Does duration of pain at baseline influence clinical outcomes of low back pain patients managed on an evidence-based pathway?

Mary-Anne Jess [1] MSc

Dr Cormac Ryan [1] PhD

Professor Sharon Hamilton [1] PhD

Dr Shaun Wellburn [1] PhD

Professor Greg Atkinson [1] PhD

Professor Charles Greenough [2] PhD

Dr Andrew Coxon [2] PhD

Diarmaid Ferguson [3] BSc

Professor Francis Fatoye [4] PhD

Dr John Dickson [5] MB ChB

Dr Andrea Jones [6] MB BS

Professor Denis Martin [1] PhD

[1] School of Health and Social Care. Teesside University. Middlesbrough.

[2] South Tees Hospitals NHS Foundation Trust. Middlesbrough.

[3] Northumbria Healthcare NHS Trust, North Tyneside General Hospital, North Shields.

[4] Department of Health Professions, Faculty of Health Psychology and Social Care. Manchester Metropolitan University.

[5] Primary Care Rheumatology Society.

[6] NHS Darlington Clinical Commissioning Group.

Mary-Anne Jess

School of Health and Social Care. Teesside University. Middlesbrough. TS1 3BX

07964234887

m.jess@tees.ac.uk

Sources of support & Acknowledgments:

Funding from The Health Foundation Scaling Up Improvement Programme, the North East Academic Health Science Network and the Primary Care Rheumatology Society, supported the NERBPP. Mary-Anne Jess was supported by funding from Teesside University. We thank all staff and patients involved in the NERBPP, including Kevin Pears for assisting with inputting and validating the data.

Abstract

Study Design: Longitudinal observational study.

Objective: To investigate the association between the duration of pain at baseline and the clinical outcomes of patients with low back pain (LBP) enrolled on the North East of England Regional Back Pain and Radicular Pain Pathway (NERBPP).

Summary of Background Data: The NERBPP is a clinical pathway based upon NICE guidelines (2009) for LBP of <1-year duration. Recent changes to NICE guidelines (2016), advocate the same management for all LBP patients regardless of pain duration.

Methods: Patients with LBP referred onto the NERBPP by their General Practitioner between May 2015 and January 2017 were included. Data from 667 patients, who provided pre-and post-data for pain (Numerical rating scale), function (Oswestry Disability Index), quality-of-life (EuroQol five-dimension, five-level questionnaire), anxiety (the Generalised Anxiety Disorder Screener) and depression (the Patient Health Questionnaire), were analysed using a series of covariate-adjusted models. Patients were categorised into four groups based upon baseline pain duration: <3 months, ≥3 to <6 months, ≥6 months to <12 months, ≥12 months.

Results: Each group showed improved outcomes greater than the minimal clinically important difference (MCID) for each measure as defined in NICE guidelines (2016). There was a trend towards better outcomes for those with shorter pain durations. The magnitude of the differences between the groups, in most instances, was below the MCID. For example, mean improvement in function for those with baseline pain duration <3 months was 20 points and 12 points for those of pain duration ≥12 months, both above the MCID of ≥10.

Conclusions: Patients with different durations of LBP at baseline improved on the NERBPP, supporting the recent modification to NICE guidelines. However, those with shorter durations of pain may have superior outcomes in the short-term, suggesting added benefit in getting patients onto the pathway in the early stages of LBP.

Key Words: low back pain, duration, disability, NICE guidelines.

Key Points:

- The NERBPP is a clinical pathway based upon NICE guidelines (2009) for LBP.
- Patients on the NERBPP of differing LBP duration at baseline improved by a clinically relevant amount, on a suite of outcomes, according to the minimal clinically important difference recommended by NICE (2016).
- There was a trend towards better outcomes for those with shorter pain durations, although the magnitude of the differences between the groups, in most instances, was below the MCID, this would suggest that there is added benefit in getting patients onto the pathway in the early stages of LBP.

Mini Abstract:

This study found that patients with differing durations of LBP at baseline, enrolled on an evidence-based clinical pathway, improved by the minimal clinically important difference recommended by NICE (2016) on a suite of outcome measures. However, those with shorter durations of pain may have superior outcomes in the short-term.

Introduction

Low back pain (LBP) presents a considerable challenge to health care systems globally. It is associated with the longest duration of time spent with disability in many countries.[1] In addition, LBP represents a significant economic burden to individuals, health systems and society including health care resource utilisation and lost work productivity. [2-4]

The updated National Institute for Health and Care Excellence (NICE) guidelines (2016) have been refocused away from traditional duration-based classification of LBP. These guidelines

now advocate the same management for all LBP patients regardless of duration of pain, recommending instead the use of a risk stratification approach to classify patients.[5] This approach contrasts with many other guidelines where the management of patients with acute pain differs from those with chronic pain.[6,7] In-order to support any change in clinical practice, it should demonstrate better outcomes and be valued by patients.[8]

Currently, there is little empirical data to support this change in clinical practice to an approach that encourages the same management of all LBP patients regardless of duration of pain. Only one study has directly investigated the role of baseline pain duration on clinical outcome. Dunn and Croft[9] demonstrated that patients with a longer pain duration (≥3 years) at baseline were associated with poorer clinical outcomes. This indicates that baseline pain duration may be of importance when considering clinical outcome, however, this work looked at patient outcomes following a broad battery of usual care from their GP. As it is difficult to know exactly what was delivered in this usual care, it is difficult to relate back to the modification of the NICE guidelines regarding pain duration.

The North East of England Regional Back Pain and Radicular Pain Pathway (NERBPP) is a clinical manifestation of the NICE LBP guidelines (2009). It has been implemented in first adopter sites in clinical commissioning group regions of the North East of England since 2015. Although targeted originally at patients with acute pain, patients with LBP of varying pain duration access the pathway. It could be proposed that if the recent shift in NICE guidelines is justified, to treat patients of different pain duration similarly, then patients with different pain durations should have similar outcomes on the NERBPP. The aim of this study was to investigate the association between the duration of pain at baseline and the clinical outcomes of patients with LBP enrolled on the NERBPP.

Methods

Study design

This study is part of a large-scale evaluation of the implementation of the NERBPP. Ethical approval for the evaluation was obtained from Teesside University (Reference number R179/15). This was a longitudinal, observational study of patients with LBP, over the age of 18, referred onto the NERBPP by their GP between May 2015 and January 2017 (n=6102).

The pathway was not initially intended for patients with more longstanding pain. Thus, GPs were encouraged to refer onto the pathway: acute, new onset, LBP patients and those experiencing a flare up of a new attack of LBP with at least 6 months since their last attack. The GP screened patients using the STarT Back stratification tool.[10] Those classified as moderate to high risk of poor outcome on the STarT Back tool were referred to a triage and treat practitioner (T&TP) (specially trained nurses and physiotherapists). They assessed all participants and referred them for investigations and/or core therapies (physiotherapy incorporating exercise, manual therapy or acupuncture). There was also the option to refer to a 100-hour residential, combined physical and psychological therapies program for a small number of patients. Patients included for analysis in this study may have received some combination of these interventions.

Data collection

Baseline data collected at the initial T&TP appointment included socio-demographic variables age, sex and socioeconomic status. The date of onset of the patient's symptoms was extracted from the T&TP notes and uploaded to System 1 (electronic patient records system). Duration of baseline pain was calculated by subtracting "date of onset" from "date of GP referral" onto

the pathway. Patients were categorised into four groups based upon their calculated baseline pain duration: <3 months, ≥3 months to <6 months, ≥6 months to <12 months, ≥12 months.

Baseline information was gathered on the STarT Back screening tool score and a battery of standardised, valid and reliable patient reported outcome measures (PROMs). A series of standard outcome measures is recommended to enable easier comparison of results of clinical studies to be made.[11] Pain intensity was measured using the 11-point numerical rating scale (NRS).[12] The level of functional disability was determined using the Oswestry Disability Index (ODI).[13] The EuroQol five-dimension, five-level questionnaire (EQ-5D-5L) was used to measure quality of life,[14] the Generalised Anxiety Disorder Screener (GAD-7)[15] and the Patient Health Questionnaire (PHQ-9)[16] were used to assess anxiety and depression, respectively. Data collected from the GAD-7 and PHQ-9 were analysed for numerical and categorical scores. The GAD-7 numerical was scored from 0-21, while the PHQ-9 was scored from 0-27. The GAD-7 and PHQ-9 categorical scales ranged from: 0= 'Not Difficult at all' to 3= 'Extremely Difficult'.

On discharge, patients were asked to complete all PROMS and give their overall perception of improvement on a six-point Likert-based Global Subjective Outcome Scale (GSOS), the descriptors for which ranged from 'completely better' to 'worse'.[17] Patients were asked to rate their satisfaction with the service they had received using the NHS Friends and Family Test (FFT); a six-point scale ranging from 'extremely likely' to 'don't know'.[18] For an indication of their readiness to self-manage, patients were asked the question: "do you feel ready to self-manage your back pain?" The response was given using a 0-10-point continuous scale, with 0 representing 'not confident' and 10 representing 'totally confident'. This scale was adapted from work by Lorig.[19]

Analysis

Initial analysis involved grouping participants into one of three categories: those discharged at their initial appointment (same-day discharge (SDD)), those discharged following an initial appointment and at least one further appointment (standard discharge (StD)) and those discharged due to non-attendance (non-attender (NA)). The StD group was the only group for whom outcome data was routinely collected and therefore included for analysis in this study. Before commencing statistical analysis, the data were checked for any data entry errors; statistical analysis was carried out using SPSS for Windows (2012).

Outcome data included the pre and post PROMS following management on the NERBPP. Change scores for pain, disability, anxiety, depression and general health status were calculated by subtracting initial scores from discharges scores for each patient that completed follow-up data. The outcomes for the four duration groups were compared using a series of covariate-adjusted models. The outcome measures of interest were the PROMS mentioned previously, duration category was the independent variable and the following were included as covariates: baseline scores, age, sex and socioeconomic status.[20] An ordinal, covariate-adjusted model was used for categorical data. For PROMs, the minimal clinically important difference (MCID), as defined by the NICE (2016) guidelines,[5] was used to establish if clinically relevant improvements were observed in mean change scores. For continuous outcome measures, the MCID is defined as a 10% improvement of a measure of clinical benefit; for the EQ-5D, 0.03 was used as a MCID.[5]

Continuous data were presented as mean (standard deviation) while categorical data were presented as percentage, mean (standard deviation) or median (interquartile range). Comparisons between the characteristics of the discharge categories were made using a

singlefactor general linear model (GLM) for continuous data, and the Kruskal-Wallis H test and/or Chi Square test for categorical data. Statistical significance was set at $p< 0.05$.

Results

Of the 6102 participants, who had baseline data present, 2268 were excluded from the analysis, as they were not yet discharged from the pathway. Of those that were discharged, 2071 were in the StD group, 1147 were SDD and 616 were categorised as NA. As outcome data was not routinely collected from SDD and NA, they were not included in the analysis.

The baseline participant characteristics for the three discharge categories are shown in Table 1. Although there were statistically significant differences, the main clinically relevant differences between the categories were that NA were younger and from lower socioeconomic backgrounds. Participants in the SDD category had the lowest levels of pain, anxiety and depression, the highest proportion of patients in the low risk grouping on the STarT Back, and the highest levels of function and quality-of-life.

Within the StD category, those that provided outcome data were labelled as complete cases (n=667), while those that did not provide any outcome data were labelled as incomplete cases (n=1404). Although the trend for baseline outcome measures for the incomplete cases were statistically poorer on all measures, these differences were not clinically relevant (supplementary table A).

Table 2 shows the mean changes in outcome measures when grouping participants in the StD group into one of the four pain duration categories from <3 months to ≥12 months. For the entire battery of PROMs, all four groups improved by clinically relevant amounts. There was a trend towards better outcomes for those with a shorter duration of pain.

On the GSOS all groups reported improvements in their overall outcome on discharge. There was a significant difference between the groups, those with shorter pain durations reporting greater improvement. For example, 64.8% of those in the <3 months category reported being a lot better/completely better, compared to 44.4% in those groups with longer durations of pain. Over 89% of those in the ≥12 months group, and 93% of those in the <3 months group, were extremely likely/likely to recommend the service to a friend or relative; the differences between the groups were not statistically significant (Table 3).

The baseline values for the PROMs for the pain duration groups are shown in Supplementary Table B. There were no statistically significant or clinically relevant differences between groups.

Discussion

Results of the analysis show that regardless of the duration of pain, all LBP patients enrolled on the NERBPP improved by a clinically relevant amount, for a suite of clinical outcomes, according to the MCID recommended by NICE (2016). Patients with a shorter duration of pain showed a trend towards statistically better outcomes, although the magnitude of the differences between the groups were, in most instances, below the MCID. Stapleton *et al.*[21] highlighted the importance of clinically significant results, rather than interpreting a p-value in isolation.

A key research priority in LBP is the translation of high-quality research findings into clinical practice.[22] The NERBPP fulfils this research priority, as it is the implementation of evidence-

based, NICE guidelines[23] directly into clinical practice. A major change in the revised NICE (2016) guidelines compared to the 2009 guidelines is the emphasis on duration of symptoms. The 2009 guidelines focused on those with LBP of more than six weeks and less than one-year duration, while the 2016 guidelines make no distinction based upon duration of symptoms. Although the NERBPP was initially intended to target patients with acute pain, the majority of patients referred onto the pathway did not meet the set duration criteria provided to GP's. The reasons for this are unclear, though it may stem from the perseverance of pre-existing local practice, which traditionally referred all patients with LBP irrespective of duration of symptoms and the ubiquitous challenge of defining a patient's duration of symptoms in the traditional dichotomy of acute and chronic. Regardless of the reason, this referral pattern points towards the practical challenge that is operationalising a pathway specifically for acute or chronic pain patients, supporting the change in guidelines from a practical perspective. Additionally, the findings of this study provide some support to the change in NICE guidelines, as LBP patients of all pain durations improved by the MCID for all outcomes following management on the NERBPP.

In contrast to this, however, is the fact that for some pain duration categories, the difference between groups, for some outcomes, was greater than the MCID, with 95% CI for the mean change not overlapping. For example, pain reduction in the <3 month group was -3.3 (-3.7, -2.8) compared to -2.0 (-2.3, -1.6) in the ≥12 month group. Using the NICE MCID of 1 on the 0-10 pain NRS, it indicates that those with more chronic pain improve by a clinically lower amount. This suggests that there is added value in getting patients onto the pathway in the early stages of back pain. Overall, although there appears to be benefit for all pain durations, there is room to explore the differential benefits for the subgroups of pain duration categories.

There is a lack of clearly defined criteria to classify chronic pain,[9,24,25] with increasing evidence demonstrating the problems of describing chronic pain based upon duration alone.[25,-27] Dunn and Croft[9] highlight there is little specific research on the association between duration and outcome, beyond the standard dichotomy. Their work found that in LBP patients, duration of pain is a predictor of outcome, regardless of baseline severity and psychological status. A key difference between the current study and that of Dunn and Croft[9] is that their participants received usual care from their GP, while participants in this study were managed on the evidence-based NERBPP.

Strengths of the current study include the use of real life clinical data, gathered as part of everyday clinical practice and the comprehensive number of standardised, valid and reliable PROMs reported. Outcome data was routinely collected only from participants in the StD category. Although there were statistically significant differences between the baseline participant characteristics for the three discharge categories, the main clinically relevant differences were in keeping with clinical expectations of these groups. Baseline values were also added to the statistical model as a covariate, so group differences were adjusted for this factor. Those in the NA category were younger and from lower socioeconomic backgrounds, common characteristics of those less likely to attend appointments.[28] Participants in the SDD category were found to have a better clinical profile, including lower levels of pain, anxiety and depression, and the highest proportion of patients in the low risk grouping on the STarT Back. This correlates with the aims of the NERBPP, using the STarT Back tool, to stratify the management of patients and prevent over-treatment of low-risk patients.[29] Within the StD category, those that did not provide outcome data, i.e. the incomplete cases, had statistically poorer baseline measures, however, these differences were small and of questionable clinical relevance.

Another strength of this study is that although the duration of symptoms were categorised into four pain duration groups, analysis of the data was initially attempted using a continuous method. However, the data was heavily skewed towards more acute durations of less than 3 months, demonstrating that this would be an inappropriate way to analyse the data, hence, a categorised approach was adopted.

A limitation of this study was the data collection process regarding the onset of LBP symptoms. Duration of symptoms was obtained from the patient's notes uploaded to System 1 by the T&TP. As no standardised question was used to obtain this information, there was no clarification as to whether this referred to the patient's first onset of LBP or whether it referred specifically to their current episode. Patients may have answered this question differently with more precise questioning. Dunn *et al*.[30] highlight the importance of the wording of questions, as this could lead to misclassification of patients. Standardised definitions of episode of pain have been proposed to ensure that the question accurately represents what is being asked;[9,31] such definitions could be used in the future for more consistency.

Future research could explore whether the clinical improvements seen in this study, for the four pain duration categories, are maintained over the longer term, as it is known that LBP is characterised by variability and change.[32]

Conclusions:

This study found that patients of differing LBP duration improved by clinically relevant amounts on the NERBPP. However, those with shorter durations of pain may have superior outcomes in the short-term.

References

1. Vos T, Barber RM, Bell B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;22:386(9995):743.
2. Ekman M, Jönhagen S, Hunsche E, et al. Burden of illness of chronic low back pain in Sweden: a cross-sectional, retrospective study in primary care setting. *Spine* 2005;1;30(15):1777-85.
3. Gore M, Sadosky A, Stacey BR, et al. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. *Spine* 2012;15:37(11):E668-77.
4. Hong J, Reed C, Novick D, et al. Costs associated with treatment of chronic low back pain: an analysis of the UK General Practice Research Database. *Spine* 2013:1:38(1):75-82.
5. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management (NICE guideline NG59). Full guideline. 2016. www.nice.org.uk/guidance/ng59/evidence.
6. Airaksinen O, Brox JI, Cedraschi C, et al. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;11:15:s192-300.
7. Van Tulder M, Becker A, Bekkering T, et al. Chapter 3 European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine Journal* 2006:11:15:s169-91.

8. Croft P, Altman DG, Deeks JJ, et al. The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. *BMC Med* 2015;30:13(1):20.
9. Dunn KM, Croft PR. The importance of symptom duration in determining prognosis. *Pain* 2006;31:121(1):126-32.
10. Hill JC, Dunn KM, Lewis M, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthrit Care Res* 2008;15:59(5):632-41.
11. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research: a proposal for standardized use. *Spine* 1998;15:23(18):2003-13.
12. Strong J, Ashton R, Chant D. Pain intensity measurement in chronic low back pain. *Clin J Pain* 1991;1:7(3):209-18.
13. Fairbank JC, Couper J, Davies JB, et al. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66(8):271-3.
14. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;1:20(10):1727-36.
15. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Medi* 2006;22:166(10):1092-7.
16. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002;1:32(9):509-15.
17. Harland NJ, Dawkin MJ, Martin D. Relative utility of a visual analogue scale vs a six-point Likert scale in the measurement of global subject outcome in patients with low back pain receiving physiotherapy. *Physiotherapy* 2015;31:101(1):50-4.
18. NHS England. The Friends and Family Test. NHS England. 2014. Available at: https://www.england.nhs.uk/ourwork/pe/fft/ Accessed January 20, 2017.

19. Lorig K. *Outcome measures for health education and other health care interventions*. Sage; 1996.
20. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *Brit Med J* 2001;10:323(7321):1123-4.
21. Stapleton C, Scott MA, Atkinson G. The 'So What' Factor: statistical versus cinical significance. *Int J Sports Med.* 2009;30(11):773-4.
22. Costa LD, Koes BW, Pransky G, et al. Primary care research priorities in low back pain: an update. *Spine* 2013;15:38(2):148-56.
23. Savigny P, Kuntze S, Watson P, et al. *Low back pain: early management of persistent non-specific low back pain*. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners; 2009.
24. Andersson GB. Epidemiological features of chronic low-back pain. *Lancet* 1999;14:354(9178):581-5.
25. Cedraschi C, Robert J, Goerg D, et al. Is chronic non-specific low back pain chronic? Definitions of a problem and problems of a definition. *Brit J Gen Pract* 1999;1:49(442):358-62.
26. Von Korff M, Dunn KM. Chronic pain reconsidered. *Pain* 2008;31:138(2):267-76.
27. Kongsted A, Kent P, Axen I, et al. What have we learned from ten years of trajectory research in low back pain? *BMC Musculoskelet Dis* 2016;21:17(1):220.
28. George A, Rubin G. Non-attendance in general practice: a systematic review and its implications for access to primary health care. *Fam Pract* 2003;1:20(2):178-84.
29. Hill JC, Whitehurst DG, Lewis M, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet* 2011;4:378(9802):1560-71.

30. Dunn KM, de Vet HC, Hooper H, et al. Measurement of back pain episode inception in questionnaires: a study combining quantitative and qualitative methods. *J Musculoskelet Pain* 2006;1:14(1):29-37.

31. de Vet HC, Heymans MW, Dunn KM, et al. Episodes of low back pain: a proposal for uniform definitions to be used in research. *Spine* 2002;1:27(21):2409-16.

32. Von Korff M, Saunders K. The course of back pain in primary care. *Spine* 1996;15:21(24):2833-7.

Table 1: Participant characteristics for group discharge categories.

	Standard n= 2071	Same-day n= 1147	Non-attender n= 616	*p*-value
Age (years)	52.8 (16.2)	52.1 (15.9)	44.5 (14.6)	<0.001
Sex (female)	59.3 %	57.5 %	61.0 %	0.340
Socioeconomic Status (1-10)	5 (1-8)	5 (1-8)	4 (1-7)	<0.001
STarTBack score (0-9)	6.4 (1.8)	5.9 (2.0)	6.5 (1.9)	<0.001
STarTBack risk (low risk)	7.7 %	13.5 %	7.6 %	<0.001
Symptom duration (months)	42 (78)	39 (80)	43 (73)	0.484
Pain NRS (0-10)*	6.9 (1.8)	6.5 (2.1)	7.1 (1.7)	<0.001
ODI (0-100%)*	44.6 (17.7	37.9 (19.4)	43.9 (16.4)	<0.001
EQ-5D (1 to -0.594) #	0.4 (0.3)	0.5 (0.3)	0.5 (0.3)	<0.001
EQ-5D VAS (0-100%) #	52.5 (24.8)	59.6 (22.4)	53.6 (23.0)	<0.001
GAD-7 (0-21)*	8.2 (6.5)	6.7 (6.5)	9.4 (6.4)	<0.001
PHQ-9 (0-27)*	9.6 (7.3)	7.7 (7.2)	10.8 (7.6)	<0.001
GAD-7 categorical *	1.1 (0.9) 1 (1-2)	0.9 (0.9) 1 (0-1)	1.1 (0.8) 1 (1-2)	<0.001
PHQ-9 categorical *	1.1 (0.9) 1 (0-2)	0.9 (0.9) 1 (0-1)	1.1 (0.9) 1 (0-2)	<0.001

Data are mean (SD), percentage (%), or median (IQR). Not all participants provided data for each of the variables, the total for each variable is provided. STarTback risk= dichotomy of scores: low risk= score of 0-3, score of >3= at risk. (n=3446). NRS= Numerical Rating Scale (n=2173). ODI= Oswestry Disability Index (n=2488). EQ-5D= EuroQol five-Dimension Questionnaire (n=2555), VAS= Visual Analogue Scale (n=2496). GAD-7= Generalised Anxiety Disorder Screener (n=2197). PHQ-9= Patient Health Questionnaire (n=2208). GAD-7 categorical= data from question 8 on the GAD-7 (n=2343); PHQ-9 categorical data from question 10 on PHQ-9 (n=2380). *Higher NRS, ODI, GAD-7 and PHQ-9 scores are worse. # Lower EQ-5D and EQ-5D VAS scores are worse.

Table 2: Mean change for patient reported outcome measure for standard discharge patients, categorisation based on duration of pain.

Variable	n	< 3 months	≥3-<6 months	≥6-<12 months	≥12 months	*p*-value
Pain NRS (0-10) *	407	-3.3 (-3.7, -2.8) a	-3.1 (-3.6, -2.6) b	-2.4 (-3.0, -1.8) c	-2.0 (-2.3, -1.6)	0.06
ODI (0-100%) *	491	-20 (-23, -17) a	-18 (-21, -15) b	-15 (-19, -12) c	-12 (-14, -10)	0.05
EQ-5D Value (1 to -0.594) #	500	0.3 (0.2, 0.3) a	0.2 (0.2, 0.3) b	0.2 (0.2, 0.3)	0.2 (0.1, 0.2)	0.04
EQ-5D VAS (0-100%) #	507	14.3 (10.9, 17.7)	17.7 (14.0, 21.4) b	13.5 (9.2, 17.8)	10.8 (8.0, 13.5)	0.02
GAD-7 (0-21) *	401	-3.7 (-4.5, -2.9) a	-2.9 (-3.8, -2)	-2.6 (-3.7, -1.6)	-2.1 (-2.7, -1.4)	0.24
PHQ-9 (0-27) *	409	-4.4 (-5.3, -3.6) a	-3.8 (-4.7, -2.8)	-2.9 (-4.0, -1.7) c	-2.8 (-3.5, -2.1)	0.24
GAD-7 categorical*	374	0.5 (0.7)	0.4 (0.6)	0.6 (0.8)	0.7 (0.7)	<0.01
		0 (0-1)	0 (0-1)	0 (0-1)	1 (0-1)	
PHQ-9 categorical*	534	0.4 (0.6)	0.3 (0.5)	0.6 (0.7)	0.6 (0.8)	0.01
		0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	
Self-management #	638	6.4 (6.0-6.8)	6.8 (6.3-7.3)	6.3 (5.8-6.8)	5.9 (5.6-6.2)	0.17

Data are Mean change (95% Confidence Interval Lower, Upper Bound), Mean (SD) and Median (IQR) by use of covariate adjusted models for: age, sex, socioeconomic status and baseline score for the outcome measure. a: statistically significant better outcome at < 3 months than at ≥12 months. b: statistically significant better outcome at ≥3-<6 months than at ≥12 months. c: statistically significant better outcome at <3 months than at ≥6- <12 months. Not all participants provided data for each of the variables, numbers are given for each duration category (n= < 3 months; ≥3-<6 months; ≥6<12 months; ≥12 months) NRS= Numerical Rating Scale (n=106; 86; 59; 156). ODI= Oswestry Disability Index (n= 125; 104; 75; 187). EQ-5D= EuroQol five-Dimension Questionnaire, Value (n=128; 108; 77; 187) VAS= Visual Analogue Scale (n= 127; 108; 79; 193). GAD-7= Generalised Anxiety Disorder Screener (n=103; 84; 60; 154). PHQ-9= the Patient Health Questionnaire (n=100; 85; 62; 162). GAD-7 categorical= data from question 8 on the GAD-7 (n=86; 85; 52; 151); PHQ-9 categorical data from question 10 on PHQ-9 (n=128; 108; 81; 217). Self-management (n= 168; 125; 93; 252) *Higher NRS, ODI, GAD-7 and PHQ-9 scores are worse. #Lower EQ-5D Value, EQ-5D VAS and self-management scores are worse.

Table 3: Categorical data for GSOS and FFT, categorisation based on duration of pain.

	< 3 months	≥3-<6 months	≥6-<12 months	≥12 months	*p*-value
GSOS	**n=162**	**n=116**	**n=91**	**n=241**	<0.01
Completely better	10.5%	4.3%	3.3%	5.0%	
A lot better	54.3%	56.0%	42.9%	39.4%	
Moderately better	14.8%	19.0%	22.0%	16.2%	
A little better	9.9%	10.3%	17.6%	14.5%	
Same	8%	7.8%	13.2%	21.6%	
Worse	2.5%	2.6%	1.1%	3.3%	
FFT	**n=166**	**n=122**	**n=97**	**n=250**	0.22
Extremely likely	75.3%	71.3%	66.0%	62.0%	
Likely	18.7%	22.1%	24.7%	27.6%	
Neither likely or unlikely	1.2%	4.9%	5.2%	5.2%	
Unlikely	1.2%	0.0%	1.0%	0.8%	
Extremely unlikely	2.4%	0.8%	0.0%	1.2%	
Don't know	1.2%	0.8%	3.1%	3.2%	

GSOS= Global Subjective Outcome Scale. FFT= Friends and Family Test

Supplementary Table A: Participant characteristics for standard discharge patients who have and have not provided outcome data.

	Incomplete cases (n=1404)	Complete case (n=667)	*p*-value
Age (years)	52.0 (16.0)	54.7 (16.5)	<0.001
Sex (female)	60.1%	57.7 %	0.300
Socioeconomic Status	5 (1-8)	5 (1-8)	0.261
Startback score	6.4 (1.9)	6.3 (1.8)	0.103
Startback risk (low risk)	8.0%	7.2%	0.528
Symptom duration (months)	43 (79)	37 (74)	0.073
Pain NRS (0-10) *	7.0 (1.7)	6.8 (1.8)	0.048
ODI (0-100%) *	45.7 (17.9)	42.9 (17.2)	0.002
EQ-5D Value (1 to -0.594) #	0.4 (0.3)	0.4 (0.3)	0.017
EQ-5D VAS (0-100%) #	50.6 (21.2)	55.1 (28.8)	0.001
GAD-7 (0-21) *	8.6 (6.7)	7.4 (6.2)	0.001
PHQ-9 (0-27) *	10.3 (7.3)	8.7 (7.2)	<0.001
GAD-7 categorical *	1.2 (0.9) 1 (1-2)	1.1 (0.8) 1 (0-2)	0.029
PHQ-9 categorical *	1.2 (0.9) 1 (1-2)	1.0 (0.8) 1 (0-1)	0.002

Data are mean score (SD), percentage (%), or median (IQR). Not all participants provided data for each of the variables, the total for each variable is provided. STarTback risk= dichotomy of scores: low risk= score of 0-3; score of >3= at risk (n=1952). NRS= Numerical Rating Scale (n=1271). ODI= Oswestry Disability Index (n=1493). EQ-5D= EuroQol five-Dimension Questionnaire, with Value (n=1524), with VAS= Visual Analogue Scale (n=1506). GAD-7= Generalised Anxiety Disorder Screener (n=1306). PHQ-9= Patient Health Questionnaire (n=1314). GAD-7 categorical= data from question 8 on the GAD-7 (n=1399); PHQ-9 categorical data from question 10 on PHQ-9 (n=1427). * Higher NRS, ODI, GAD-7 and PHQ-9 scores are worse. # Lower EQ-5D and EQ-5D VAS scores are worse.

Supplementary Table B: Baseline values for patient reported outcome measures, categorisation based on duration of pain.

Variable	n	**< 3 months**	**≥3-<6 months**	**≥6-<12 months**	**≥12 months**	***p*-value**
Pain NRS (0-10) *	407	6.6	6.7	7.2	6.9	0.15
ODI (0-100%) *	491	43.0	40.0	42.4	42.9	0.42
EQ5D Value (1 to -0.594) #	500	0.44	0.49	0.43	0.45	0.43
EQ5D VAS (0-100%) #	507	59.2	57.5	56.0	53.7	0.42
GAD7 (0-21) *	401	6.4	6.4	6.5	7.4	0.50
PHQ9 (0-27) *	409	8.4	7.6	6.9	9.0	0.16
GAD-7 categorical*	374	1.1	1.0	1.0	1.0	0.55
PHQ-9 categorical*	534	1.1	0.9	0.8	1.0	0.16

Data are Mean score. Not all participants provided data for each of the variables, numbers are given for each duration category (n= < 3 months; ≥3-<6 months; ≥6<12 months; ≥12 months) NRS= Numerical Rating Scale (n=106; 86; 59; 156). ODI= Oswestry Disability Index (n= 125; 104; 75; 187). EQ-5D= EuroQol five-Dimension Questionnaire, Value (n=128; 108; 77; 187) VAS= Visual Analogue Scale (n= 127; 108; 79; 193). GAD-7= Generalised Anxiety Disorder Screener (n=103; 84; 60; 154). PHQ-9= the Patient Health Questionnaire (n=100; 85; 62; 162). GAD-7 categorical= data from question 8 on the GAD-7 (n=86; 85; 52; 151); PHQ-9 categorical data from question 10 on PHQ-9 (n=128; 108; 81; 217). *Higher NRS, ODI, GAD-7 and PHQ-9 scores are worse. #Lower EQ-5D Value, EQ-5D VAS and self-management scores are worse.