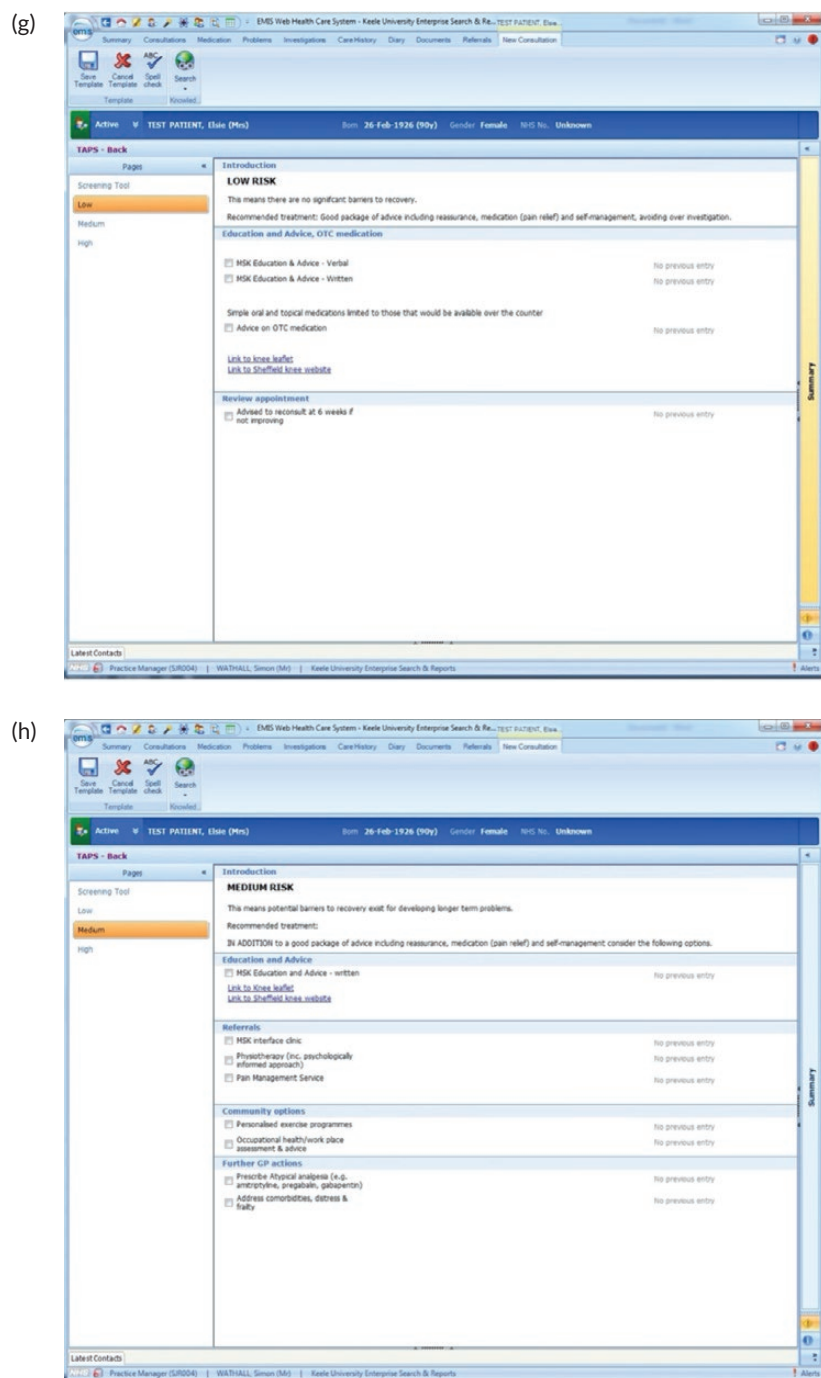


**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)

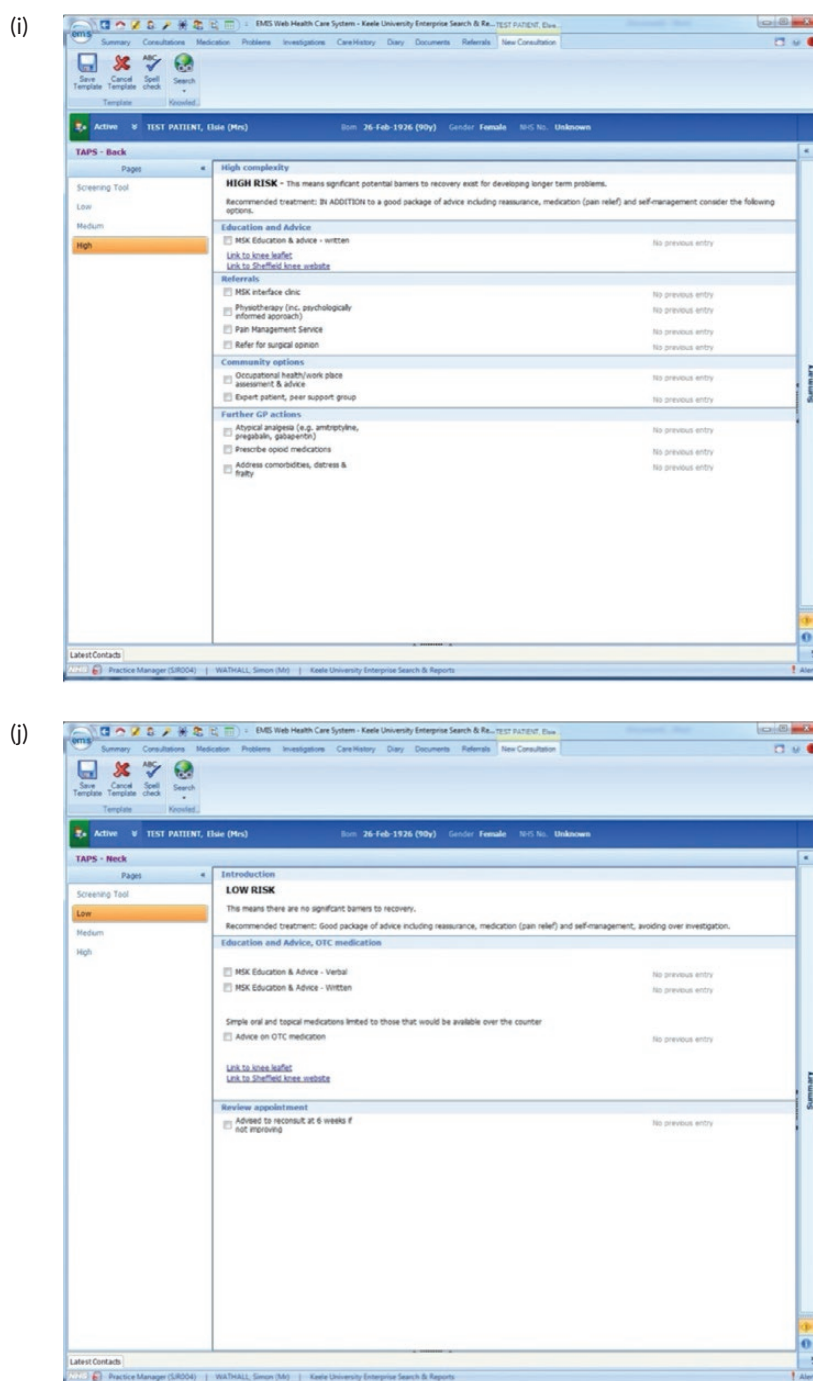




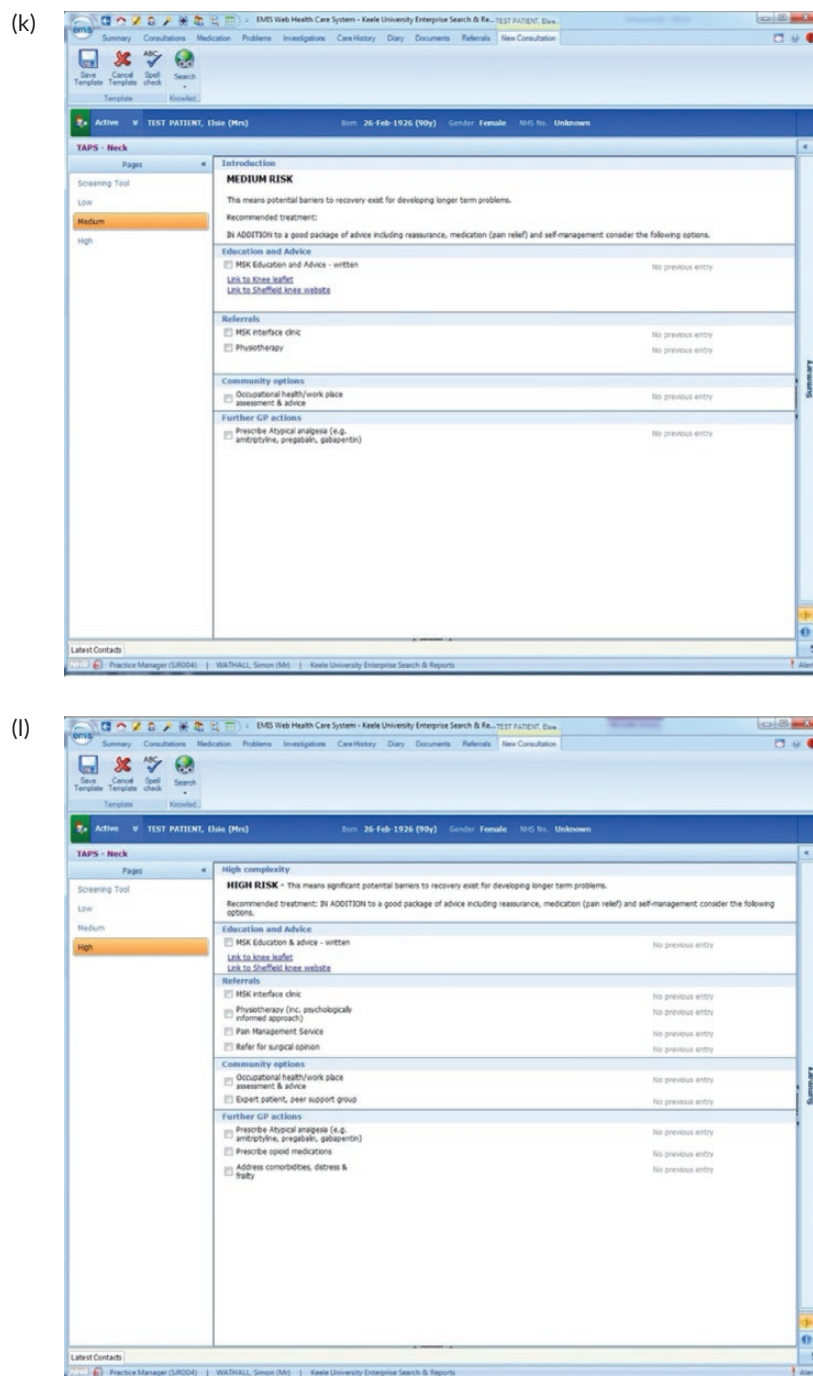
**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)



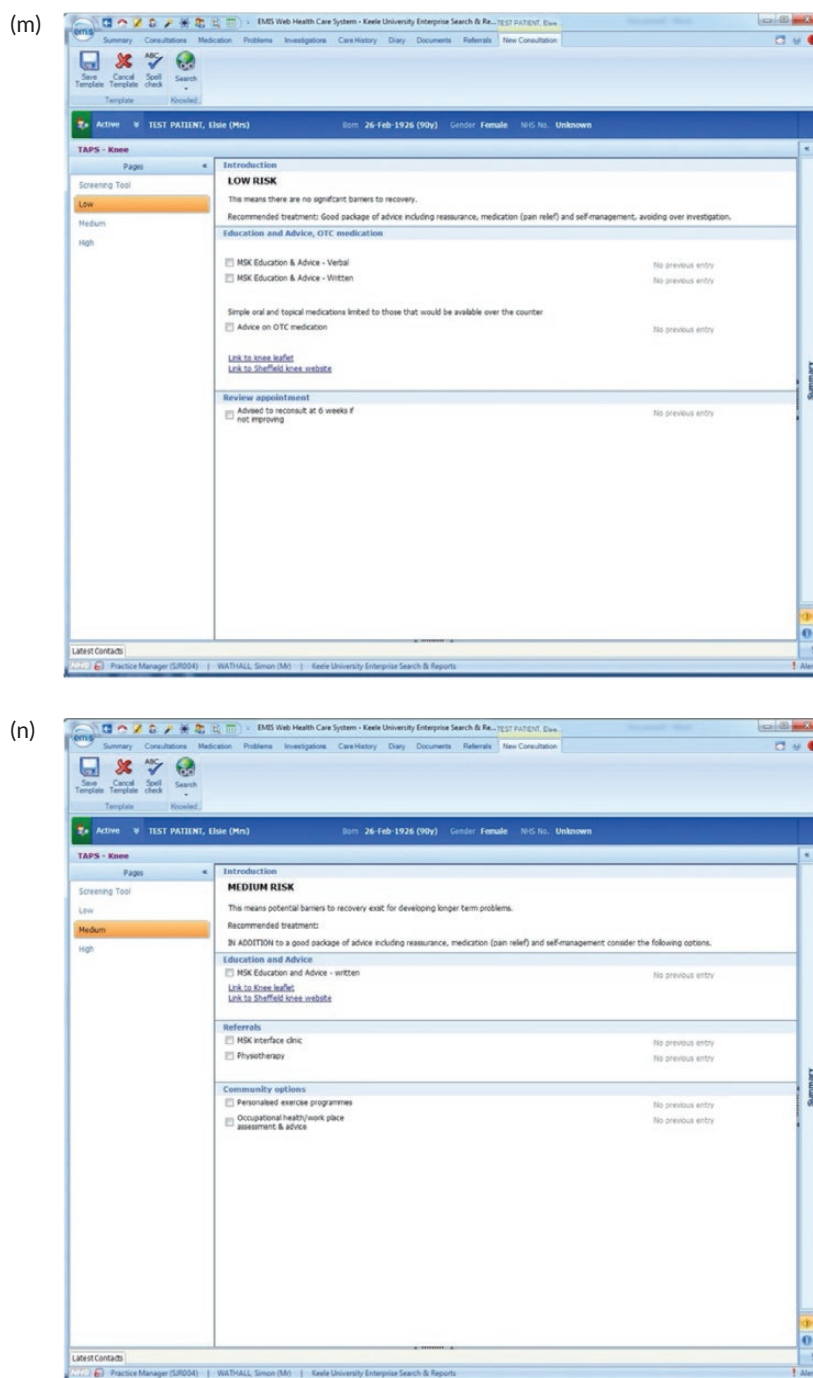
**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)



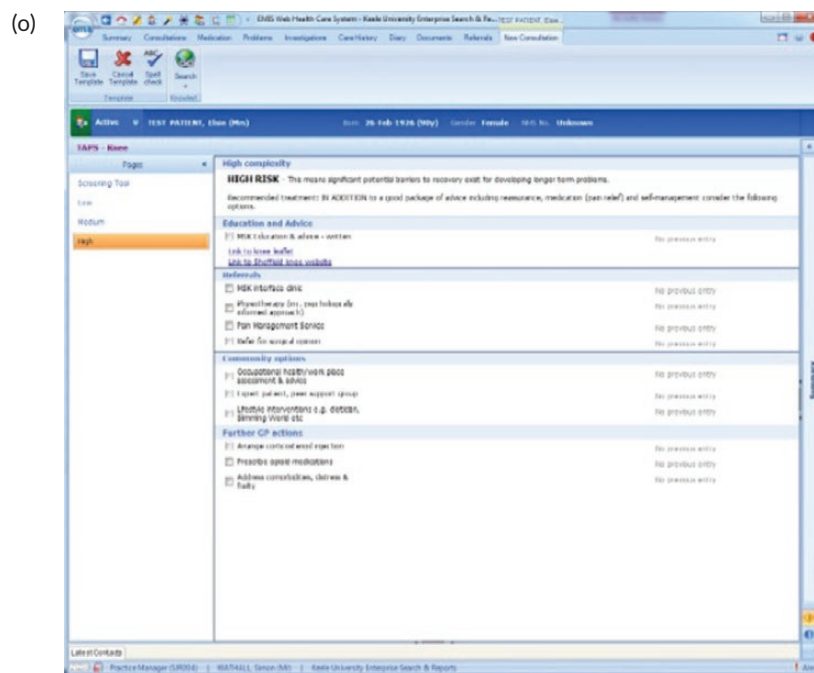
**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)



**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)



**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk. (*continued*)



**FIGURE 2** EMIS templates for matched treatment options. (a) Multisite pain patients classified as low risk; (b) multisite pain patients classified as medium risk; (c) multisite pain patients classified as high risk; (d) shoulder pain patients classified as low risk; (e) shoulder pain patients classified as medium risk; (f) shoulder pain patients classified as high risk; (g) back pain patients classified as low risk; (h) back pain patients classified as medium risk; (i) back pain patients classified as high risk; (j) neck pain patients classified as low risk; (k) neck pain patients classified as medium risk; (l) neck pain patients classified as high risk; (m) knee pain patients classified as low risk; (n) knee pain patients classified as medium risk; and (o) knee pain patients classified as high risk.



## Appendix 3 Tables and figures from work package 3

### BOX 2 Per-protocol recommended matched treatment options for each risk subgroup developed in work package 2

#### Low risk matched treatment options

- Musculoskeletal education and advice: for example, exercise, activity modification, weight loss.
- Advice on OTC medication (simple oral and topical medications limited to those that would be available over the counter).
- Refer to supported self-management and locally available community resources, for example walking group or exercise on prescription.
- Review by primary care practitioner if not improving after 6 weeks.

Per protocol: GPs selected at least one low-risk option with none of the medium- or high-risk options.

#### Medium risk matched treatment options

- Musculoskeletal education and advice: for example, to exercise, activity modification, weight loss.
- Advice on OTC medication (simple oral and topical medications limited to those that would be available over the counter).
- Refer to supported self-management and locally available community resources, for example walking group or exercise on prescription.
- Review by primary care practitioner if not improving after 6 weeks.
- Refer to musculoskeletal interface clinic.
- Refer to physiotherapy.
- Refer to psychosocial intervention or multidisciplinary pain management service.
- Personalised exercise programmes.
- Occupational health/workplace assessment and advice.
- Prescribe atypical analgesia (e.g. amitriptyline, pregabalin and gabapentin). Consider if neuropathic pain present.
- Arrange corticosteroid injection (not recommended for neck or back pain).
- Refer to rheumatology.

Per protocol: GPs selected at least one medium-risk option with none of the high-risk options.

#### High risk matched treatment options

- Musculoskeletal education and advice: for example, to exercise, activity modification, weight loss.
- Advice on OTC medication (simple oral and topical medications limited to those that would be available over the counter).
- Refer to supported self-management and locally available community resources, for example walking group or exercise on prescription.
- Review by primary care practitioner if not improving after 6 weeks.
- Refer to musculoskeletal interface clinic.
- Refer to physiotherapy.
- Refer to psychosocial intervention or multidisciplinary pain management service.
- Personalised exercise programmes.
- Occupational health/workplace assessment and advice.
- Prescribe atypical analgesia (e.g. amitriptyline, pregabalin and gabapentin). Consider if neuropathic pain present.
- Arrange corticosteroid injection (not recommended for neck or back pain).
- Refer to rheumatology.
- Sign-post to expert patient or peer support group.
- Sign-post to lifestyle interventions [e.g. dietitian or Slimming World (Alfreton, UK), etc.].
- Prescribe opioid medication (consider weak opioid if acute pain as alternative to NSAIDs).
- Refer to surgical opinion.
- Address comorbidities, distress and frailty.

Per protocol: GPs selected at least one high-risk treatment option or referral to physiotherapy/musculoskeletal interface clinic with tool subgroup information within their referral so that services were aware that an onward referral to a high-risk treatment option might be required.

TABLE 5 Example monthly report showing peer-to-peer comparisons of GP engagement with recruitment template in work package 3

Anonymised GP identifier					Patients, n (invited, n = 127; consented, n = 65)							
	Back	Knee	Multisite	Neck	Shoulder	No time	Patient not present	Patient declined	Suspected serious pathology	Vulnerable patient	GP escaped	Grand total
a	5	1	1	3	1	46	23		2	1	1	84
b						1	1				1	3
c						1		2				3
d	11	3		2	4	8	21	2	2	1	5	59
d	7	10		3	1	19	17	17			3	77
e	1	2	2			45	13	7	1		3	74
f								1				1
g	1		1			51		30			1	84
h	5	1				2		4	1	2	5	20
i	1							5			1	7
j	7	5	2	2		51	2	2		4	9	84
k	2			1	1	1	1	9				15
l								2				2
m	16	6		4		13	11	12	2	1	6	71
n	5	8		1	1	18	44	3		2	3	85
o						10		4		1	1	16
p						5						5
q						1						1
Grand total	61	36	6	16	8	272	133	100	8	12	39	691
Percentage for general practice (%)	9	5	1	2	1	39	19	14	1	2	6	100
Percentage for all general practices in the trial (%)	12	8	3	3	3	18	12	27	1	3	9	100



**TABLE 6** Example monthly report showing peer to peer comparisons of GP clinical decision-making from intervention general practices in work package 3

Decision made	Patients (n)															Proportion of total from all TAPS practices (%)
	Anonymised GP identifier															
	a	b	c	d	e	f	g	h	i	j	k	Grand total	Proportion of total from this GP surgery (%)			
Low – per protocol	1	1						3		8	3	16	13	16		
Medium – per protocol	5	6	13		2	4		10		10	7	57	45	44		
High – per protocol	2	2	6	1		1	1	3		5	3	24	19	18		
Low – no treatments selected	3	1		1					2		1	8	6	3		
Medium – no treatments selected		7		2		1		2				12	9	5		
High – no treatments selected		3		1								4	3	3		
Low – with medium treatments			1								1	2	2	1		
Low – with high treatments													0	1		
Medium – low treatments only			1							2		3	2	6		
Medium – with high treatment													0	2		
High – low treatments only													0	0		
High – with medium options selected										1		1	1	1		
Grand total	11	20	21	5	2	6	1	16	4	26	15	127	100	100		
TAPS, Treatment for Aches and Pains Study.																

**TABLE 7** Summary of participant self-reported measures used in work package 3

Conceptual domain	Operational definition	Empirical measure used	Number of items	Time point of data collection
Age	Age at index consultation	Date of birth	1	GP EMR audit
Sex	Sex	Male/female	1	GP EMR audit
Index pain location	Site of index pain complaint	Choice of anatomical region	1	GP EMR audit
Pain intensity	Usual pain intensity	0–10-point NRS	1	GP EMR audit, I, 6FU, MF, MDC
Socioeconomic status (IMD)	The individual's (1) current or (2) most recent job title	Job title: categorised as manual/non-manual	2	GP EMR audit
GP practice	GP practice consulted for musculoskeletal pain	Taken from medical record	1	GP EMR audit
Episode duration	Time since last whole month pain free	Episode duration	1	I
Health literacy screen	Health literacy	Single question: Likert scale	1	I
Comorbidities	Self-reported diagnosed comorbidities from a provided list	Yes	1	I
Widespread pain	Presence of widespread pain	Yes/no	1	I
Support needed	Support to complete questionnaire	Yes/no	1	I
Living arrangements	Lives alone	Yes/no	1	I
Previous episodes	Number of previous pain episodes	Number	1	I
Perceived reassurance from GP consultation	ECRQ	12 items with 7-point Likert scale	12	I
Receipt of written education material from GP	Single item to ask if patient received written information at their GP visit	Yes/no/do not remember	1	I
Pain self-efficacy	Single item: confidence to manage pain	0–10-point NRS	1	I, MF
Psychological distress	Single item regarding level of distress	0–10-point NRS	1	I, MF
Employment status and absence from work	Employment status at time of questionnaire	Yes/no and details	1	I, 6FU
Risk status: development version of the Keele STarT MSK Tool	Risk of persistent disabling pain	Yes/no	9	I, 6FU
Musculoskeletal health	Impact from musculoskeletal symptoms	MSK-HQ	14	I, 6FU
Overall rating of change	Change since index pain consultation	Single question: –5 to +5 scale	1	I, 6FU
Physical activity level	Days past week of moderate activity	1–7 days	1	I, 6FU
Fear avoidance beliefs	Fear of movement	TSK-11	11	I, 6FU
Satisfaction	Satisfaction with care	Single question – Likert scale	1	I, 6FU

**TABLE 7** Summary of participant self-reported measures used in work package 3 (*continued*)

Conceptual domain	Operational definition	Empirical measure used	Number of items	Time point of data collection
Physical function				
Back pain patients	Site-specific physical function	RMDQ: original version	24	I, 6FU
Neck pain patients	Site-specific physical function	NDI	10	I, 6FU
Shoulder pain patients	Site-specific physical function	SPADI	13	I, 6FU
Knee pain patients	Site-specific physical function	KOOS-PS	7	I, 6FU
Multisite pain	Site-specific physical function	SF-12 PCS	12	I, 6FU
Health-related quality of life	Utility-based quality of life	EQ-5D-5L	5	I, 6FU, MDC
Health-care costs				
Performance at work	How productivity at work is affected	0–10-point NRS	1	I, 6FU
Work absence	Number of days absent from work	Yes/no and details	1	I, 6FU
Health-care resource use	Use of primary care, other NHS services and private health care	Yes/no and, if yes, details of resources used	3	6FU

6FU, 6-month participant follow-up questionnaire; ECRQ, Effective Consultation and Reassurance Questionnaire; I, initial participant questionnaire; IMD, Index of Multiple Deprivation; KOOS-PS, Knee Injury and Osteoarthritis Outcome Score-Physical Function Short-Form; MDC, minimal data collection; MF, monthly participant follow-up questionnaire; MSK-HQ, Musculoskeletal Health Questionnaire; NDI, Neck Disability Index; SF-12, Short Form questionnaire-12 items; SPADI, Shoulder Pain and Disability Index; TSK-11, Tampa Scale for Kinesiophobia-11 item version.

## The Keele STarT MSK Tool<sup>®</sup> Clinician-completed version

For questions 1-9, think about just the last two weeks:

### Pain intensity

1) On average, how intense was your pain? [where 0 is “no pain”, 10 is “pain as bad as it could be”]

0	1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/> 0	<input type="checkbox"/> 0	<input type="checkbox"/> 0	<input type="checkbox"/> 0	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 3

Yes No

<b>Pain self-management</b> 2) Have you been struggling to manage or control this pain by yourself? (e.g. using medication or exercises etc...)	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Pain impact</b> 3) Over the last 2 weeks, have you been bothered a lot by your pain?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Walking short distances only</b> 4) Have you only been able to walk short distances because of your pain?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Pain elsewhere</b> 5) Are you having troublesome pain in more than one part of your body?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Long-term expectations</b> 6) Are you concerned you're developing a long-term problem?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Other important health problems</b> 7) Are you also having to deal with other important health problems at present?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Emotional well-being</b> 8) Have you felt anxious or low in your mood because of your pain?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Fear of harm</b> 9) Do you worry that physical activity could make your condition worse?	<input type="checkbox"/> 1	<input type="checkbox"/> 0
<b>Pain duration</b> 10) Have you had your current pain problem for 6 months or more?	<input type="checkbox"/> 1	<input type="checkbox"/> 0

Red text represents coding applied if box ticked

Total maximum score = 12; 0–4 = Low Risk, 5–8 = Medium Risk, 9–12 High Risk.

**FIGURE 3** The Keele STarT MSK Tool developed in work package 3 (clinician-completed version).

## Appendix 4 Stratified care intervention developed in work package 4

### BOX 3 Per-protocol recommended matched treatment options for each risk subgroup used in work package 3

#### Low risk matched treatment options

- Musculoskeletal education and advice: verbal.
- Musculoskeletal education and advice: written.
- Advice on OTC medication (simple oral and topical medications limited to those that would be available over the counter).

Per protocol: Selected at least one low-risk option with none of the medium- or high-risk options.

#### Medium risk matched treatment options

- Musculoskeletal education and advice: verbal.
- Musculoskeletal education and advice: written.
- Refer to musculoskeletal interface clinic.
- Refer to physiotherapy.
- Sign-post to locally available exercise programme.
- Sign-post to expert patient or peer support group.
- Sign-post to lifestyle interventions (e.g. dietitian or Slimming World, etc.).
- Prescribe opioid medication (consider weak opioid if acute pain as alternative to NSAIDs).
- Arrange corticosteroid injection (not recommended for neck or back pain).

Per protocol: Selected at least one medium risk option with no high risk option.

#### High risk matched treatment options

- Musculoskeletal education and advice: verbal.
- Musculoskeletal education and advice: written.
- Refer to musculoskeletal interface clinic.
- Refer to physiotherapy.
- Sign-post to locally available exercise programme.
- Sign-post to expert patient or peer support group.
- Sign-post to lifestyle interventions (e.g. dietitian or Slimming World, etc.).
- Prescribe opioid medication (consider weak opioid if acute pain as alternative to NSAIDs).
- Arrange corticosteroid injection (not recommended for neck or back pain).
- Refer to pain management service
- Refer to secondary care
- Refer to imaging
- Prescribe atypical analgesia (e.g. amitriptyline, pregabalin and gabapentin). Consider if neuropathic pain.
- Address comorbidities, distress and frailty.

Per protocol: Selected at least one high risk treatment option or referral to physiotherapy/musculoskeletal interface clinic with tool subgroup information within their referral so that services were aware that an onward referral to a high-risk treatment option might be required.



# Appendix 5 Summary health economic findings from main trial

## Detailed Health Economics Report for the main trial

### Overview

The within-trial cost-utility analysis estimated the cost-effectiveness of stratified care (SC) versus with usual care (UC) for adults consulting in primary care with the five most common MSK pain presentations. Costs were expressed in British pounds sterling (2019 price year) and health outcomes in quality-adjusted life-years (QALYs). The base case analysis was based on the intention-to-treat population and conducted from the perspective of UK National Health Service and Personal Social Services (NHS/ PSS). The time horizon covered the period from randomisation to end of follow-up at 6 months post-randomisation. Costs and outcomes were not discounted as the trial was limited to 6 months follow-up. Trial findings are reported in accordance with the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) (Husereau et al 2013).<sup>44</sup>

### Methods

#### Resource use and costs

Resource use and cost data were collected on: i) health care resource use during the 6 month follow-up period ii) broader societal resource use and costs including private healthcare and productivity loss related to me off work and reduced productivity over 6 months. Categories of resource use collected over the 6 months post randomisation period included: 1) primary and community care contacts (face-to-face general practice doctor, practice nurse, community therapy, and other primary contacts); 2) hospital based services including consultants, outpatient appointments, physiotherapy, inpatient admissions, diagnostic tests, scans and surgical procedures; 3) prescribed medication use from medical record review 4) private costs incurred by patients including private physiotherapy and other private medical use; and 5) time off work related to their musculoskeletal problem and reduced work performance (presenteeism) (Kigozi et al 2014).<sup>1a</sup>

Unit costs were obtained from a number of standard sources. Prescription data and costs were obtained from medical record review databases and the British National Formulary (BNF, 2019).<sup>2a</sup> Primary and community health social services were derived from the PSSRU Unit Costs of Health and Social Care compendium, whilst unit costs for hospital based services, diagnostics and scans were obtained primarily from NHS Reference costs (Curs and Burns, 2019; DoH 2019).<sup>3a,4a</sup> Resource use and cost summary statistics were generated by treatment group over 6 months follow-up. In order to estimate productivity costs, self-reported days off work were multiplied by the average wage rate (Office for National Statistics Annual survey of hours and earnings).<sup>5a</sup> The analysis used the human capital approach (Krol and Brouwer, 2014).<sup>6a</sup> Total health care costs over the study period were calculated by multiplying the resource items used by the respective unit cost and summing over all items. Between group differences were compared using generalised linear models adjusting for clustering. Bootstrapped 95% confidence intervals for the between-group differences in cost estimates were also reported.

#### Measurement of outcomes

Health-related quality of life was assessed at baseline and 6 months post randomisation using the EuroQol EQ-5D-5L questionnaire. In line with current guidelines, responses were converted into index scores using the interim cross-walk value set for mapping from the EQ-5D-5L to the EQ-5D-3L (Van Hout et al 2012).<sup>7a</sup> Quality-adjusted life-years (QALYs) were generated for each patient using the area under the baseline-adjusted utility curve, assuming linear interpolation between two follow-up me

points (Manca et al 2005).<sup>8a</sup> EQ-5D values and QALYs over 6 months were then reported by treatment group and presented as means and standard deviations. Between group differences were compared using multilevel regression modelling techniques. QALYs were adjusted for clustering, baseline EQ-5D, age and gender variables in the multivariate regression model.

### Statistical Analysis

The statistical analysis was conducted on an intention-to-treat basis, and in accordance with current cRCT guidelines (Gomes et al 2012; Ng et al 2013).<sup>9a,10a</sup> A multi-level modelling statistical approach taking into consideration clustering in cost and effect data was adopted. Missing EQ-5D-5L and cost data were imputed using multiple imputation techniques (Schafer, 1999)<sup>11a</sup> in order to ensure that all trial participants were included in the final analysis. The imputation was performed by the predictive mean matching method to account for the non-normality of the distribution of costs and EQ-5D values for missing total costs and EQ-5D items (Schafer, 1999).<sup>12a</sup> The imputed dataset informed the base-case and all subgroup and sensitivity analyses, with the exception of the complete-case analysis. An imputation included 25 imputed datasets and Rubin's rule was used to combine the imputed datasets into one final imputed variable (Rubin and Schenker, 1991).<sup>13a</sup> Statistical analysis was performed using Stata V.16 (StataCorp, 2019).<sup>14a</sup>

### Base case cost-effectiveness analysis

Separate generalised equation models, controlling for clustering were used to estimate the mean incremental costs and QALYs for SC relative to UC. Uncertainty around these estimates was estimated using cost-effectiveness acceptability curves (CEACs), which link the probability of SC being cost-effective to a range of potential threshold values ( $\lambda$ ) that the health system may be willing to pay for an additional QALY gained (see [Figures 4–6](#)). CEACs were estimated using a NMB regression approach (Gomes et al 2012).<sup>10</sup> NMB was defined as  $\lambda \times (\Delta \text{effect}_i) - \Delta \text{cost}_i$ , where  $\Delta \text{effect}_i$  is the incremental person-level outcome associated with the stratified care intervention, and  $(\Delta \text{cost}_i)$ , the additional costs due to stratified care, and  $\lambda$ =willingness to pay per unit of outcome gain. Using the output, we plotted CEACs, to provide a graphical display of the probability that stratified care is cost effective across a wide range of cost-effectiveness thresholds. Results for probability of cost-effectiveness were reported for the £20,000 per QALY threshold, currently used by NICE to determine cost-effectiveness of interventions for the NHS (NICE, 2013).<sup>15a</sup>

### Sensitivity analyses

A number of sensitivity analyses were conducted to investigate additional methodological and data collection aspects, and these included: i) A complete-case analysis to investigate bias on the findings resulting from missing data; ii) broadening the perspective of the analysis to healthcare and societal perspectives by capturing private costs and wider societal costs and their impact on relative cost-effectiveness of the interventions; iii) additional exploratory subgroup analysis to investigate the cost-effectiveness of stratified care compared with usual care for participants in each subgroup (low, medium and high risk). Details of the three subgroups were reported in the final report.

## Results

### Response rates and data completion

Resource use and cost per participant are reported by category, for all those with complete data at 6 months. Estimates for the base-case NHS/PSS perspective were based on the imputed participants' data for NHS cost and EQ-5D data ( $n=1211$ ). The proportion of returned questionnaires at 6 months follow-up is reported (See [Table 17](#): Descriptive and incremental health outcomes over 6 months for the base-case analysis and the complete case analyses). Complete cost data were available for 978(81%) participants at 6 months, while outcome data were available for 1070 (88%) including MDC data at 6 months. After excluding missing EQ-5D items, 1019 (84%) participants had a valid EQ-5D-5L score at 6 months.



## Health-related quality of life outcomes

Health-related outcomes over 6 months are reported in [Table 17](#). At baseline, participants in the SC arm had higher EQ-5D scores compared to those in UC (0.5572 and 0.5542 respectively) and a slightly higher score at 6 months than those in the UC group (0.6715 and 0.6512 respectively). At 6 months, the mean adjusted QALY difference between the two groups was 0.0041 (-0.0013, 0.0095) in favour of SC, and this difference was not significant. Adjustments were made for clustering, age, gender and baseline EQ-5D. There was overall improvement over the 6 months reflected by increased EQ-5D scores from baseline through to 6 months in both trial arms.

## Primary care consultations and medications

There were minimal differences in primary resource use and costs between the two groups, but overall, the primary care costs for UC were slightly higher than SC (see [Tables 8](#) and [9](#)). Costs related to prescribed medication were obtained from medical records. The mean prescription costs of prescribed medications were slightly lower for SC than UC (£1.07(5.10), £2.44 (11.60)), respectively (see [Table 9](#)).

## Hospital-based care

Hospital-based care resource use and costs of health professional contacts were very similar between the two groups with the exception of NHS physiotherapy treatment in the SC arm which was higher than UC (£95.18 and £41.00 respectively) (see [Tables 8](#) and [9](#)).

## Private healthcare use

The proportions of individuals reporting private healthcare use are given in [Table 8](#) (Resource components analysed in the trial) and [Table 9](#) (Cost components analysed in the trial). The biggest area of difference in resource use and cost was in the use of private physiotherapy care among those allocated to UC (£12.44 and £21.43 respectively).

## Imaging and other tests

Information about resource use and costs associated with hospital tests and investigations is shown in [Tables 8](#) and [9](#). Data from the procedures and investigations showed more UC participants received NHS scans and injections than SC, with costs of overall NHS investigations and treatments £31.61 and £61.13 for SC and UC respectively.

## Work-related outcomes

Results regarding paid employment, work status, MSK-related work absence and reduced productivity are reported in [Table 10](#) (Productivity costs analysed in the trial). Overall, reduced productivity (presenteeism) was very similar between the two groups (3.41 and 3.57 for SC and UC respectively) and the costs of work absence were slightly less in SC compared to UC (£698 and £738 respectively).

## Incremental costs and cost-effectiveness analysis

[Table 11](#) (Cost-Utility analysis using the net-benefit regression approach) also provides a summary of the incremental costs, outcomes, net monetary benefit results and cost-effectiveness analysis using data from the base-case NHS perspective. Results from the healthcare and societal perspective are also reported.

The results show that, although the SC and UC costs to the NHS perspective were very similar between the groups, the cost of SC was slightly higher overall than the UC cost (SC: £356.36 and UC: £343.44 difference £6.85 (95% CI: (-107.82, 121.54)), and was associated with minimal health gains of 0.0041 QALYs (-0.0013, 0.0094). The net monetary benefit (NMB) was £132 if society's willingness to pay for a QALY ( $\lambda$ ) is valued at £20,000 (see [Figure 4](#); Incremental net benefits (INB) of SC vs UC (95% CI) (Base-case) and [Figure 6](#); Incremental net benefits (INB) of SC vs UC (95% CI) complete case). Uncertainty around the values is illustrated in [Table 12](#) (Cost-effectiveness results for the within-trial economic analysis with 6-month horizon) for the NHS perspective. The uncertainty analysis showed that SC was more costly and slightly more effective, and is likely to be cost-effective at the £20,000 per QALY

threshold range with a 0.73 probability for the NHS perspective. The CEACs show the probability that the SC intervention is cost-effective at different levels of willingness-to-pay for a QALY (See [Figure 3](#); Cost Effectiveness Acceptability Curve (Base-case) and [Figure 5](#); Cost Effectiveness Acceptability Curve (Complete-case) analysis). Uncertainty around the values is illustrated in [Table 12](#) for the healthcare and societal perspective.

## Sensitivity analyses

### *Healthcare and societal perspectives*

Details of the sensitivity analysis from alternative perspectives are presented in [Table 11](#) and [12](#). The results between the two groups showed that SC and UC were very similar, but overall SC was slightly more costly than UC from a societal perspective. The probability that SC was cost-effective at £20,000 per QALY was 0.76 and 0.61 from a healthcare and societal perspective respectively.

### *Complete-case analysis*

Under this scenario from an NHS perspective, SC was slightly more costly (£11.61) but generated more QALYs on average (Diff, CI: 0.0053; -0.0005, 0.0112) when the analysis was restricted to participants with complete cost and outcome data ([Table 12](#)). The net monetary benefit value was £162 ( $\lambda$ =£20,000); and the ICER was £2,190 per QALY gained. The probability of SC being cost-effective at  $\lambda$  = £20,000 was 0.75. The details of the incremental cost-effectiveness and uncertainty analysis are reported in [Table 12](#).

## Subgroup analysis

### *Impact of stratified care on health-related quality of life by study group*

Quality of life data (EQ-5D-5L and QALY scores) at baseline and 6 months follow-up by treatment arm and MSK subgroup are provided in [Table 13](#) (Subgroup analysis: Costs and health outcomes mean (SD) scores by STarT MSK risk subgroups (Imputed analysis). At baseline and at the end of 6 month follow-up, participants in the SC arm had slightly lower EQ-5D scores compared to UC, in low risk and medium risk, but higher in high risk group. At 6 months, QALY outcomes in the low risk and medium risk groups slightly favoured UC: (difference SC-UC: -0.0015, -0.0001 respectively) and favoured SC for the high risk group (0.0117) with all CIs crossing zero.

### *Impact of stratified care on costs and cost-effectiveness results by study group*

Point estimates of incremental NHS costs, QALYs and ICERs by risk group are reported in [Table 13](#). Results for NHS perspective costs showed that SC was slightly more expensive for participants in the low risk and medium risk groups but slightly cheaper for those in the high-risk group, due to the higher number of spinal injections and scans in the UC arm. Relatively similar results were observed in the healthcare perspective estimates incorporating private healthcare resource use. From a societal perspective, SC was more costly than UC due to higher productivity costs in the medium and high-risk groups.

There was some uncertainty due to the reduced sample size in each group, but overall SC was associated with slightly fewer QALYs on average, and was slightly more expensive than UC for the low and medium risk group, and was therefore dominated by UC from an NHS perspective. For those in the high-risk group however, SC dominated UC over 6 month follow-up, that is, better health benefits (0.0129 additional QALYs) and slightly reduced NHS costs (-£65.96) (NICE, 2013).<sup>15a</sup> At a willingness-to-pay threshold of £20,000 per QALY, the probability that SC is cost-effective compared with UC from an NHS perspective was 0.23, 0.29 and 0.87 in the low, medium and high-risk groups, respectively (See [Table 14](#); Cost-effectiveness outcomes for STarT MSK risk subgroups). From healthcare and societal perspectives the probability that SC is cost-effective compared with UC was 0.16, 0.40, and 0.89 and 0.02, 0.47, 0.69 respectively (See [Table 14](#)). Further details of the incremental cost-effectiveness analysis and uncertainty analysis of the subgroup analysis are reported in [Table 14](#).

### Overall summary

After adjusting for clustering in costs and outcomes, SC was associated with increased healthcare costs of £6.85 but with slightly better QALY outcomes (0.0041 health gain); the ICER for SC was £1,670 per QALY gained. At a willingness-to-pay threshold of £20,000 per QALY, the incremental net monetary benefit was £132 and the probability of SC being cost-effective was approximately 73%. In the context of the findings in this economic analysis, the SC intervention is potentially a cost-effective option from an NHS and healthcare perspective. However, in subgroup analyses, SC was only likely to be cost-effective in the high risk subgroup.

### Strengths and limitations

The economic analysis was based on a large sample (n=1211) with resource use information from a combination of self-reported data, including information outside the main NHS perspective and GP records prescription data, and therefore reports comprehensive resource use data. Also, the analysis was performed using recommended statistical approaches for analysing cost-effectiveness data alongside cluster trials. However, there were some limitations. Resource use data were primarily obtained from self-report data. A limitation with this approach is that respondents could potentially under-report utilisation, particularly over longer periods of recall (Petrou et al 2002).<sup>16a</sup> Additionally, EQ-5D data were only collected at two time-points; therefore changes in quality of life related to interventions at intermediate time points could not be incorporated into the QALY calculation.

### References

- 1a. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.* on behalf of the ISPOR Health Economic Evaluation Publication Guidelines - CHEERS Good Reporting Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) - Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013; **16**:231–50.
- 2a. Kigozi J, Lewis M, Jowett S, Barton P, Coast J. Construct Validity and Responsiveness of the single-item presenteeism question in patients with lower back pain for the measurement of presenteeism. *Spine* 2014; **39**:409–16.
34. BNF, 2019. Joint Formulary Committee. *BNF - British National Formulary*. London: BMJ Group and Pharmaceutical press; 2019.
- 4a. Curtis L, Burns A. *PSSRU: Unit Costs of Health & Social Care*. Canterbury: Personal Social Service Research Unit (PSSRU), University of Kent; 2019.
- 5a. Department of Health. *National Schedule of Reference Costs: 2018–2019*. London: Department of Health; 2019.
- 6a. Office for National Statistics. *Annual survey of hours and earnings*. URL: <http://www.ons.gov.uk>. 2017. (Accessed 1 March 2019).
- 7a. Krol M, Brouwer W. How to Estimate Productivity Costs in Economic Evaluations. *PharmacoEconomics* 2014; **4**:335–44.
- 8a. van Hout B, Janssen M, Feng V, Kohlmann T, Busschbach J, Golicki D, *et al.* Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012; **15**:708–15.
- 9a. Manca A, Hawkins N, Sculpher M. Estimating mean QALYs in trial-based cost-effectiveness analysis: The importance of controlling for baseline utility Article in Health Economics. *Health Econ* 2005; **14**:487–96.
- 10a. Gomes M, Ng ESW, Grieve R, Nixon R, Carpenter J, and Thompson SG: Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Decis Making* 2012, **32**:350–361. OpenURL

- 11a. Ng ES, Diaz-Ordaz K, Grieve R, Nixon RM, Thompson SG, Carpenter JR. Multilevel models for cost-effectiveness analyses that use cluster randomised trial data: an approach to model choice. *Stat Methods Med Res* 2013. doi: 10.1177/0962280213511719.
- 12a. Schafer J. Multiple imputation: a primer. *Stat Methods Med Res* 1999; **8**:3–15.
- 13a. Rubin D, Schenker N. Multiple imputation in health-care databases: An overview and some applications. *Stat Med* 1991; **10**:585–98.
- 14a. StataCorp. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC. 2019.
- 15a. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal*. London: NICE; 2013.
- 16a. Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic studies. *Int J Technol Assess Health Care* 2002; **18**:705–10.

**TABLE 8** Descriptive and incremental health outcomes over 6 months for the base-case analysis and the complete case analyses. Values are mean (SD) scores unless stated otherwise

	Total questionnaires returned	Total questionnaires returned (Including MDC)	
6 month follow-up	978 (80.8%)	1070 (88.4%)	
Health outcomes	SC n =534	UC n =677	Mean difference <sup>b</sup> (95% CI)
Primary (Imputed) analysis <sup>a</sup>			
Baseline EQ-5D	0.5572 (0.2293)	0.5542 (0.2349)	(-0.04530.0011, 0.0476)
6-month EQ-5D	0.6715 (0.2192)	0.6512 (0.2308)	(-0.02330.0166, 0.05662)
Unadjusted QALYs	0.3072 (0.1012)	0.3014 (0.1042)	(-0.01340.0054, 0.0242)
Adjusted QALYs <sup>c</sup>	0.3063	0.3021	0.0041 (-0.0013, 0.0095)
Complete-case analysis <sup>d</sup>	n = 445	n = 574	
Unadjusted QALYs	0.3095 (0.1034)	0.3032 (0.1063)	(-0.01410.0068, 0.0277)
Adjusted QALYs <sup>c</sup>	0.3080	0.3027	0.0053 (-0.0005, 0.0112)

QALYs=quality-adjusted-life-years.

a Base-case imputed dataset

b Difference=SC-UC by generalized linear latent and mixed (gllamm) models adjusting for clustering. Reported CIs were generated using generalised linear regression methods.

c Adjusted for baseline utility

**TABLE 9** Resource components analysed in the trial. Values are mean (SD) resource per patient, by treatment group, for patients providing health care utilisation data at 6 months (n=978), unless stated otherwise

Resource use component	SC (n=415)	UC (n=563)
Primary care contacts:		
Primary care general practitioner	0.96 (1.60)	1.15(2.11)
Primary care nurse	0.08 (0.57)	0.13 (0.65)
Primary care physiotherapist	0.35 (1.67)	0.29 (1.17)
Hospital-based care:		
NHS consultant	0.33 (1.32)	0.37 (0.99)
NHS other physiotherapist	1.64 (2.64)	0.71 (1.87)
NHS acupuncturist	0.06 (0.59)	0.03 (0.44)
NHS osteopath	0.002 (0.05)	0.02 (0.22)
Private consultant	0.02 (0.23)	0.06 (0.40)
Private physiotherapist	0.21 (1.24)	0.37 (1.99)
Private acupuncturist	0.06 (0.55)	0.06 (0.42)
Private osteopath	0.29 (1.28)	0.26 (1.57)
Other Hospital-based care (n, %)		
NHS surgery	10 (2.2%)	11 (2.0%)
Shoulder	2	1
Neck	0	1
Spine/Back	1	1
Hip	1	4
Knee	6	4
Private surgery	1(0.2%)	2 (0.4%)
Back	1	0
Knee	0	1
Shoulder	0	1
NHS Scans	13 (3.1%)	27 (4.8%)
NHS MRI investigations	64 (15.4%)	67 (11.9%)
NHS Blood tests	1 (0.2%)	5 (0.9%)
NHS injections	4 (1.0%)	24 (4.3%)
Private Scans	1 (0.2%)	1 (0.2%)
Private MRI investigation	12 (2.9%)	13 (2.3%)
Private Injections	1 (0.2%)	0
SC=Stratified Care; UC=Usual Care.		

**TABLE 10** Cost components analysed in the trial. Values are mean (SD) costs per patient, by treatment group, for patients providing health care utilisation data at 6 months (n=978), unless stated otherwise

Cost component	SC (n=415)	UC (n=563)
Primary care contacts:		
Primary care general practitioner	31.81 (52.81)	37.87 (69.58)
Primary care nurse	3.14 (23.77)	5.52 (27.11)
Primary care physiotherapist	20.40 (96.94)	17.10 (68.11)
Prescriptions	1.07 (5.10)	2.44 (11.60)
Hospital-based care		
NHS consultant	51.72 (209.25)	59.31 (158.67)
NHS other physiotherapist	95.18 (152.96)	41.00 (108.72)
NHS Acupuncturist	3.49 (34.15)	1.95 (25.702)
NHS Osteopath	0.14 (2.847)	0.93 (12.67)
NHS other professionals	3.54 (41.36)	5.25 (78.07)
Private consultant	3.83 (36.45)	9.32 (64.00)
Private Physiotherapist	12.44 (72.11)	21.43 (115.565)
Private Acupuncturist	3.63 (31.79)	3.50 (24.21)
Private Osteopath	16.63 (74.38)	15.25 (91.05)
NHS private other	2.65 (33.98)	2.37 (32.27)
Other hospital-based care		
NHS surgery <sup>†</sup>	110.07 (728.99)	106.34 (808.99)
NHS investigations and treatments <sup>††</sup>	31.61(86.22)	61.13(254.14)
Private surgery <sup>†</sup>	2.80 (57.04)	16.38 (284.23)
Private investigations/treatments <sup>††</sup>	6.86 (63.76)	3.93 (31.63)

# Between-group difference in mean scores (SC– UC) by generalized linear latent and mixed (gllamm) models adjusting for clustering,

†† Includes investigations such as MRI scans, CT scans and ultrasound scans, and injections; †Includes shoulder, neck, spine, back, hip, knee injury. SC=Stratified Care; UC=Usual Care.

**TABLE 11** Productivity costs analysed in the trial. Values are mean (SD) costs per patient, by treatment group, for patients providing work outcomes data at 6 months

Work-related outcomes <sup>a</sup>	SC	UC
Baseline: working in paid employment (n, %)	275 (53.8)	286 (43.3)
Baseline: reported me off work during the last 6 months (n, %)	91 (30.6)	101 (30.2)
Working in paid employment at 6-months (n, %)	177 (44.1)	219 (39.8)
Performance at work at 6 months (mean, SD) <sup>b</sup>	3.41 (2.71)	3.57 (2.89)
Reported me off work at 6-months (n, %)	47 (23.6)	57 (23.2)
Days off-work during the last 6 months (mean, SD) <sup>a</sup>	6.13 (20.05)	6.48 (19.66)
Mean difference <sup>c</sup> (95% CI)	-1.09 (-5.87, 3.68)	
Productivity costs during the last 6 months (mean, SD) <sup>a</sup>	698.81 (2285.24)	738.14 (2240.95)
Mean difference <sup>d</sup> (95% CI)	-124.89 (-669.73, 419.93)	

a The evaluation of work-related outcomes and the estimation of indirect costs focussed on the subsample of respondents in paid employment at 6 months (396/1211).

b Mean performance at work on a scale of 0 to 10 where 0 indicates work performance not affected.

c Mean difference of reported days off-work over 6 months and adjusted for paid employment at baseline and clustering.

d Productivity costs obtained from days off-work over 6 months and adjusted for paid employment at baseline and clustering.

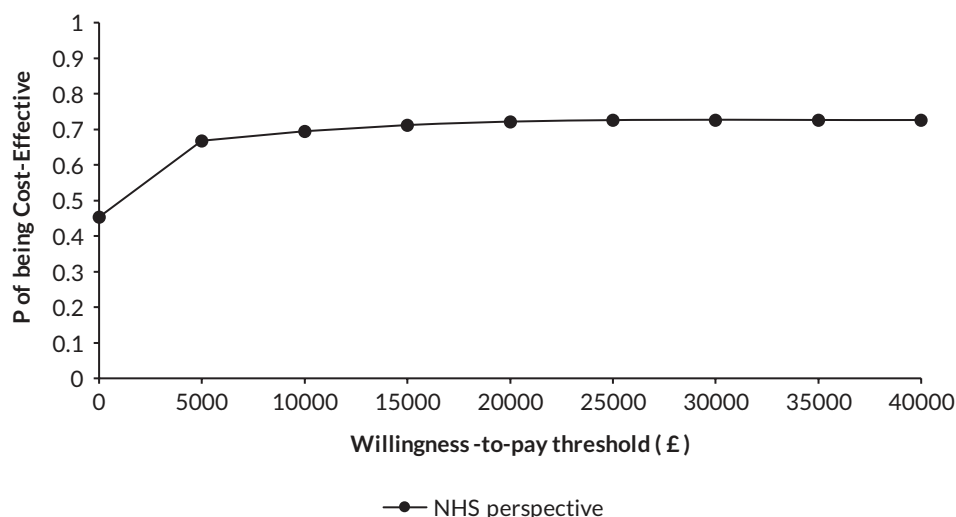


**TABLE 12** Cost-Utility analysis using the net-benefit regression approach<sup>†</sup>

	SC N=534	UC N=677	Mean differences or ICER (95% CI's)
Imputed dataset analysis			
<b>Cost Analysis<sup>†</sup></b>			
NHS cost (£) Mean (Standard Deviation)	356.36 (864.01)	343.44 (942.92)	6.85 (-107.82, 121.54)
Healthcare cost (£) Mean (Standard Deviation)	411.17 (890.161)	417.11 (1035.28)	-5.92 (-116.3216, 104.46)
Societal cost (£) Mean (Standard Deviation)	838.53 (2102.12)	762.05 (2028.20)	65.01 (-195.56, 325.59)
<b>Effectiveness Analysis<sup>†</sup></b>			
Unadjusted QALYs gained <sup>*</sup> Mean (Standard Deviation)	0.3072 (0.1012)	0.3014 (0.1042)	0.0053 (-0.0134, 0.0241)
Adjusted QALYs gained <sup>*</sup>	0.3065	0.3018	0.0041 (-0.0013, 0.0094)
Complete-case analysis			
<b>Cost Analysis<sup>†</sup></b>			
NHS cost (£) Mean (Standard Deviation)	351.99 (910.58)	338.89 (988.83)	11.61 (-118.84, 142.08)
Healthcare cost (£) Mean (Standard Deviation)	400.85 (940.16)	411.06 (1082.03)	-10.21 (-139.95, 119.53)
Societal cost (£) Mean (Standard Deviation)	762.07 (2195.45)	745.16 (2111.96)	19.05 (-295.94, 334.04)
<b>Effectiveness Analysis<sup>†</sup></b>			
Unadjusted QALYs gained <sup>*</sup> Mean (Standard Deviation)	0.3095 (0.1034)	0.3032 (0.1063)	0.0068 (-0.0141, 0.0277)
Adjusted QALYs gained <sup>*</sup>	0.3092	0.3033	(-0.00050.0053, 0.0112)
NHS, National Health Service, QALY, Quality-Adjusted Life Year.			
<sup>*</sup> Mean differences adjusted for clustering in cost and QALY outcomes			
<sup>†</sup> Complete-case dataset			

**TABLE 13** Cost-effectiveness results for the within-trial economic analysis with 6-month horizon

	Cost-effectiveness outcomes – NHS			Probability SC is cost- effective at cost-effectiveness threshold of		
	Mean incremental costs (95% CI), £	Mean incremental QALYs(95% CI)	ICER	£20,000 per QALY	£30,000 per QALY	£50,000 per QALY
Base-case analysis	6.85 (-107.82, 121.54)	0.0041 (-0.0013, 0.0094)	£1,670	0.73	0.73	0.73
Sensitivity analysis 1: Alternative-perspectives						
Healthcare perspective	-5.92 (-116.3216, 104.46)	0.0041 (-0.0013, 0.0094)	Dominant	0.75	0.75	0.74
Societal perspective	65.01 (-195.56, 325.59)	0.0041 (-0.0013, 0.0094)	£15,856	0.62	0.61	0.63
Sensitivity analysis 2: Complete-case analysis <sup>††</sup>						
NHS cost (£) Mean (SD)	11.61 (-118.84, 142.08)	0.0053 (-0.0005, 0.0112)	£2,190	0.75	0.74	0.73
Healthcare cost (£) Mean (SD)	-10.21 (-139.95, 119.53)	0.0053 (-0.0005, 0.0112)	Dominant	0.78	0.77	0.75
Societal cost (£) Mean (SD)	19.05 (-295.94, 334.04)	0.0053 (-0.0005, 0.0112)	£3,594	0.74	0.72	0.71
Adjusted for age, sex, treatment allocation, study site, impingement type, baseline health-related quality of life and baseline costs.						
<sup>†</sup> Mean ICERs in north east quadrant of the cost-effectiveness plane where SC is more costly and more effective. NHS, National Health Service, QALY, Quality-Adjusted Life Year.						
<sup>*</sup> Mean differences adjusted for clustering in cost and QALY outcomes.						
<sup>††</sup> Complete-case dataset. ICER = Incremental cost-effectiveness ratio; CI = confidence interval						

**FIGURE 4** Cost Effectiveness Acceptability Curve (Base-case).

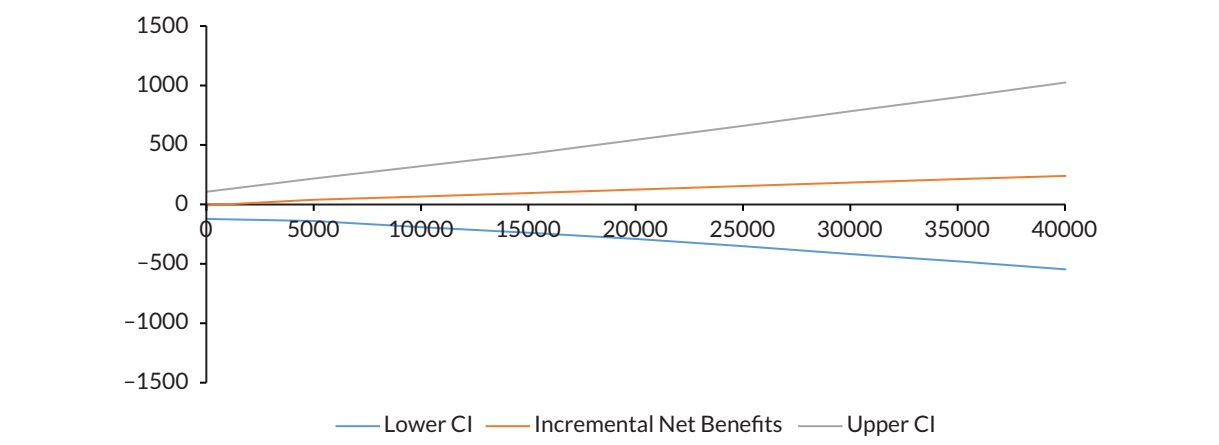


FIGURE 5 Incremental net benefits (INB) of SC vs UC (95% CI) (Base-case).

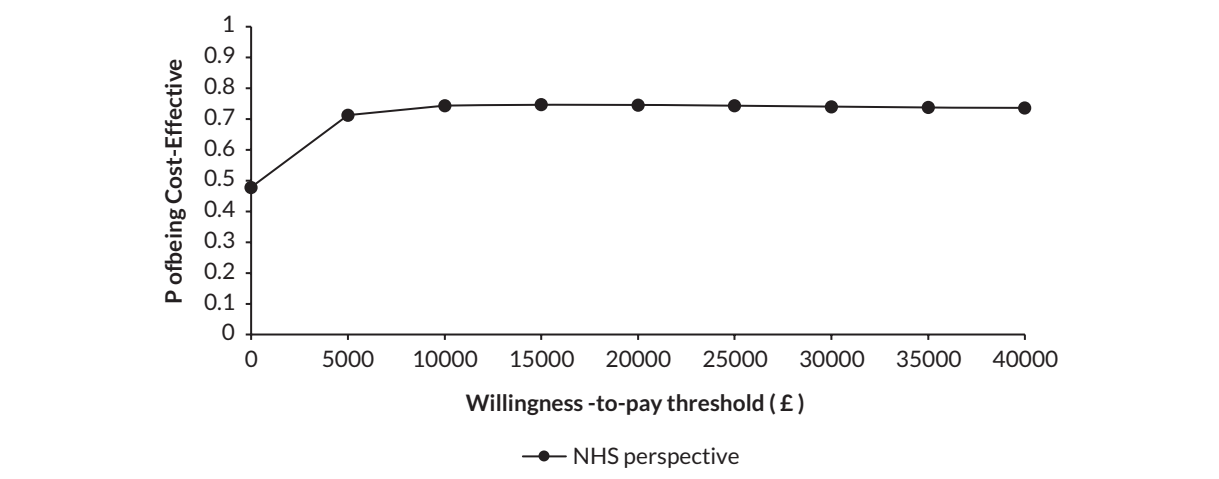


FIGURE 6 Cost Effectiveness Acceptability Curve (Complete-case) analysis.

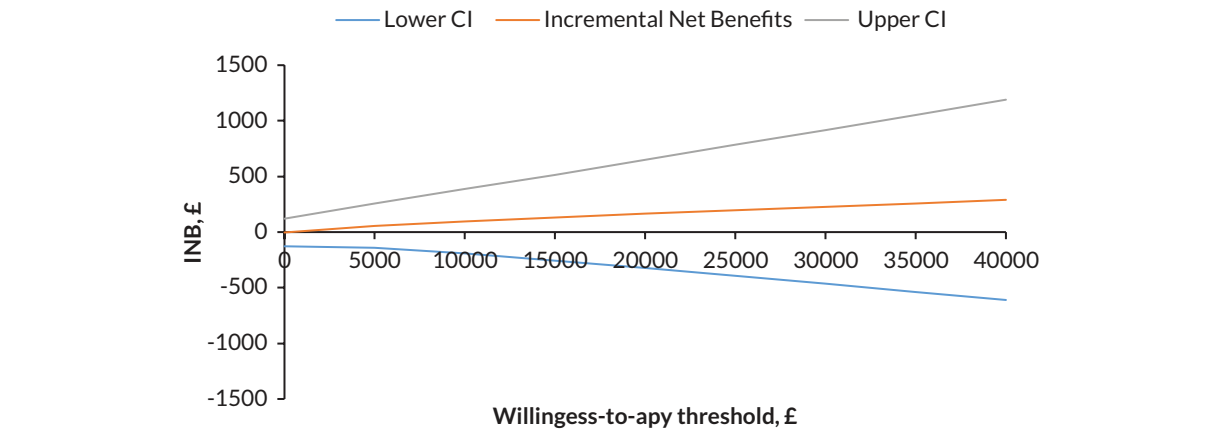


FIGURE 7 Incremental net benefits (INB) of SC vs UC (95% CI) complete case.

**TABLE 14** Subgroup analysis: Costs and health outcomes mean (SD) scores by STaT MSK risk subgroups (Imputed analysis)

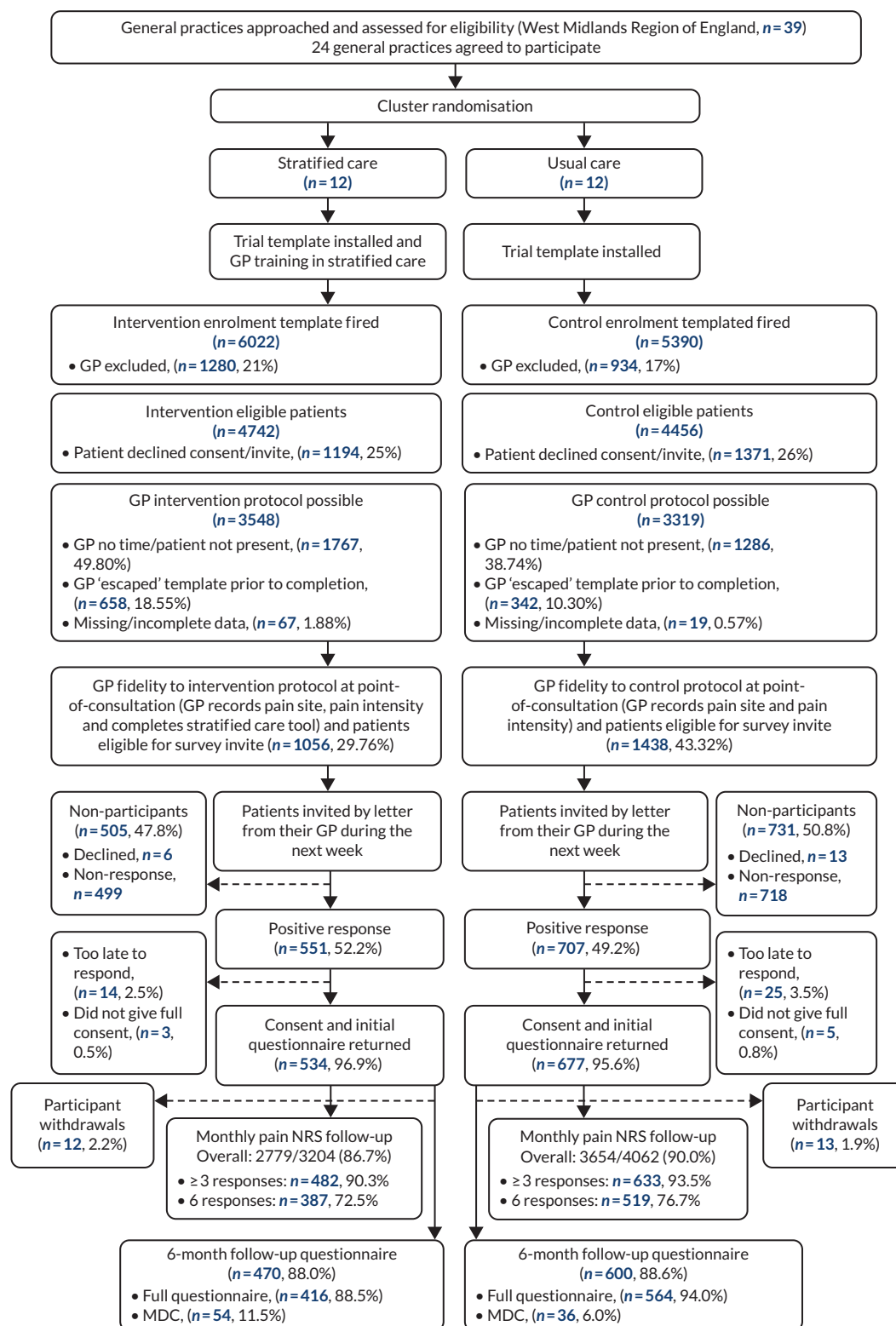
	Low risk		Medium risk		High risk	
	SC	UC	SC	UC	SC	UC
Total NHS cost	250.62 (800.71)	173.11 (586.11)	359.85 (764.10)	324.274 911.25	429.87 (1077.50)	495.83 (1193.57)
Mean difference (95% CI)	77.50 (-103.34, 258.35)		35.57 (-109.90, 181.06)		-65.96 (-296.69, 164.77)	
Total Health care cost	314.87 (846.92)	204.42 (600.38)	406.23 (793.68)	403.962 1041.25	493.95 (1092.04)	595.94 (1276.64)
Mean difference (95% CI)	110.45 (-78.38, 299.29)		2.27 (-158.55, 163.09)		-101.99 (-343.85, 139.87)	
Total Societal cost	586.49 (1730.83)	250.44 (642.97)	682.59 (1333.82)	759.47 (2128.87)	1293.22 (3089.00)	1175.37 (2510.19)
Mean difference (95% CI)	336.05 (9.74, 662.36)		-76.87 (-387.82, 234.07)		105.97 (-498.20, 710.14)	
Baseline EQ-5D	0.7446 (0.1076)	0.7603 (0.1146)	0.5895 (0.1726)	0.5986 (0.1628)	0.3979 (0.2473)	0.3885 (0.2432)
Month 6 EQ-5D	0.8113 (0.1442)	0.8177 (0.1260)	0.6983 (0.1778)	0.6986 (0.1609)	0.5595 (0.2487)	0.5088 (0.2619)
QALYs	0.3890 (0.0504)	0.3945 (0.0489)	0.3220 (0.0767)	0.3243 (0.0635)	0.2394 (0.1111)	0.2243 (0.1104)
Mean difference (95% CI)	-0.0068 (-0.0224,0.0088)		-0.0023 (-0.0143, 0.0097)		0.0129 (-0.0145,0.0404)	
Adjusted QALYs *	0.3917	0.3924	0.3233	0.3232	0.2375	0.2258
Mean difference (95% CI)	-0.0015 (-0.0115, 0.0085)		-0.0001 (-0.0069,0.0066)		0.0117 (0.0009,0.0225)	
SC = Stratified Care; UC = Usual Care						
* Incremental QALY estimates following multiple regression-based adjustment for age, gender and baseline EQ-5D						

**TABLE 15** Cost-effectiveness outcomes for STarT MSK risk subgroups (Imputed analysis)

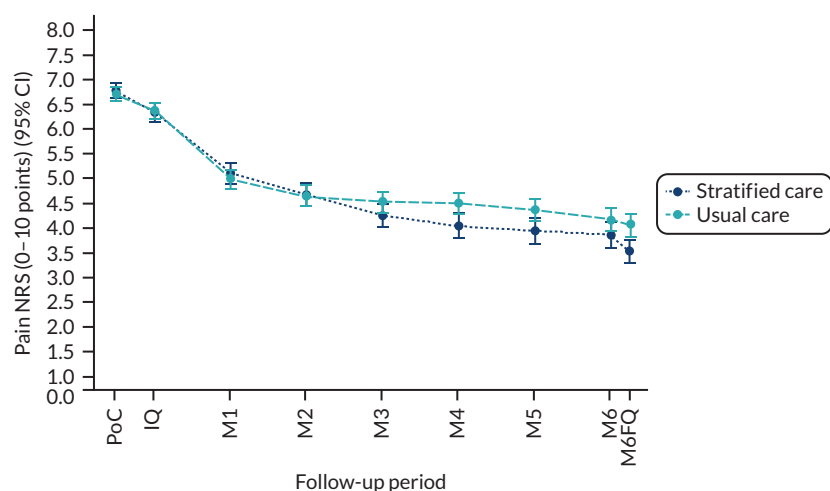
Cost-effectiveness outcomes – NHS				Probability Stratified care is cost-effective at cost-effectiveness threshold of		
NHS Perspective						
Subgroup analysis	Mean incremental costs (95% CI), £	Mean incremental QALYs(95% CI)	ICER	£20,000 per QALY	£30,000 per QALY	£50,000 per QALY
Low risk	77.50 (-103.34, 258.35)	-0.0068 (-0.0224,0.0088)	Dominated	0.23	0.21	0.19
Medium risk	35.57 (-109.90, 181.06)	-0.0023 (-0.0143, 0.0097)	Dominated	0.29	0.30	0.32
High risk	-65.96 (-296.69, 164.77)	0.0129 (-0.0145,0.0404)	Dominant	0.87	0.86	0.85
Healthcare Perspective						
Low risk	110.45 (-78.38, 299.29)	-0.0068 (-0.0224,0.0088)	Dominated	0.16	0.16	0.17
Medium risk	2.27 (-158.55, 163.09)	-0.0023 (-0.0143, 0.0097)	Dominated	0.40	0.38	0.37
High risk	-101.99 (-343.85, 139.87)	0.0129 (-0.0145,0.0404)	Dominant	0.89	0.88	0.86
Societal Perspective						
Low risk	336.05 (9.74, 662.36)	-0.0068 (-0.0224,0.0088)	Dominated	0.02	0.04	0.08
Medium risk	-76.87 (-387.82, 234.07)	-0.0023 (-0.0143, 0.0097)	£33,421	0.47	0.49	0.43
High risk	105.97 (-498.20, 710.14)	0.0129 (-0.0145,0.0404)	£8,214	0.69	0.73	0.63
Further breakdown of excluded/declined data (total=8918; Stratified Care=6022, Usual Care=5390): *GP no me (1866 (16%); 1171 (19%), 695 (13%))/ Patient not present (1187 (10%); 596 (10%), 591 (11%)); **Not clinically relevant based on GP opinion (1628 (14%); 985 (16%), 643 (12%))/ Vulnerable patient (302 (3%); 151 (3%), 151 (3%))/ Not trial-specific pain site consultation (163 (1%); 50 (1%), 113 (2%))/ Suspected serious pathology (121 (1%); 94 (2%), 27 (1%)); ***Incomplete data (57 (0.6%); 56 (0.9%), 1 (<0.1%))/ IT processing error (29 (0.3%; 11 (0.2%), 18 (0.3%)). \$ For the primary analysis if the last monthly SMS/brief questionnaire response was missing it was imputed using the corresponding pain response from the returned 6-month questionnaire (if completed within 20 days of the date of issue): this gave an overall number of available scores for the analysis of 2791/3204 (87.1%) for Stratified Care and 3668/4062 (90.3%) for Usual Care.						



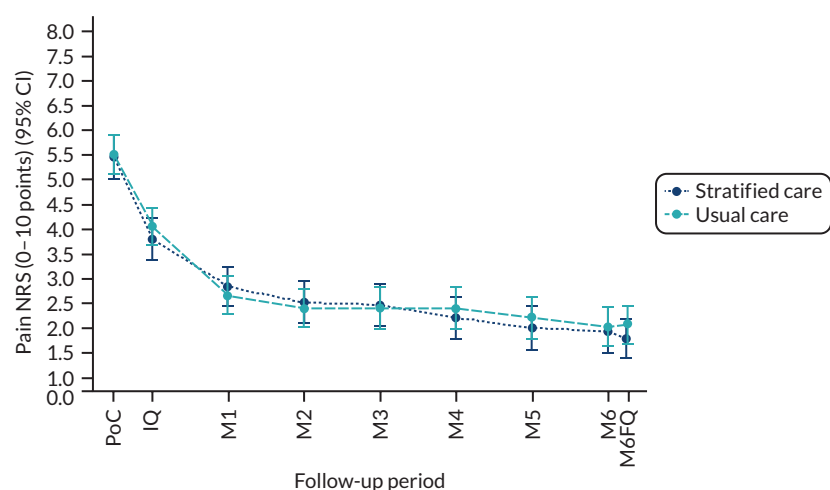
## Appendix 6 Figures and tables from main trial in work package 4



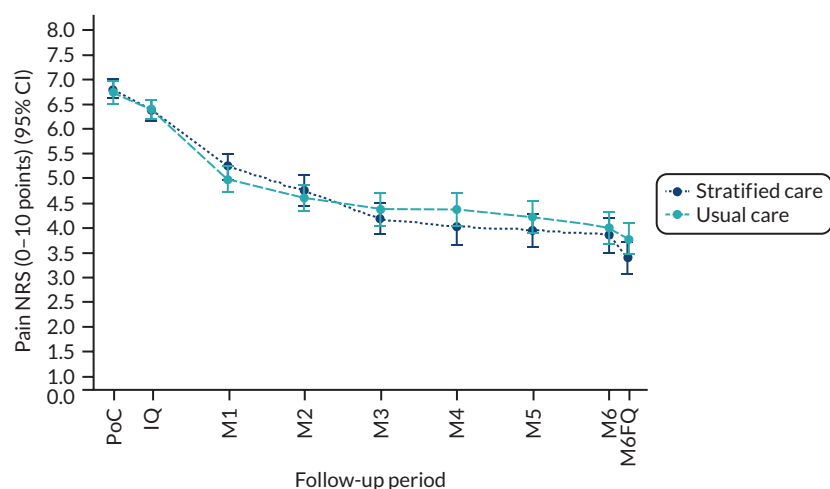
**FIGURE 8** Main trial flowchart from work package 4. MDC, minimum data collection.



**FIGURE 9** Overall pain intensity (NRS) scores per trial arm, over time. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.

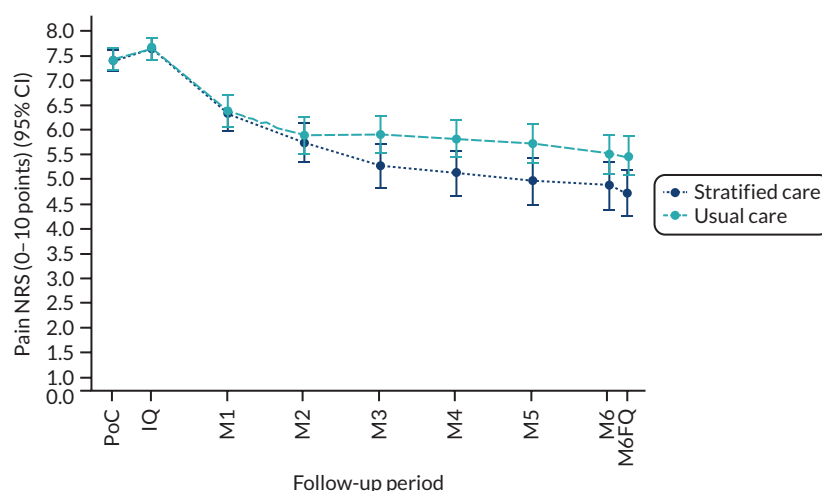


**FIGURE 10** Pain intensity (NRS) scores per trial arm for patients at low risk of poor outcome, over time. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.

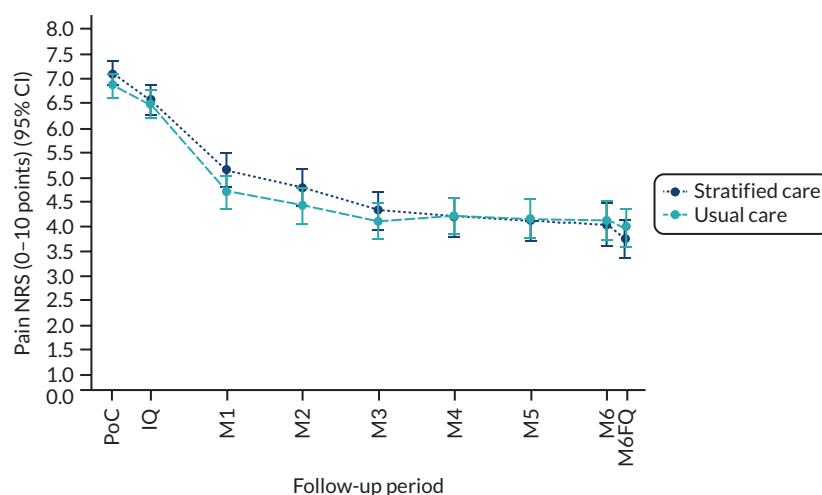


**FIGURE 11** Pain intensity (NRS) scores per trial arm for patients at medium risk of poor outcome, over time. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.

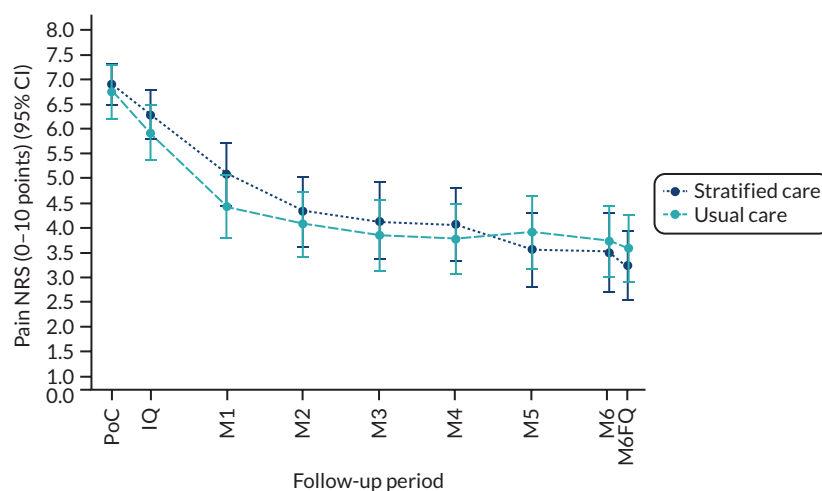




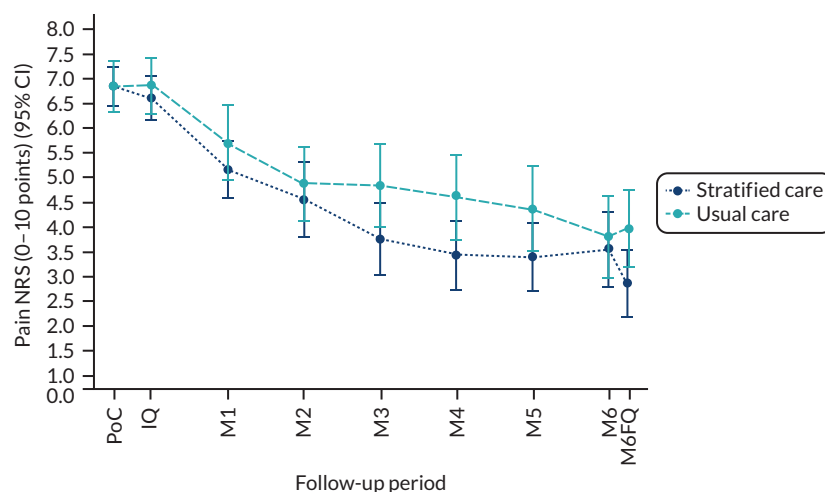
**FIGURE 12** Pain intensity (NRS) scores per trial arm for patients at high risk of poor outcome, over time. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.



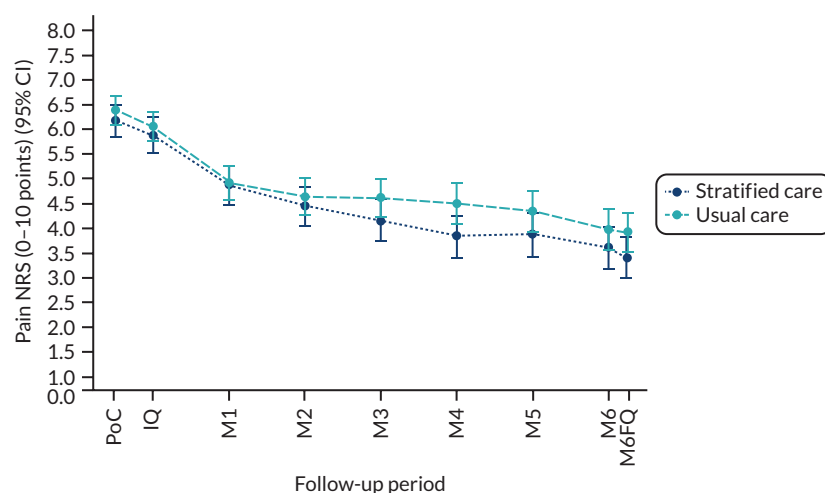
**FIGURE 13** Pain intensity (NRS) scores per trial arm for patients with back pain. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.



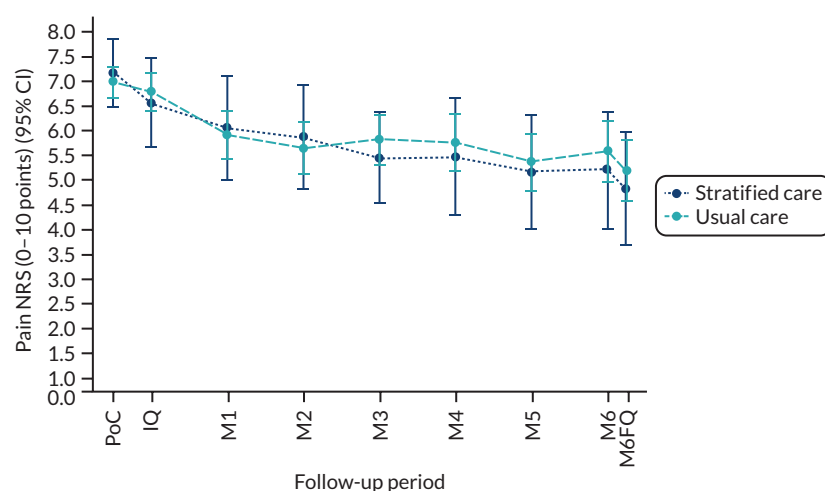
**FIGURE 14** Pain intensity (NRS) scores per trial arm for patients with neck pain. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.



**FIGURE 15** Pain intensity (NRS) scores per trial arm for patients with shoulder pain. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.



**FIGURE 16** Pain intensity (NRS) scores per trial arm for patients with knee pain. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.



**FIGURE 17** Pain intensity (NRS) scores per trial arm for patients with multisite pain. IQ, initial questionnaire; M1–6, months 1–6 SMS or brief questionnaire scores; M6FQ, month 6 follow-up questionnaire; PoC, point of consultation.

**Table 16** shows that there were some minor differences between the characteristics of patients who agreed to participate in data collection and those who did not. These differences were in musculoskeletal pain body site (patients with knee pain were more likely to participate than those with back pain) and risk subgroups (patients at medium risk were more likely to participate than those at low risk). There were no differences for the primary outcome, risk group allocation or the proportion of stratified care patients treated as per the intervention protocol.

**TABLE 16** Characteristics of participants vs. non-participants in the main trial

Characteristic	Participants (N = 1211)	Non-participants (N = 1283)	All invited patients (N = 2494)
Trial arm, n (%)			p = 0.085
Stratified care	534 (44.1)	522 (40.7)	1056 (42.3)
Usual care	677 (55.9)	761 (59.3)	1438 (57.7)
Pain site, n (%)			p < 0.001
Back	457 (37.7)	584 (45.5)	1041 (41.7)
Neck	129 (10.7)	151 (11.8)	280 (11.2)
Shoulder	130 (10.7)	117 (9.1)	247 (9.9)
Knee	379 (31.3)	327 (25.5)	706 (28.3)
Multisite	116 (9.6)	104 (8.1)	220 (8.8)
Pain score, mean (SD)	6.7 (2.0)	6.8 (1.9)	p = 0.119
<b>Stratified care arm only</b>			
Characteristic	Participants (N = 534)	Non-participants (N = 522)	All invited patients in the stratified care arm (N = 1056)
Treatment as per protocol, n (%) <sup>a</sup>			p = 0.219
No	113 (21.2)	127 (24.3)	240 (22.7)
Yes	421 (78.8)	395 (75.7)	816 (77.3)
Risk subgroup, n (%)			p = 0.043 (overall test) p = 0.109 (trend test)
Low risk	98 (18.4)	128 (24.5)	226 (21.4)
Medium risk	311 (58.2)	274 (52.5)	585 (55.4)
High risk	125 (23.4)	120 (23.0)	245 (23.2)
The Keele STarT MSK Tool score, mean (SD)	6.7 (2.4)	6.4 (2.5)	6.5 (2.4); p = 0.069

<sup>a</sup> Treatment as per protocol based on data GPs entered into the trial-specific electronic template.

#### Note

p-values by chi-squared test (test for trend as indicated) for categorical variables and t-test for numerical variables for the difference in pain scores and STarT MSK tool scores. **Table 17** shows that there were some differences in the population characteristics of general practices randomised to the stratified care intervention compared with usual care: as Crime and Living Environment deprivation were worse in those allocated to stratified care. The mean and median population sizes of practices and the overall Index of Multiple Deprivation were similar.

**TABLE 17** General practice characteristics per trial arm in the main trial

Key characteristics	Stratified care (n = 12)	Usual care (n = 12)
Registered population size		
Mean (SD)	8033 (3214)	7391 (3574)
Median (IQR)	7361.5 (6277.0–9954.5)	6981.5 (5616.5–8628.0)
Minimum, maximum	1994, 13248	2031, 13894
English indices of area deprivation, median (IQR) <sup>a</sup>		
Income deprivation	19,390.5 (14,171.0–25,490.5)	17,483.5 (10,423.5–25,405.5)
Employment deprivation	14,676.0 (11,254.5–24,009.5)	14,010.0 (9897.0–24,563.0)
Education, skills and training deprivation	20,409.0 (10,228.0–28,632.5)	18,106.0 (8492.5–25,586.5)
Health deprivation and disability	15,656.0 (9502.0–24,854.0)	17,587.5 (12,628.0–22,859.5)
Crime	15,095.0 (52,53.5–24,809.5)	19,234.5 (12,500.5–26,251.5)
Barriers to housing and services	23,271.5 (17,355.5–26,325.5)	26,846.5 (19,260.0–30,273.0)
Living environment deprivation	9448.0 (2717.0–25,406.5)	15,491.0 (7144.5–21,202.5)
IMD	17,995.0 (11,240.0–23,126.5)	15,626.0 (11,146.0–24,248.5)

IMD, Index of Multiple Deprivation.

a For the English indices of area deprivation, the 32,844 Lower layer Super Output Areas in England are ranked according to their deprivation score. For each of the neighbourhood-level indices, the most deprived Lower layer Super Output Area in England is given a rank of 1, and the least deprived a rank of 32,844 (IMD is a weighted aggregate of the seven sub-indices).

**TABLE 18** Baseline participant characteristics, per trial arm in the main trial

Key characteristics	Stratified care arm (N = 534)	Usual care arm (N = 677)	p-value
<i>Point of consultation with GP</i>			
Pain intensity (0–10-point NRS), <sup>a</sup> mean (SD)	6.8 (1.9)	6.7 (2.0)	0.726
Back	7.1 (1.8)	6.9 (1.9)	0.309
Neck	6.9 (1.6)	6.7 (2.2)	0.555
Shoulder	6.7 (1.8)	6.8 (1.9)	0.697
Knee	6.2 (2.1)	6.4 (2.3)	0.578
Multisite	7.2 (1.9)	7.0 (1.5)	0.524
<i>Baseline questionnaire</i>			
Age (years), mean (SD)	57.8 (15.3)	61.8 (15.0)	<b>0.004</b>
Sex (female), n (%)	313 (58.6)	400 (59.1)	0.869
Days between consultation and returning questionnaire, mean (SD)	16.6 (27.3)	16.8 (28.0)	0.896
Ethnicity (white), n (%) <sup>b</sup>	513 (96.8)	659 (97.6)	0.400
Lives alone (yes), n (%)	91 (17.2)	121 (18.0)	0.936
Currently employed (yes), n (%)	275 (53.8)	286 (43.7)	<b>0.001</b>
Performance at work in past 6 months (current workers only, 0–10-point NRS), mean (SD) <sup>c</sup>	4.9 (3.1)	4.8 (3.1)	0.725

**TABLE 18** Baseline participant characteristics, per trial arm in the main trial (*continued*)

Key characteristics	Stratified care arm (N = 534)	Usual care arm (N = 677)	p-value
Performance at work in past 6 months (total study population, 0–10-point NRS), mean (SD) <sup>c</sup>	4.8 (3.2)	4.6 (3.2)	0.529
Time off work (yes), n (%)	99 (32.7)	203 (31.0)	0.803
Number of days off (from yes subgroup), median (IQR)	5 (2.5–15)	10 (4–20)	0.341
Health literacy (need help), n (%)			0.580
Never	434 (82.5)	539 (81.1)	
Rarely	44 (8.4)	58 (8.7)	
Sometimes	31 (5.9)	42 (6.3)	
Often	14 (2.7)	18 (2.7)	
Always	3 (0.6)	8 (1.2)	
Pain area affected, n (%)			0.007
Back	214 (40.1)	243 (35.9)	
Neck	61 (11.4)	68 (10.0)	
Shoulder	71 (13.3)	59 (8.7)	
Knee	157 (29.4)	222 (32.8)	
Multisite pain	31 (5.8)	85 (12.6)	
Pain intensity (0–10-point NRS), <sup>a</sup> by pain area, mean (SD)	6.3 (2.2)	6.4 (2.2)	0.775
Back	6.6 (2.2)	6.5 (2.2)	0.876
Neck	6.3 (2.0)	5.9 (2.2)	0.325
Shoulder	6.6 (1.9)	6.8 (2.2)	0.905
Knee	5.9 (2.3)	6.1 (2.4)	0.412
Multisite	6.5 (2.5)	6.8 (1.7)	0.600
Distress (0–10-point NRS), mean (SD) <sup>d</sup>	5.9 (2.6)	5.8 (2.6)	0.914
Confidence to manage (0–10-point NRS), mean (SD) <sup>e</sup>	5.1 (2.5)	5.3 (2.6)	0.186
Pain duration, n (%)			0.674
< 3 months	126 (23.9)	180 (26.7)	
3–6 months	106 (20.1)	101 (15.0)	
7–12 months	65 (12.3)	86 (12.8)	
1–2 years	64 (12.1)	83 (12.3)	
3–5 years	74 (14.0)	88 (13.1)	
6–10 years	34 (6.4)	49 (7.3)	
> 10 years	59 (11.2)	87 (12.9)	
Overall pain change (–5 to +5 points), mean (SD) <sup>f</sup>	0.3 (2.1)	0.3 (2.0)	0.833

continued

**TABLE 18** Baseline participant characteristics, per trial arm in the main trial (*continued*)

Key characteristics	Stratified care arm (N = 534)	Usual care arm (N = 677)	p-value
Previous episodes in last 3 years, n (%)			0.052
0	135 (25.5)	125 (18.5)	
1	69 (13.0)	75 (11.1)	
2–3	98 (18.5)	133 (19.7)	
4–9	70 (13.2)	115 (17.0)	
≥ 10	158 (29.8)	227 (33.6)	
Previous surgery related to problem, n (%)			0.162
0	455 (89.6)	564 (85.5)	
1	38 (7.5)	61 (9.2)	
2	5 (1.0)	24 (3.6)	
≥ 3	10 (2.0)	11 (1.7)	
Days of moderate activity in last week, median (IQR)	2 (1–4)	2 (0–5)	0.716
Physical function, mean (SD)			
Back (RMDQ) <sup>g</sup>	9.9 (5.8)	9.2 (5.6)	0.159
Neck (NDI) <sup>h</sup>	17.5 (8.7)	15.7 (8.7)	0.241
Shoulder (SPADI-Disability subscale) <sup>i</sup>	46.8 (24.1)	50.9 (26.1)	0.776
Knee (KOOS-PS) <sup>j</sup>	56.5 (15.2)	55.8 (17.9)	0.711
Multisite (SF-12 PCS) <sup>k</sup>	37.6 (8.8)	33.2 (10.3)	0.035
Standardised function scale (overall mean 0, SD 1)	0.01 (0.97)	0.00 (1.02)	0.844
MSK-HQ (0–56), mean (SD) <sup>l</sup>	28.7 (9.9)	29.5 (10.3)	0.604
The Keele STaT MSK Tool (clinical version), mean (SD) <sup>m</sup>	7.1 (2.7)	7.0 (2.9)	0.969
The Keele STaT MSK Tool (clinical version): risk subgroup, n (%)			0.964
Low risk (0–4 score)	98 (19.5)	126 (20.1)	
Medium risk (5–8 score)	238 (47.3)	286 (45.6)	
High risk (9–12 score)	167 (33.2)	215 (34.3)	
Health-related quality of life (EQ-5D-5L), mean (SD) <sup>n</sup>	0.56 (0.23)	0.55 (0.24)	0.754
Fear avoidance beliefs (TSK-11), mean (SD) <sup>o</sup>	25.4 (6.4)	24.8 (6.5)	0.262
Listed long term conditions, n (%)			0.237
0	184 (34.5)	202 (29.8)	
1	192 (36.0)	247 (36.5)	
2	103 (19.3)	147 (21.7)	
≥ 3	55 (10.3)	81 (12.0)	

**TABLE 18** Baseline participant characteristics, per trial arm in the main trial (*continued*)

Key characteristics	Stratified care arm (N = 534)	Usual care arm (N = 677)	p-value
Perceived reassurance from GP consultation (RQ), mean (SD) <sup>p</sup>			
Total	58.7 (16.0)	59.8 (16.6)	0.686
Data gathering	16.5 (4.3)	16.6 (4.6)	0.885
Relationship building	16.7 (4.3)	16.9 (4.5)	0.748
Generic	10.7 (5.0)	10.6 (5.2)	0.784
Cognitive	14.9 (5.1)	15.5 (5.1)	0.378
Satisfaction with GP care in last 6 months, n (%)			0.258
Very satisfied	123 (23.7)	190 (28.3)	
Quite satisfied	173 (33.3)	244 (36.3)	
No opinion	149 (28.7)	153 (22.8)	
Not very satisfied	58 (11.2)	73 (10.9)	
Not at all satisfied	16 (3.1)	12 (1.8)	
Preferential mode of follow-up, n (%)			<b>0.013</b>
Text message	285 (53.4)	310 (45.8)	
Post	249 (46.6)	367 (54.2)	

KOOS-PS, Knee Injury and Osteoarthritis Outcome Score Physical Function Short-Form; MSK-HQ, Musculoskeletal Health Questionnaire; NDI, Neck Disability Index; RQ, Reassurance Questionnaire; SF-12, Short Form questionnaire-12 items; SPADI, Shoulder Pain and Disability Index; TSK-11, Tampa Scale for Kinesiophobia-11 item version.

a NRS-pain: 0 points = 'no pain', 10 points = 'worst ever pain'.

b Non-white group includes mixed, Asian, black and other.

c Performance at work (0–10 NRS): 0 = problem 'not at all' affected performance over last 6 months, 10 = 'so bad I am unable to do my job'.

d NRS-distress: 0 points = 'no distress', 10 points = 'extreme distress'.

e NRS-confidence to manage: 0 points = 'not at all confident', 10 points = 'extremely confident'.

f Pain change on 11-point NRS scale (–5 to +5 points): –5 points = very much worse, 0 points = unchanged, +5 points = completely recovered (change from clinic appointment to time of self-report baseline completion).

g RMDQ (0–24 scale): 0 = no low back pain/disability, 24 = maximum low back/pain disability.

h NDI (0–50 scale): 0 = no disability, 50 = maximum disability.

i SPADI–Disability subscale (0–100): 0 = no disability, 100 = maximum disability.

j KOOS-PS (0–100): 0 = extreme disability, 100 = no disability.

k SF-12 PCS (0–100): 0 = worst physical health score, 100 = best physical health score.

l MSK-HQ (0–56 scale) based on summation of 14-items on a 0–4 scale and where 0 = worst musculoskeletal healthstatus and 56 = best musculoskeletal health-status.

o MSK-HQ (0–56 scale) based on summation of 14-items on a 0–4 scale and where 0 = worst musculoskeletal health-status and 56 = best musculoskeletal health-status.

m The Keele STarT MSK Tool score (0–12): 0 = lowest risk, 12 = highest risk.

n EQ-5D-5L (–0.59 to 1.00): –0.59 = worst health status, 1.00 = best health status.

o TSK-11 (11–44): 11 = minimum fear avoidance, 44 = maximum fear avoidance.

p RQ (12–84 scale): 12 = no reassurance, 84 = high reassurance (subscales all recorded on 3–21 scale: 3 = no reassurance, 21 = high reassurance).

#### Note

p-values derived through linear or generalised mixed model accounting for GP–practice clustering (random factor). Bold indicates where there is a difference between the two groups for that variable.

TABLE 19 Pain intensity scores (0–10-point NRS) and between-arm differences in the main trial

Time point	Stratified care, <i>n</i>	Usual care, <i>n</i>	Stratified care mean (SD)	Usual care mean (SD)	Mean difference (95% CI)	SMD (95% CI)	<i>p</i> -value
Point of consultation with GP	534	675	6.8 (1.9)	6.7 (2.0)	–	–	–
Baseline	533	675	6.3 (2.2)	6.4 (2.2)	–	–	–
Month 1	491	632	5.1 (2.5)	5.0 (2.6)	0.17 (–0.34 to 0.68)	0.09 (–0.17 to 0.35)	0.51
Month 2	470	623	4.7 (2.6)	4.6 (2.7)	0.09 (–0.43 to 0.60)	0.05 (–0.22 to 0.31)	0.74
Month 3	478	615	4.2 (2.7)	4.5 (2.8)	–0.21 (–0.73 to 0.30)	–0.11 (–0.37 to 0.15)	0.42
Month 4	455	604	4.0 (2.7)	4.5 (2.9)	–0.40 (–0.92 to 0.13)	–0.21 (–0.47 to 0.07)	0.14
Month 5	451	601	3.9 (2.8)	4.4 (2.9)	–0.35 (–0.88 to 0.18)	–0.18 (–0.45 to 0.09)	0.2
Month 6 <sup>b</sup>	446	593	3.8 (2.8)	4.2 (2.9)	–0.24 (–0.78 to 0.30)	–0.12 (–0.40 to 0.15)	0.39
Average (1–6 months) <sup>c,d,e</sup>	515	663	4.4 (2.3)	4.6 (2.5)	–0.16 (–0.65 to 0.34) <sup>f</sup>	–0.08 (–0.33 to 0.17)	0.54
Average (4–6 months) <sup>g,h</sup>	486	642	4.0 (2.6)	4.3 (2.7)	–0.33 (–0.84 to 0.19) <sup>f</sup>	–0.17 (–0.43 to 0.10)	0.21

a The between-group differences were calculated through use of linear mixed models with practice and participants specified as random factors, and practice size, participants' age, sex and point-of-consultation pain score specified as fixed factors.

b If the last monthly SMS/brief questionnaire response was missing it was imputed using the corresponding pain response from the returned 6-month questionnaire (if completed within 20 days of the date of issue).

c Primary end point (average of available data for 1–6 months follow-up). Average summary mean (SD) relates to the mean of available 1–6-month follow-up data.

d Sensitivity analysis (1–6 months).

e Sensitivity analysis comparing those with/without complete monthly follow-up.

f Kenward-Roger/Satterthwaite adjustment to CI: average 1–6 months –0.68 to 0.37; average 4–6 months –0.87 to 0.22.

g Post hoc analysis (not pre-specified).

h Sensitivity analysis (4–6 months); post hoc analysis due to patients not receiving their intervention immediately owing to waiting lists.

#### Note

Contrast arm × time: chi-squared test 19.8, 5 degrees of freedom; *p* = 0.0014.

SMD (effect size) relative to NRS-pain point-of-consultation SD (1.95).

(i) Constrained baseline model –0.18, 95% CI –0.53 to 0.16; *p* = 0.302.

(ii) Per-Protocol analysis (based on consultation template) –0.11, 95% CI –0.62 to 0.41; *p* = 0.681.

(iii) Complier-average causal effect analysis (compliance based on consultation template) 0.47, 95% CI –2.16 to 3.10; *p* = 0.725.

(iv) MI inclusive of additional baseline covariates –0.03, 95% CI –0.48 to 0.42; *p* = 0.900.

(v) Practice-level analysis –0.29, 95% CI –0.95 to 0.36; *p* = 0.358.

(i) Constrained baseline model –0.35, 95% CI –0.74 to 0.04; *p* = 0.082.

(ii) Per-protocol analysis (based on consultation template) –0.30, 95% CI –0.84 to 0.23; *p* = 0.270.

(iii) Complier-average causal effect analysis (compliance based on consultation template) 0.93, 95% CI –1.98 to 3.83; *p* = 0.532.

(iv) MI inclusive of additional baseline covariates –0.15, 95% CI –0.64 to 0.33; *p* = 0.536.

(v) Practice-level analysis –0.45, 95% CI –1.16 to 0.27; *p* = 0.208.

(i) Average between-group difference in mean scores for participants with 6/6 (100%) monthly responses: –0.24, 95% CI –0.76 to 0.39; *p* = 0.389.

(ii) Average between-group difference in mean scores for participants with incomplete (< 6) monthly responses: 0.01, 95% CI 0.68 to –0.65; *p* = 0.971.

The additional analyses presented were additional sensitivity analyses carried out across 1–6 months, and 4–6 months, based on recommendations from the DMC. The first five are 1–6 months, the next five are 4–6 months (due to potential delayed impact of the intervention), and the final two were sensitivity analyses for related to participants who did/did not complete all follow-up time points.



TABLE 20 Number (%) achieving the MCIC for pain intensity NRS (at least 1-point change)

Month	Stratified care, n/N (%)	Usual care, n/N (%)	OR (95% CI) <sup>b</sup>	p-value	Percentage difference (95% CI)	NNT (95% CI) <sup>c</sup>
1	319/491 (65.0)	404/632 (63.9)	0.89 (0.48 to 1.66)	0.72	-2.7 (-18.0 to 10.7)	36.6 <sup>UC</sup> (5.6 <sup>UC</sup> to 9.3 <sup>SC</sup> )
2	332/470 (70.6)	418/623 (67.1)	1.11 (0.59 to 2.10)	0.74	2.3 (-12.5 to 14.0)	44.2 <sup>SC</sup> (8.0 <sup>UC</sup> to 7.2 <sup>SC</sup> )
3	361/478 (75.5)	414/615 (67.3)	1.92 (1.01 to 3.67)	0.047	12.5 (0.2 to 21.0)	8.0 <sup>SC</sup> (457 <sup>SC</sup> to 4.8 <sup>SC</sup> )
4	346/455 (76.0)	397/604 (65.7)	2.47 (1.29 to 4.74)	0.007	16.9 (5.5 to 24.4)	5.9 <sup>SC</sup> (18.2 <sup>SC</sup> to 4.1 <sup>SC</sup> )
5	349/451 (77.4)	408/601 (67.9)	2.17 (1.12 to 4.20)	0.021	14.2 (2.4 to 22.0)	7.0 <sup>SC</sup> (41.3 <sup>SC</sup> to 4.5 <sup>SC</sup> )
6 <sup>d</sup>	347/446 (77.8)	412/593 (69.5)	2.03 (1.04 to 3.94)	0.037	12.7 (0.8 to 20.5)	7.9 <sup>SC</sup> (121 <sup>SC</sup> to 4.9 <sup>SC</sup> )
Mean 1–6 months <sup>e</sup>	371/515 (72.0)	434/663 (65.5)	1.66 (0.98 to 2.82)	0.061	10.4 (-0.5 to 18.8) <sup>d</sup>	9.6 <sup>SC</sup> (218 <sup>UC</sup> to 5.3 <sup>SC</sup> ) <sup>d</sup>
Mean 4–6 months <sup>f,g</sup>	367/486 (75.5)	423/642 (65.9)	2.22 (1.26 to 3.89)	0.006	15.2 (5.0 to 22.4) <sup>d</sup>	6.6 <sup>SC</sup> (20.0 <sup>SC</sup> to 4.5 <sup>SC</sup> ) <sup>d</sup>

NNT, number needed to treat.

a The between-group differences were calculated through use of linear mixed models with practice and participants specified as random factors, and practice size, participants' age, sex and point-of-consultation pain score specified as fixed factors.

b Absolute percentage difference derived from the odds ratio using the observed proportion in the control arm as the reference.

c NNT (derived as the reciprocal of the percentage difference) where SC superscript denotes the NNT in respect of favourable stratified care arm and UC superscript denotes the NNT in respect of favourable usual care arm.

d If the last monthly SMS/brief questionnaire response was missing it was imputed using the corresponding pain response from the returned 6-month questionnaire (if completed within 20 days of the date of issue).

e Sensitivity analysis using different cut-off points (1–6 months).

f Post hoc analysis (not pre-specified).

g Sensitivity analysis using different cut-off points (4–6 months).

**Note**Contrast arm × time: chi-squared test 20.3, 5 degrees of freedom;  $p = 0.0011$ .(i) MCIC 2-point change OR = 1.34, 0.75 to 2.42;  $p = 0.322$ .(ii) MCIC 3-point change OR = 1.28, 0.71 to 2.32;  $p = 0.410$ .(i) MCIC 2-point change OR = 1.87, 1.01 to 3.43;  $p = 0.045$ .(ii) MCIC 3-point change OR = 1.88, 1.02 to 3.48;  $p = 0.044$ .

The additional analyses presented were additional sensitivity analyses carried out across 1–6 months, and 4–6 months, based on recommendations from the DMC. The first five are 1–6 months, the next five are 4–6 months (due to potential sensitivity analyses for related to participants who did/did not complete all follow-up time points).

**TABLE 21** Pain intensity (0–10-point NRS) scores per patient risk subgroup (clinical version)

Month	Low risk, mean (SD) [n]		Medium risk, mean (SD) [n]		High risk, mean (SD) [n]	
	Stratified care	Usual care	Stratified care	Usual care	Stratified care	Usual care
1	2.8 (2.0) [93]	2.6 (2.1) [123]	5.2 (2.0) [218]	5.0 (2.2) [267]	6.3 (2.3) [156]	6.4 (2.3) [199]
2	2.5 (2.0) [88]	2.4 (2.2) [119]	4.7 (2.3) [209]	4.6 (2.3) [262]	5.7 (2.5) [150]	5.9 (2.6) [199]
3	2.4 (2.0) [88]	2.4 (2.3) [120]	4.2 (2.4) [216]	4.4 (2.6) [258]	5.3 (2.8) [151]	5.9 (2.7) [195]
4	2.2 (1.9) [88]	2.4 (2.3) [116]	4.0 (2.6) [205]	4.4 (2.7) [259]	5.1 (2.7) [139]	5.8 (2.7) [188]
5	2.0 (2.1) [87]	2.2 (2.3) [117]	3.9 (2.5) [201]	4.2 (2.7) [252]	5.0 (2.9) [142]	5.7 (2.7) [192]
6 <sup>a</sup>	1.9 (2.1) [85]	2.0 (2.1) [114]	3.8 (2.6) [198]	4.0 (2.7) [249]	4.9 (2.9) [143]	5.5 (2.8) [191]
Mean (1–6 months) <sup>b</sup>	2.3 (1.7) [93]	2.3 (1.9) [124]	4.4 (2.0) [233]	4.4 (2.1) [282]	5.4 (2.2) [164]	5.9 (2.3) [210]
Interaction <sup>c</sup>			–0.01, 95% CI –0.64 to 0.62; <i>p</i> = 0.98		–0.30, 95% CI –0.97 to 0.36; <i>p</i> = 0.37	

a If the last monthly SMS/brief questionnaire response was missing it was imputed using the corresponding pain response from the returned 6-month questionnaire (if completed within 20 days of the date of issue).

b Summary mean (SD) relates to the mean of available 1–6-month follow-up data.

c Between-group difference in mean scores (stratified care minus usual care) by linear mixed model with practice and participants (random factors) and practice size and participants' age, sex, point-of-consultation pain score, time as well as interaction of The Keele STarT MSK Tool clinical subgroup dummy variables (low-risk subgroup = reference) × treatment arm (fixed factors).

TABLE 22 Pain intensity (0–10-point NRS) scores per subgroup of musculoskeletal pain site

Month	Back, mean (SD) [n]		Neck, mean (SD) [n]		Shoulder, mean (SD) [n]		Knee, mean (SD) [n]		Multisite, mean (SD) [n]	
	Stratified care	Usual care	Stratified care	Usual care	Stratified care	Usual care	Stratified care	Usual care	Stratified care	Usual care
1	5.2 (2.5) [195]	4.7 (2.5) [221]	5.1 (2.4) [58]	4.4 (2.6) [67]	5.1 (2.4) [67]	5.7 (2.8) [53]	4.9 (2.4) [146]	4.9 (2.6) [210]	6.0 (2.5) [25]	5.9 (2.2) [81]
2	4.8 (2.5) [186]	4.4 (2.7) [221]	4.3 (2.6) [52]	4.1 (2.6) [63]	4.5 (3.0) [62]	4.9 (2.7) [55]	4.4 (2.4) [144]	4.6 (2.7) [205]	5.8 (2.6) [26]	5.6 (2.3) [79]
3	4.3 (2.7) [192]	4.1 (2.8) [221]	4.1 (2.8) [52]	3.9 (2.8) [60]	3.8 (2.9) [62]	4.8 (3.0) [52]	4.2 (2.6) [145]	4.6 (2.8) [206]	5.4 (2.4) [27]	5.8 (2.3) [76]
4	4.2 (2.7) [180]	4.2 (2.8) [213]	4.1 (2.7) [53]	3.8 (2.8) [63]	3.4 (2.8) [62]	4.6 (3.1) [53]	3.8 (2.5) [136]	4.5 (3.0) [200]	5.5 (2.8) [24]	5.7 (2.5) [75]
5	4.1 (2.9) [176]	4.2 (2.8) [213]	3.6 (2.7) [50]	3.9 (2.8) [61]	3.4 (2.7) [62]	4.4 (3.0) [50]	3.9 (2.6) [137]	4.3 (3.1) [201]	5.2 (2.9) [26]	5.4 (2.5) [76]
6 <sup>a</sup>	4.0 (2.8) [171]	4.1 (2.9) [214]	3.5 (2.8) [50]	3.7 (2.8) [61]	3.6 (3.0) [61]	3.8 (2.8) [48]	3.6 (2.6) [138]	4.0 (3.0) [198]	5.2 (2.9) [26]	5.6 (2.6) [72]
Average (1–6 months) <sup>b</sup>	4.4 (2.3) [205]	4.3 (2.5) [236]	4.4 (2.4) [61]	3.9 (2.5) [68]	4.1 (2.5) [67]	4.8 (2.5) [57]	4.2 (2.2) [154]	4.6 (2.5) [219]	5.6 (2.4) [28]	5.6 (2.2) [83]
Reference site	Interaction: back (95% CI)		Interaction: neck (95% CI)		Interaction: shoulder (95% CI)		Interaction: knee (95% CI)		Interaction: multisite (95% CI)	
Back	-		0.37 (-0.50 to 1.24); p = 0.40		-0.56 (-1.46 to 0.34); p = 0.22		-0.48 (-1.10 to 0.14); p = 0.13		-0.16 (-1.20 to 0.89); p = 0.77	
Neck	-0.37 (-1.24 to 0.50); p = 0.40		-		-0.93 (-2.03 to 0.17); p = 0.096		-0.85 (-1.75 to 0.04); p = 0.061		-0.53 (-1.75 to 0.70); p = 0.40	
Shoulder	0.56 (-0.34 to -1.46); p = 0.22		0.93 (-0.17 to 2.03); p = 0.096		-		0.08 (-0.83 to 0.99); p = 0.87		0.41 (-0.84 to 1.65); p = 0.52	
Knee	0.48 (-0.14 to 1.10); p = 0.13		0.85 (-0.04 to 1.75); p = 0.061		-0.08 (-0.99 to 0.83); p = 0.87		-		0.33 (-0.74 to 1.39); p = 0.55	

<sup>a</sup> If the last monthly SMS/brief questionnaire response was missing it was imputed using the corresponding pain response from the returned 6-month questionnaire (if completed within 20 days of the date of issue).

<sup>b</sup> Average summary mean (SD) relates to the mean of available 1–6 month follow-up data.

<sup>c</sup> Pre-specified reference site category = back. Between-group difference in mean scores (stratified care minus usual care) by linear mixed model with practice and participants (random factors) and practice size and participants' age, sex, point-of-consultation pain score as well as interaction of pain site dummy variables × treatment arm (fixed factors).

**TABLE 23** Pain intensity (0–10-point NRS) scores per musculoskeletal pain region

	Single-site, mean (SD) [n]		Multisite, mean (SD) [n]	
Month	Stratified care	Usual care	Stratified care	Usual care
1	5.1 (2.4) [466]	4.8 (2.6) [551]	6.0 (2.5) [25]	5.9 (2.2) [81]
2	4.6 (2.6) [444]	4.5 (2.7) [544]	5.8 (2.6) [26]	5.6 (2.3) [79]
3	4.2 (2.7) [451]	4.3 (2.9) [539]	5.4 (2.4) [27]	5.8 (2.3) [76]
4	4.0 (2.7) [431]	4.3 (2.9) [529]	5.5 (2.8) [24]	5.7 (2.5) [75]
5	3.9 (2.8) [425]	4.2 (3.0) [525]	5.2 (2.9) [26]	5.4 (2.5) [76]
6 <sup>a</sup>	3.8 (2.8) [420]	4.0 (2.9) [521]	5.2 (2.9) [26]	5.6 (2.6) [72]
Mean (1–6 months) <sup>b</sup>	4.3 (2.3) [487]	4.4 (2.5) [580]	5.6 (2.4) [28]	5.6 (2.2) [83]
Interaction <sup>c</sup>			0.02, 95% CI –0.98 to 1.02; p = 0.972	
a If the last monthly SMS/brief questionnaire response was missing it was imputed using the corresponding pain response from the returned 6-month questionnaire (if completed within 20 days of the date of issue).				
b Summary mean (SD) relates to the mean of available 1–6 month follow-up data.				
c Between-group difference in mean scores (stratified care minus usual care) by linear mixed model with practice and participants (random factors) and practice size and participants' age, sex, point-of-consultation pain score as well as interaction of pain site [single-site (reference) vs. multisite] × treatment arm (fixed factors).				

**TABLE 24** Secondary outcome measures at 6 months, per trial arm in the main trial

Key characteristics	Stratified care (n = 534)	Usual care (n = 677)	Mean difference (95% CI) <sup>a</sup>	p-value
Pain intensity reported in questionnaire (0–10-point NRS), mean (SD) <sup>b,1</sup>	3.5 (2.7)	4.1 (2.8)	–0.45 (–0.97 to 0.07)	0.088
Overall global change (–5 to +5), mean (SD) <sup>c,1</sup>	1.7 (2.4)	1.2 (2.6)	0.29 (–0.10 to 0.68)	0.143
Days of moderate activity in last week, mean (SD). <sup>1</sup>	3.2 (2.1)	3.2 (2.3)	–0.06 (–0.35 to 0.22)	0.656
Physical Function, mean (SD). <sup>1</sup>				
Back (RMDQ) <sup>d</sup>	6.4 (5.7)	6.1 (5.5)	–0.30 (–1.30 to 0.70)	0.558
Neck (NDI) <sup>e</sup>	11.5 (8.9)	12.3 (9.1)	–1.01 (–4.81 to 2.80)	0.604
Shoulder (SPADI-Disability subscale) <sup>f</sup>	25.5 (27.3)	39.5 (31.4)	–11.1 (–19.8 to –2.3)	0.013
Knee (KOOS-PS) <sup>g</sup>	68.1 (14.7)	65.6 (20.0)	0.35 (–4.94 to 5.64)	0.896
Multisite (SF-12 PCS) <sup>h</sup>	40.4 (9.9)	35.8 (11.5)	0.31 (–4.40 to 5.01)	0.899
Standardised function score (0, 1)	–0.06 (0.94)	0.05 (1.04)	–0.07 (–0.22 to 0.08)	0.341
MSK-HQ (0–56), mean (SD) <sup>i,1</sup>	39.2 (11.0)	37.4 (12.1)	1.57 (–0.30 to 3.45)	0.100
The Keele STaT MSK tool (clinical version), mean (SD) <sup>i,1</sup>	4.8 (2.8)	5.1 (3.1)	–0.27 (–0.73 to 0.20)	0.265
The Keele STaT MSK (clinical version): risk subgroup, n (%)			OR 0.76 (0.51 to 1.13)	0.174
Low risk (0–4 score)	211 (53.0)	263 (48.6)		
Medium risk (5–8 score)	136 (34.2)	180 (33.3)		
High risk (9–12 score)	51 (12.8)	98 (18.1)		

**TABLE 24** Secondary outcome measures at 6 months, per trial arm in the main trial (*continued*)

Key characteristics	Stratified care (n = 534)	Usual care (n = 677)	Mean difference (95% CI) <sup>a</sup>	p-value
Health-related quality of life (EQ-5D-5L), mean (SD) <sup>k,1</sup>	0.67 (0.23)	0.65 (0.24)	0.022 (−0.003 to 0.048)	0.082
Fear avoidance beliefs (TSK-11), mean (SD) <sup>l,1</sup>	22.7 (7.0)	23.3 (7.4)	−0.64 (−1.70 to 0.42)	0.240
Currently employed (yes), n (%)	177 (44.1)	219 (39.8)	OR 0.83 (0.43 to 1.59)	0.574
Performance at work over last 6 months (0–10-point NRS), mean (SD)	3.4 (2.8)	3.5 (3.0)	−0.18 (−0.62 to 0.27)	0.434
Time off work in last 6 months (yes), n (%)	47 (23.6)	57 (23.2)	OR 0.97 (0.53 to 1.78)	0.919
Satisfaction with care, n (%)			OR 0.74 (0.57 to 0.98)	<b>0.033</b>
Very satisfied	125 (30.3)	138 (24.7)		
Quite satisfied	141 (34.1)	205 (36.7)		
No opinion	97 (23.5)	134 (24.0)		
Not very satisfied	39 (9.4)	64 (11.5)		
Not at all satisfied	11 (2.7)	18 (3.2)		

KOOS-PS, Knee Injury and Osteoarthritis Outcome Score-Physical Function Short-Form; MSK-HQ, Musculoskeletal Health Questionnaire; NDI, Neck Disability Index; SF-12, Short Form questionnaire-12 items; SPADI, Shoulder Pain and Disability Index; TSK-11, Tampa Scale for Kinesiophobia-11 item version.

a The between-group differences were calculated through use of linear mixed models with practice specified as a random factor, and practice size, participants' age, sex, point-of-consultation pain score and corresponding baseline measure (if available) specified as fixed factors. Generalised mixed models were used for categorical outcomes, labelled as OR.

b NRS-pain: 0 points = 'no pain', 10 points = 'worst ever pain'.

c Pain change on 11-point NRS scale (−5 to +5 points): −5 points = very much worse, 0 points = unchanged, +5 points = completely recovered (change from clinic appointment to time of self-report baseline completion).

d RMDQ (0–24 scale): 0 = no low back pain/disability, 24 = maximum low back/pain disability.

e NDI (0–50 scale): 0 = no disability, 50 = maximum disability.

f SPADI- Disability subscale (0–100): 0 = no disability, 100 = maximum disability.

g KOOS-PS (0–100): 0 = extreme disability, 100 = no disability.

h SF-12 PCS (0–100): 0 = worst physical health score, 100 = best physical health score.

i MSK-HQ (0–56 scale) based on summation of 14-items on a 0–4 scale and where 0 = worst musculoskeletal health-status and 56 = best musculoskeletal health-status.

j The Keele STarT MSK Tool score (0–12): 0 = lowest risk, 12 = highest risk.

k EQ-5D-5L (−0.59 to 1.00): −0.59 = worst health status, 1.00 = best health status.

l TSK-11 (11–44): 11 = minimum fear avoidance, 44 = maximum fear avoidance.

m Performance at work (0–10 NRS): 0 = problem 'not at all' affected performance over last 6 months, 10 = 'so bad I am unable to do my job'.

#### Note

Number of days absent restricted to 120 days (over a 6-month period). Bold shows results that fall below the 0.05 significance level.

**TABLE 25** Number and proportion of invited patients for whom GPs recorded an intention to provide a recommended matched treatment (stratified care arm) based on trial EMR template

GP treatment option	Low risk (n = 226, 21.4%), n (%)	Medium risk (n = 585, 55.4%), n (%)	High risk (n = 245, 23.2%), n (%)	Grand total (N = 1056), n
Verbal advice given	160 (71)	393 (67)	142 (58)	695
Written advice given	146 (65)	338 (58)	119 (49)	603
OTC medication	119 (53)	13 (2)	2 (1)	134
Musculoskeletal interface service referral	1 (0)	73 (12)	40 (16)	114
Physiotherapy referral	21 (9)	378 (65)	148 (60)	547
Exercise programme	0 (0)	56 (10)	30 (12)	86
Expert peer support	0 (0)	26 (4)	23 (9)	49
Lifestyle advice/intervention	1 (0)	52 (9)	32 (13)	85
Opioid medication	6 (3)	135 (23)	77 (31)	218
Corticosteroid injection	1 (0)	19 (3)	11 (4)	31
Pain management service referral	0 (0)	0 (0)	7 (3)	7
Referral to secondary care	5 (2)	37 (6)	33 (13)	75
Referral for imaging	9 (4)	20 (3)	71 (29)	100
Prescribed atypical analgesia	0 (0)	2 (0)	49 (20)	51
Addressed comorbidities	0 (0)	0 (0)	28 (11)	28
Total treatments	469	1542	812	2823
The shading denotes treatment options matched to low (purple shading), medium (no shading) and high-risk (green shading) subgroups.				

**TABLE 26** Number and proportion of invited patients for whom the GP recorded selection of a recommended matched treatment (based on GP entered information onto the trial specific EMR template)

	n	Proportion of patient risk subgroup (%)	Proportion of total (%)
Per protocol (overall)	816	–	77.3
Low risk: per protocol	176	77.2	16.7
Medium risk: per protocol	457	77.7	43.3
High risk: per protocol	183	76.7	17.3
Treatment not in line with the stratified care protocol			
Low risk: given treatments for patients at medium risk	17	7.5	1.6
Low risk: given treatments for patients at high risk	8	3.5	0.8
Low risk: only tool used (no treatments selected)	27	11.8	2.6
Medium risk: given treatments for patients at low risk	66	11.2	6.3
Medium risk: given treatments for patients at high risk	15	2.6	1.4
Medium risk: only tool used (no treatments selected)	50	8.5	4.7
High risk: given treatments for patients at low risk	2	0.8	0.2

**TABLE 26** Number and proportion of invited patients for whom the GP recorded selection of a recommended matched treatment (based on GP entered information onto the trial specific EMR template) (*continued*)

	<i>n</i>	Proportion of patient risk subgroup (%)	Proportion of total (%)
High risk: given treatments for patients at medium risk	27	11.3	2.6
High risk: only tool used (no treatments selected)	27	11.3	2.6
Incomplete tool	1	–	0.1
Grand total	1056		

The shading denotes treatment options matched to low (purple shading), medium (no shading) and high-risk (green shading) subgroups.

**TABLE 27** Number and proportion of trial participants for whom the GP recorded selection of a recommended matched treatment (Stratified Care arm) (based on GP entered information onto the trial specific electronic template)

	Low risk ( <i>n</i> = 98, 18.4%), <i>n</i> (%)	Medium risk ( <i>n</i> = 311, 58.2%), <i>n</i> (%)	High risk ( <i>n</i> = 125, 23.4%), <i>n</i> (%)	Grand total 534, <i>n</i>
Verbal advice	71 (72)	213 (68)	70 (56)	354
Written advice	68 (69)	185 (59)	63 (50)	316
Over-the-counter Medication	57 (58)	6 (2)	0 (0)	63
Musculoskeletal interface service referral	1 (1)	37 (12)	23 (18)	61
Physiotherapy referral	10 (10)	210 (68)	80 (64)	300
Exercise programme	0 (0)	32 (10)	14 (11)	46
Expert peer support	0 (0)	17 (5)	11 (9)	28
Lifestyle advice/intervention	1 (1)	27 (9)	16 (13)	44
Opioid medication	2 (2)	69 (22)	35 (28)	106
Corticosteroid injection	1 (1)	14 (5)	8 (6)	23
Pain management referral	0 (0)	0 (0)	5 (4)	5
Referral to secondary care	2 (2)	22 (7)	13 (10)	37
Referred for imaging	4 (4)	12 (4)	35 (28)	51
Prescribed atypical analgesia	0 (0)	0 (0)	18 (14)	18
Address comorbidities	0 (0)	0 (0)	13 (10)	13
Total treatments	217	844	404	1465

The shading denotes treatment options matched to low (purple shading), medium (no shading) and high-risk (green shading) subgroups.

**TABLE 28** Number and proportion of trial participants for whom the GP recorded selection of a recommended matched treatment (Stratified Care arm) (based on GP entered information onto the trial specific electronic template)

	<i>n</i>	Proportion of patient risk subgroup (%)	Proportion of total (%)
Per protocol (overall)	421	–	78.8
Low risk: per protocol	78	78.8	14.6
Medium risk: per protocol	250	80.6	46.8
High risk: per protocol	93	75.0	17.4
Treatment not in line with the stratified care protocol			
Low risk: given treatments for patients at medium risk	10	10.1	1.9
Low risk: given treatments for patients at high risk	2	2.0	0.4
Low risk: only tool used (no treatments selected)	9	9.1	1.7
Medium risk: given treatments for patients at low risk	31	10.0	5.8
Medium risk: given treatments for patients at high risk	8	2.6	1.5
Medium risk: only tool used (no treatments selected)	21	6.8	3.9
High risk: given treatments for patients at low risk	0	0.0	0.0
High risk: given treatments for patients at medium risk	14	11.3	2.6
High risk: only tool used (no treatments selected)	17	13.7	3.2
Incomplete tool	1	–	0.2
Grand total	534		
The shading denotes treatment options matched to low (purple shading), medium (no shading) and high-risk (green shading) subgroups.			



**TABLE 29** Comparison of GP decision-making per arm (0–182 days) taken from the anonymised EMR audit

Anonymised EMR audit data on all patients with a completed trial recruitment template (N = 2494)				Trial participants with known baseline risk subgroup taken from the baseline questionnaire (N = 1130)					
	Both usual care and stratified care, all patients (n = 2494)	Usual care practices, all patients (n = 1438)	Stratified care practices, all patients (n = 1056)	Low risk (n = 224, 20%)		Medium risk (n = 524, 46%)		High risk (n = 382, 34%)	
				Usual care (n = 126)	Stratified care (n = 98)	Usual care (n = 286)	Stratified care (n = 238)	Usual care (n = 215)	Stratified care (n = 167)
Prescription, n (mean per patient)									
Simple analgesics									
0–182 days (overall)	1375 (0.55)	861 (0.60)	514 (0.49)	19 (0.15)	38 (0.39)	117 (0.41)	136 (0.57)	282 (1.31)	165 (0.99)
0–7 days (PoC)	312 (0.13)	116 (0.08)	196 (0.19)	5 (0.04)	34 (0.35)	12 (0.04)	45 (0.19)	39 (0.18)	22 (0.13)
≥ 1 in 0–7 days (PoC),n (%)	266 (10.7)	90 (6.3)	176 (16.7)	5 (4.0)	32 (32.7)	11 (3.8)	37 (15.5)	28 (13.0)	17 (10.2)
8–182 days	1063 (0.43)	745 (0.52)	318 (0.30)	14 (0.11)	4 (0.04)	105 (0.37)	91 (0.38)	243 (1.13)	143 (0.86)
Anti-inflammatories									
0–182 days (overall)	1965 (0.79)	1145 (0.80)	820 (0.78)	52 (0.41)	35 (0.36)	212 (0.74)	138 (0.58)	286 (1.33)	232 (1.39)
0–7 days (PoC)	580 (0.23)	331 (0.23)	249 (0.24)	19 (0.15)	21 (0.21)	54 (0.19)	45 (0.19)	63 (0.29)	36 (0.22)
≥ 1 in 0–7 days (PoC), n (%)	511 (20.5)	300 (20.9)	211 (20.0)	18 (14.3)	17 (17.3)	51 (17.8)	41 (17.2)	51 (23.7)	28 (16.8)
8–182 days	1385 (0.56)	814 (0.57)	571 (0.54)	33 (0.26)	14 (0.14)	158 (0.55)	93 (0.39)	223 (1.04)	196 (1.17)
Neuromodulators									
0–182 days (overall)	2929 (1.17)	1859 (1.29)	1070 (1.01)	51 (0.40)	31 (0.32)	244 (0.85)	238 (1.00)	619 (2.88)	376 (2.25)
0–7 days (PoC)	378 (0.15)	218 (0.15)	160 (0.15)	7 (0.06)	8 (0.08)	37 (0.13)	26 (0.11)	54 (0.25)	45 (0.27)
≥ 1 in 0–7 days (PoC), n (%)	269 (10.8)	159 (11.1)	110 (10.4)	5 (4.0)	7 (7.1)	26 (9.1)	16 (6.7)	37 (17.2)	30 (18.0)
8–182 days	2551 (1.02)	1641 (1.14)	910 (0.86)	44 (0.35)	23 (0.23)	207 (0.72)	153 (0.64)	565 (2.63)	331 (1.98)
Muscle relaxants									
0–182 days (overall)	667 (0.27)	452 (0.31)	215 (0.20)	6 (0.05)	6 (0.06)	46 (0.16)	32 (0.13)	61 (0.28)	49 (0.29)
0–7 days (PoC)	191 (0.08)	123 (0.09)	68 (0.06)	3 (0.02)	5 (0.05)	20 (0.07)	12 (0.05)	16 (0.07)	13 (0.08)
≥ 1 in 0–7 days (PoC), n (%)	164 (6.6)	101 (7.0)	63 (6.0)	3 (2.4)	5 (5.1)	19 (6.6)	12 (5.0)	16 (7.4)	12 (7.2)
8–182 days	476 (0.19)	329 (0.23)	147 (0.14)	3 (0.02)	1 (0.01)	26 (0.09)	20 (0.08)	45 (0.21)	36 (0.22)
									continued

TABLE 29 Comparison of GP decision-making per arm (0–182 days) taken from the anonymised EMR audit (continued)

		Anonymised EMR audit data on all patients with a completed trial recruitment template (N = 2494)		Trial participants with known baseline risk subgroup taken from the baseline questionnaire (N = 1130)					
		Both usual care and stratified care, all patients (n = 2494)	Usual care patients (n = 1438)	Stratified care practices, all patients (n = 1056)	Usual care (n = 126)	Stratified care (n = 98)	Medium risk (n = 524, 46%)	High risk (n = 382, 34%)	Stratified care (n = 167)
Weak opioids									
0–182 days (overall)		3803 (1.52)	2241 (1.56)	1562 (1.48)	69 (0.55)	33 (0.34)	333 (1.16)	305 (1.28)	588 (2.73)
0–7 days (PoC)		1035 (0.41)	458 (0.32)	577 (0.55)	13 (0.10)	25 (0.26)	79 (0.28)	111 (0.47)	93 (0.43)
≥ 1 in 0–7 days (PoC), n (%)		714 (28.6)	376 (26.1)	338 (32.0)	13 (10.3)	16 (16.3)	69 (24.1)	71 (29.8)	70 (32.6)
8–182 days		2768 (1.11)	1783 (1.24)	985 (0.93)	56 (0.44)	8 (0.08)	254 (0.89)	194 (0.82)	495 (2.30)
Long-term prescribing, n (%) <sup>b</sup>		308 (12.3)	179 (12.4)	129 (12.2)	3 (2.4)	4 (4.1)	38 (13.3)	30 (12.6)	42 (19.5)
Strong opioids									
0–182 days (overall)		595 (0.24)	244 (0.17)	351 (0.33)	0 (0.00)	11 (0.11)	18 (0.06)	55 (0.23)	89 (0.41)
0–7 days (PoC)		250 (0.10)	24 (0.02)	226 (0.21)	0 (0.00)	11 (0.11)	0 (0.00)	40 (0.17)	7 (0.03)
≥ 1 in 0–7 days (PoC), n (%)		228 (9.1)	14 (1.0)	214 (20.3)	0 (0)	11 (11.2)	0 (0)	40 (16.8)	6 (1.0)
8–182 days		345 (0.14)	220 (0.15)	125 (0.12)	0 (0.00)	0 (0.00)	18 (0.06)	15 (0.06)	82 (0.38)
Long-term prescribing, n (%) <sup>b</sup>		26 (1.0)	13 (0.9)	13 (1.2)	0 (0)	0 (0)	0 (0)	1 (0.4)	6 (2.8)
Corticosteroid injection									
0–182 days (overall)		244 (0.10)	155 (0.11)	89 (0.08)	13 (0.10)	8 (0.08)	36 (0.13)	31 (0.13)	39 (0.18)
0–7 days (PoC)		83 (0.03)	39 (0.03)	44 (0.04)	4 (0.03)	5 (0.05)	8 (0.03)	20 (0.08)	7 (0.03)
≥ 1 in 0–7 days (PoC), n (%)		72 (2.9)	39 (2.7)	33 (3.1)	4 (3.2)	3 (3.1)	8 (2.8)	14 (5.9)	7 (3.3)
8–182 days		161 (0.06)	116 (0.08)	45 (0.04)	9 (0.07)	3 (0.03)	28 (0.10)	11 (0.05)	32 (0.15)
									12 (0.07)

**TABLE 29** Comparison of GP decision-making per arm (0–182 days) taken from the anonymised EMR audit (continued)

Anonymised EMR audit data on all patients with a completed trial recruitment template (N = 2494)				Trial participants with known baseline risk subgroup taken from the baseline questionnaire (N = 1130)															
Both usual care and stratified care, all patients (n = 2494)		Usual care practices (n = 1438)		Stratified care practices, all patients (n = 1056)		Low risk (n = 224, 20%)		Medium risk (n = 524, 46%)		High risk (n = 382, 34%)									
						Usual care (n = 126)		Stratified care (n = 98)		Usual care (n = 215)		Stratified care (n = 167)							
Referral, n (mean per patient)																			
Physiotherapy or musculoskeletal interface clinic																			
0–182 days (overall)		1176 (0.47)		303 (0.21)		873 (0.83)		26 (0.21)		60 (0.61)		73 (0.26)		230 (0.97)		43 (0.2)		164 (0.98)	
0–7 days (PoC)		815 (0.33)		143 (0.10)		672 (0.64)		15 (0.12)		41 (0.42)		32 (0.11)		166 (0.70)		18 (0.08)		127 (0.76)	
≥ 1 in 0–7 days (PoC), n (%)		810 (32.5)		142 (9.9)		668 (63.3)		15 (11.9)		41 (41.8)		32 (11.2)		166 (69.7)		18 (8.4)		126 (75.4)	
8–182 days		361 (0.14)		160 (0.11)		201 (0.19)		11 (0.09)		2 (0.02)		41 (0.14)		64 (0.27)		25 (0.12)		37 (0.22)	
Specialist orthopaedics																			
0–182 days (overall)		575 (0.23)		357 (0.25)		218 (0.21)		24 (0.19)		11 (0.11)		65 (0.23)		52 (0.22)		72 (0.33)		44 (0.26)	
0–7 days (PoC)		184 (0.07)		120 (0.08)		64 (0.06)		8 (0.06)		4 (0.04)		19 (0.07)		11 (0.05)		26 (0.12)		11 (0.07)	
≥ 1 in 0–7 days (PoC), n (%)		176 (7.1)		112 (7.8)		64 (6.1)		7 (5.6)		4 (4.1)		18 (6.3)		11 (4.6)		22 (10.2)		11 (6.6)	
8–182 days		391 (0.16)		237 (0.16)		154 (0.15)		16 (0.13)		7 (0.07)		46 (0.16)		41 (0.17)		46 (0.21)		33 (0.20)	
Pain clinic																			
0–182 days (overall)		43 (0.02)		36 (0.03)		7 (0.01)		0 (0.00)		0 (0.00)		3 (0.01)		1 (0.004)		8 (0.04)		4 (0.02)	
0–7 days (PoC)		13 (0.01)		12 (0.008)		1 (0.001)		0 (0.00)		0 (0.00)		1 (0.003)		0 (0.00)		3 (0.01)		1 (0.006)	
≥ 1 in 0–7 days (PoC), n (%)		13 (0.5)		12 (0.8)		1 (0.1)		0 (0)		0 (0)		1 (0.3)		0 (0)		3 (1.4)		1 (0.6)	
8–182 days		30 (0.01)		24 (0.02)		6 (0.01)		0 (0.00)		0 (0.00)		2 (0.01)		1 (0.004)		5 (0.02)		3 (0.02)	
Rheumatology																			
0–182 days (overall)		42 (0.02)		25 (0.02)		17 (0.02)		1 (0.01)		0 (0.00)		7 (0.02)		3 (0.01)		9 (0.04)		2 (0.01)	
0–7 days (PoC)		7 (0.003)		4 (0.003)		3 (0.003)		0 (0.00)		0 (0.00)		2 (0.007)		1 (0.004)		2 (0.01)		0 (0.00)	
≥ 1 in 0–7 days (PoC), n (%)		7 (0.3)		4 (0.3)		3 (0.3)		0 (0)		0 (0)		2 (0.7)		1 (0.4)		2 (0.9)		0 (0)	
8–182 days		35 (0.01)		21 (0.02)		14 (0.01)		1 (0.008)		0 (0.00)		5 (0.02)		2 (0.01)		7 (0.03)		2 (0.01)	
														continued					

TABLE 29 Comparison of GP decision-making per arm (0–182 days) taken from the anonymised EMR audit (continued)

Anonymised EMR audit data on all patients with a completed trial recruitment template (N = 2494)				Trial participants with known baseline risk subgroup taken from the baseline questionnaire (N = 1130)															
Both usual care and stratified care, all patients (n = 2494)		Usual care practices, all patients (n = 1438)		Stratified care practices, all patients (n = 1056)		Low risk (n = 224, 20%)		Medium risk (n = 524, 46%)		High risk (n = 382, 34%)									
						Usual care (n = 126)		Stratified care (n = 98)		Usual care (n = 215)		Stratified care (n = 167)							
Imaging, n (mean per patient)																			
X-ray or MRI for musculoskeletal disorder																			
0–182 days (overall)		604 (0.24)		468 (0.33)		136 (0.13)		28 (0.22)		3 (0.03)		84 (0.29)		33 (0.14)		82 (0.38)		26 (0.16)	
0–7 days (PoC)		125 (0.05)		99 (0.07)		26 (0.02)		10 (0.08)		1 (0.01)		22 (0.08)		8 (0.03)		18 (0.08)		2 (0.01)	
≥ 1 in 0–7 days (PoC), n (%)		96 (3.8)		75 (5.2)		21 (2.0)		8 (6.3)		1 (1.0)		15 (5.2)		6 (2.5)		15 (7.0)		2 (1.2)	
8–182 days		479 (0.19)		369 (0.26)		110 (0.10)		18 (0.14)		2 (0.02)		62 (0.22)		25 (0.11)		64 (0.30)		24 (0.14)	
Ultrasound scan for musculoskeletal disorder																			
0–182 days (overall)		97 (0.04)		70 (0.05)		27 (0.03)		5 (0.04)		1 (0.01)		9 (0.03)		8 (0.03)		10 (0.05)		6 (0.04)	
0–7 days (PoC)		2 (0.001)		1 (0.001)		1 (0.001)		1 (0.008)		0 (0.00)		0 (0.00)		0 (0.00)		0 (0.00)		0 (0.00)	
≥ 1 in 0–7 days (PoC), n (%)		2 (0.1)		1 (0.1)		1 (0.1)		1 (0.8)		0 (0)		0 (0)		0 (0)		0 (0)		0 (0)	
8–182 days		95 (0.04)		69 (0.01)		26 (0.03)		4 (0.03)		1 (0.01)		9 (0.03)		8 (0.03)		10 (0.05)		6 (0.04)	
Bone density scan																			
0–182 days (overall)		63 (0.03)		46 (0.03)		17 (0.02)		0 (0.00)		2 (0.02)		9 (0.03)		2 (0.01)		12 (0.06)		10 (0.06)	
0–7 days (PoC)		1 (0.001)		1 (0.001)		0 (0)		0 (0.00)		0 (0.00)		1 (0.003)		0 (0.00)		0 (0.00)		0 (0.00)	
≥ 1 in 0–7 days (PoC), n (%)		1 (0.1)		1 (0.1)		0 (0)		0 (0)		0 (0)		1 (0.3)		0 (0)		0 (0)		0 (0)	
8–182 days		62 (0.03)		45 (0.03)		17 (0.02)		0 (0.00)		2 (0.02)		8 (0.03)		2 (0.01)		12 (0.06)		10 (0.06)	
Sick certification, n (mean per patient)																			
0–182 days (overall)		1181 (0.47)		728 (0.51)		453 (0.43)		33 (0.26)		6 (0.06)		69 (0.24)		52 (0.22)		154 (0.72)		187 (1.12)	
0–7 days (PoC)		325 (0.13)		205 (0.14)		120 (0.11)		17 (0.13)		5 (0.05)		16 (0.06)		26 (0.11)		27 (0.13)		24 (0.14)	
≥ 1 in 0–7 days (PoC), n (%)		213 (8.5)		128 (8.9)		85 (8.0)		7 (5.6)		4 (4.1)		14 (4.9)		16 (6.7)		16 (7.4)		15 (9.0)	
8–182 days		1033 (0.41)		598 (0.42)		435 (0.41)		16 (0.13)		1 (0.01)		53 (0.19)		26 (0.11)		127 (0.59)		163 (0.98)	

TABLE 29 Comparison of GP decision-making per arm (0–182 days) taken from the anonymised EMR audit (continued)

	Anonymised EMR audit data on all patients with a completed trial recruitment template (N = 2494)		Trial participants with known baseline risk subgroup taken from the baseline questionnaire (N = 1130)						
	Both usual care and stratified care, all patients (n = 2494)	Usual care practices, all patients (n = 1438)	Stratified care practices, all patients (n = 1056)	Low risk (n = 224, 20%)		Medium risk (n = 524, 46%)		High risk (n = 382, 34%)	
				Usual care (n = 126)	Stratified care (n = 98)	Usual care (n = 286)	Stratified care (n = 238)	Usual care (n = 215)	Stratified care (n = 167)
Repeat musculoskeletal GP consultations over 6 months									
(8–182 days), n (mean per patient)	2404 (0.96)	1389 (0.97)	1015 (0.96)	51 (0.40)	32 (0.33)	233 (0.81)	217 (0.91)	383 (1.78)	284 (1.7)
PoC, point of consultation.									
a The Keele STarT MSK tool scored as 0–4 = low-risk, 5–8 = medium risk and 9–12 = high risk.									
b Long-term prescribing was calculated as the number of patients who received at least one prescription in both the 0–7-days period, and the 8–182-days period.									
c Receipt of written information by the GP was also asked in the baseline patient questionnaire. Overall 'yes' responses [expressed as n/N (%)] were 306/528 (58.0) in stratified care and 175/672 (26.0) in usual care ( $p < 0.001$ ). 'Yes' responses by subgroup were 65/98 (66.3) in stratified care and 31/126 (24.6) in usual care in the low-risk subgroup; 139/236 (58.9) in stratified care and 79/283 (27.9) in usual care in the medium-risk subgroup; and 84/164 (51.2) in stratified care and 53/213 (24.9) in usual care in the high-risk subgroup.									
Note									
Statistical testing of these process measures was carried out as post hoc analyses with no adjustment for multiple testing owing to their exploratory nature. Analyses were carried out by negative binomial mixed model with practice (random factor) and practice size and participants' point-of-consultation pain score (fixed factors) for count data, except where logistic mixed modelling with the same fixed/random factors was used instead (for binary data, or owing to lack of model convergence and/or small counts). Ratios are for stratified care relative to usual care. Statistically significant $p$ -values ( $p < 0.05$ ) are shown in bold.									





EME  
HSDR  
HTA  
**PGfAR**  
PHR

Part of the NIHR Journals Library  
[www.journalslibrary.nihr.ac.uk](http://www.journalslibrary.nihr.ac.uk)

*This report presents independent research funded by the National Institute for Health and Care Research (NIHR).  
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the  
Department of Health and Social Care*

***Published by the NIHR Journals Library***