



## Abstract

Choosing the most appropriate treatment for individual patients with low back pain (LBP) can be challenging, and clinical guidelines recommend taking into account patients' preferences. However, no tools exist to assess or compare patients' views about LBP treatments. We report the development and validation of the LBP Treatment Beliefs Questionnaire (LBP-TBQ) for use across different treatments in clinical practice and research. Using qualitative data we developed a pool of items assessing perceived credibility, effectiveness, concerns about and individual 'fit' of specific treatments. These items were included in a survey completed by 429 primary care patients with LBP, of whom 115 completed it again 1 to 2 weeks later. We performed psychometric analyses using non-parametric item response theory and classical test theory. The four subscales of the resulting 16-item LBP-TBQ showed good homogeneity ( $H=.46-.76$ ), internal consistency ( $\alpha=.73-.94$ ), and stability ( $r=.63-.83$ ), confirmed most convergent and discriminant validity hypotheses, and had acceptable structural validity for four guideline-recommended treatments: pain medication, exercise, manual therapy and acupuncture. Participants with stronger positive treatment beliefs were more likely to rank that treatment as their first choice, indicating good criterion validity ( $t$  values=3.11-9.80, all  $p<.01$ , except pain medication effectiveness beliefs,  $t(339)=1.35$ ;  $p=.18$ ). A short 4-item version also displayed good homogeneity ( $H=.43-.66$ ), internal consistency ( $\alpha=.70-.86$ ), and stability ( $r=.82-.85$ ), and was significantly related to treatment choice ( $t$  values=4.33-9.25, all  $p<.01$ ). The LBP-TBQ can be used to assess treatment beliefs in primary care patients with LBP and to investigate the effects of treatment beliefs on treatment uptake and adherence.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial No Derivatives 3.0 License, which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially.

## Introduction

Low back pain (LBP) is a leading cause of disability world-wide [42] and is managed mostly within primary care. Most patients have non-specific LBP [1;15] and 75% may continue to have pain and/or disability 12 months after the initial consultation [13]. Clinical guidelines recommend several treatments, including conventional (e.g. education, exercise, pain medication), complementary/alternative (CAMs) (e.g. acupuncture, manual therapy) and combined physical and psychological treatments (pain management courses) [1;11;49;50]. The clinical challenge is to choose optimal treatments for individuals; clinical guidelines explicitly encourage considering patients' preferences [1;11;49;49;50;50], but offer no recommendations on how to elicit and integrate them into decision-making. Clear conceptualization and standardized assessment of patients' preferences would facilitate further research and possible subsequent integration into practice.

Treatment preferences can be understood within the extended Common-Sense Model (CSM) of illness representations [33]. This model stipulates that, when confronted with a medical problem, patients develop cognitive and emotional representations of their condition and beliefs about possible treatments ("treatment beliefs" [26] based on information from various sources), which guide their behaviours (e.g., treatment choice) and can predict subsequent clinical outcomes (e.g., pain). Significant relationships have been found between illness representations, treatment beliefs, and outcomes such as adherence and satisfaction in various chronic conditions [7;22;24;27;28;44] including LBP [17;20]. According to the CSM, treatment preferences develop when patients attempt to "match" treatments to their condition, aiming for coherence between illness representations and treatment beliefs. For example, patients who believe their LBP is caused by a mechanical problem may prefer treatments they believe can remedy mechanical dysfunctions and choose manual therapy; patients who see

LBP as essentially a pain symptom may prefer treatments they consider appropriate to reduce pain, and choose pain medication. Reliable and valid measurement of treatment beliefs in LBP is needed to further test such hypotheses derived from the CSM and facilitate shared decision-making.

Illness perceptions have been examined extensively: validated questionnaires are available [10;41;63] and have been used in LBP research [17;20]. However, we could not identify a treatment beliefs questionnaire applicable to different LBP treatments that concomitantly assesses several relevant beliefs. Existing measures are treatment-specific [9;18;29;34;37;60;62], and previous studies in LBP have focused on single belief dimensions, e.g. expectations of effectiveness [19;43;52;58], or perceived credibility [55]. However, qualitative research suggests that LBP treatment beliefs are multidimensional [21;25]. In our recent qualitative study, patients evaluated LBP treatments according to four specific dimensions: perceived credibility, individual fit, concerns, and effectiveness [16]. Here we report the development and validation of a questionnaire, the Low Back Pain Treatment Beliefs Questionnaire (LBP-TBQ), which assesses patients' beliefs about four practitioner-delivered primary care treatments: pain medication, exercise, manual therapy, and acupuncture. We focused on these treatments as they are the frontline treatments named in the National Institute for Health and Care Excellence (NICE) care pathway for persistent non-specific LBP [49;50], and are also recommended by the American College of Physicians and the American Pain Society LBP guidelines [9]; pain medication, exercise, and manual therapy are also endorsed in European guidelines for chronic non-specific LBP [1].

## **Methods**

### **Instrument Development**

In our previous qualitative study [16] we showed how 75 patients participating in 13 focus groups evaluated specific LBP treatments according to whether they: perceived them to be believable and to ‘make sense’ (credibility); expected them to lead to symptom improvements (effectiveness); had concerns that treatments might cause further damage or have side-effects (concerns); felt the treatment would be a suitable solution for them personally (individual fit). Because patients expressed these beliefs about specific treatments (e.g. pain medication, acupuncture) we termed these Specific Treatment Beliefs. Themes reflecting the context of treatment decision-making also emerged and highlighted the importance of understanding patients’ more general treatment-seeking beliefs: their need for a clear diagnosis, their willingness to try different treatments, their interest in self-management and their expectations regarding the healthcare system [16]. We developed an item pool comprising 71 items, 27 items assessing the Specific Treatment Beliefs (the focus of this paper) and 44 assessing the contextual themes (to be reported elsewhere).

We reviewed our qualitative data to choose item content and wording that reflected topics and terminology used by participants. To facilitate comparisons between patients’ beliefs about different treatments, items assessing Specific Treatment Beliefs were designed to be answered four times, once each in relation to: pain medication, exercise, manual therapy, and acupuncture. Therefore, issues specific to particular treatments (e.g. fear of needles in acupuncture) were not included. Remaining items were worded more generically in order to capture these specific issues (e.g. “I have concerns about [acupuncture] for my back pain”).

We pre-tested the initial pool of 27 items using cognitive “think aloud” interviews [64] with 10 adults with LBP. This pre-test allowed us to select the most appropriate items for further testing and to adjust item content and wording to enhance face validity and acceptability. After the pre-test we retained 20 items on Specific Treatment Beliefs in the

Draft LBP-TBQ for psychometric testing. The reasons for exclusion were: 1) participants interpreted the item in a different way from the intended meaning (1 item); 2) the item was too similar to another item that was perceived as clearer (4 items); 3) the respondents had difficulties applying it to all four treatments (1 item); or 4) the item was more related to the context of care than to the treatment itself (1 item). We opted for a lower number of items and a confirmatory approach to psychometric testing (instead of a higher number of items and an exploratory approach) because of the increased patient burden involved in answering questions repeatedly for each of the four treatments and to facilitate analysis of the structure of the questionnaire across all four treatments.

### **Design and Procedure**

We included the Draft LBP-TBQ, items on the context of treatment decision-making, several validating measures, and questions on demographic and clinical characteristics in a self-report survey of adults (at least 18 years) with LBP. We included adults who reported LBP for at least 6 weeks because our prior qualitative work revealed that, although the NICE guidelines particularly focus on persistent non-specific LBP (i.e. pain not caused by malignancy, infection, fracture, inflammatory disorders, nerve root compression, and lasting between 6 weeks and 12 months), the distinction between persistent and chronic LBP is rarely used in practice by clinicians or patients [6;16]. We aimed to examine whether our questionnaire applies to all people experiencing LBP for more than 6 weeks, irrespective of duration of complaint, whether LBP is non-specific (e.g. report a diagnosis of sciatica, or symptoms that can be clinically interpreted as nerve compression), or whether patients are treatment-experienced or treatment naïve. Therefore, we did not apply additional exclusion criteria but compared responses to our questionnaire across different sub-groups of patients.

Participants were recruited between November 2011 and March 2012 from public sector primary care physicians (General Practitioners, GPs) and private sector CAM clinics in three South England counties (Hampshire, Wiltshire and Dorset), and advertisements on online UK-based patient forums. We aimed for 400 participants, a statistically-acceptable sample size for our planned psychometric analyses, acknowledging that statistical power also depends on data properties that could not be estimated prior to analysis [38;57;65]. Physicians and CAM clinicians forwarded paper-based surveys to their eligible patients by post. Online advertisements linked directly to an identical web-based survey. To enable examination of test-retest reliability, participants were asked to volunteer to complete the LBP-TBQ again; all such volunteers were sent a second survey by post or email one week later. We obtained ethics approval from Southampton and South West Hampshire REC B (10/H0504/78).

### **Draft Low Back Pain Treatment Beliefs Questionnaire (LBP-TBQ)**

In the 20-item Draft LBP-TBQ, 4 items assessed perceptions of credibility (2 negatively-worded, i.e. described in terms of doubting the credibility of the treatment), 5 items assessed perceived effectiveness (2 negatively-worded), 6 items assessed concerns (4 negatively-worded) and 5 items assessed perceived individual fit (3 negatively-worded) (see Table 2 for item content). A 5-point verbal response scale was used for all items (Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree), and scored (1-5) such that a high score represented positive beliefs about the LBP treatment. Each set of 20 items was presented with respect to each of the four treatments – pain medication, exercise, manual therapy and acupuncture (so each participant responded to 80 LBP-TBQ items). Definitions of these treatments, based on the UK NICE guidelines [49;50], were provided to limit variability in interpreting treatment labels and encourage answers that can be interpreted

within the context of UK clinical practice; these definitions may need to be adjusted for different purposes in future research (subject to confirmation of psychometric properties in specific other contexts and populations).

### **Validating Measures**

We developed hypotheses about relationships between each validating measure and the Draft LBP-TBQ subscales (see below and Table 3). In brief, to demonstrate convergent validity, we required at least medium or strong significant correlations (i.e.,  $r \geq 0.3$ ). To demonstrate discriminant validity, we required at most small to moderate significant correlations ( $r < 0.3$ ) [12]. Cronbach's alpha values reported below for the validating measures were all calculated in the present sample.

**Beliefs about Medicine Questionnaire (BMQ) - General Beliefs [29].** Respondents completed 5-point scales to rate their agreement with statements representing beliefs about the potential harmful effects of medicines (BMQ-Harm, 4 items, e.g. 'medicines do more harm than good', Cronbach's  $\alpha = .68$  in the present sample) and about medicines being over-prescribed by doctors (BMQ-Overuse, 4 items, e.g. 'doctors use too many medicines',  $\alpha = .76$ ). High scores indicated more negative beliefs about medicines. Both BMQ scales were used to assess the convergent validity of the LBP-TBQ Concerns subscale for pain medication.

**The Brief Illness Perceptions Questionnaire (BIPQ) [10].** Single items with 11-point response scales assessed 8 dimensions of illness perceptions: consequences (the extent to which LBP affects one's life), timeline (the expected duration of LBP), personal control (the extent to which one perceives control over one's LBP), treatment control (the extent to which one perceives one's treatment controls one's LBP), identity (the number of symptoms associated with LBP), coherence (the extent to which one understands one's LBP), concern



(the extent of concerns about LBP), emotional response (the extent of emotional distress attributed to LBP). We worded all items to refer to 'your low back pain' instead of 'your illness.' Perceptions of causes of LBP were investigated using an adapted version of the 'perceptions of illness causes' subscale of the revised IPQ [41] that requires respondents to agree or disagree (on 5-point response scales) that each of 18 factors was a cause of their LBP (reliability not applicable as no total scores were computed). We retained existing items potentially relevant to patients' perceptions of the causes of their LBP and replaced other items (pollution in the environment; alcohol, smoking; my personality; altered immunity) with commonly perceived causes of LBP (malformation of the spine; pregnancy or giving birth; wear and tear; a physical problem in my back, e.g. a 'slipped disc'; a specific disease in my back, e.g. osteoporosis), using data from a previous questionnaire-based study [8], our qualitative work [16], and our clinical and research experience. The BIPQ concern and emotional response items were used to assess the divergent validity of all four LBP-TBQ subscales for all four treatments.

**Credibility Expectancy Questionnaire (CEQ) [14].** Two subscales assessed perceptions of treatment credibility (CEQ-Credibility, 3 items with 9-point response scales, e.g. 'At this point, how logical does [treatment] seem?') and outcome expectancy (CEQ-Expectancy, 1 item with 9-point response scale and 2 items with a 11-point response scale, from 0% to 100%, e.g. 'By the end of a course of [treatment], how much improvement in your back pain do you think would occur?'). To reduce response burden, each respondent answered the CEQ in relation to one of the four treatments only (randomised allocation). Good internal consistency was shown in our sample for credibility ( $\alpha$  range .85 - .94) and expectancy ( $\alpha$  range .85-.96) scales for all treatments. High scores indicated perceiving the treatment as more believable, convincing and logical, and as leading to bigger improvements. CEQ-Credibility was used to assess the convergent validity of the LBP-TBQ credibility

subscale for all four treatments. CEQ-Expectancy was used to assess the convergent validity of the LBP-TBQ expectancy subscale for all four treatments.

### **Holistic Complementary and Alternative Medicine Questionnaire (HCAMQ)**

**[31] Attitudes to CAM subscale.** Six statements assessed general attitudes towards CAM using 6-point agree/disagree response scales (e.g. 'It is worthwhile trying complementary medicine before going to the doctor',  $\alpha = .71$ ). High scores indicated stronger beliefs that CAM is ineffective and unscientific compared to mainstream medicine. The HCAMQ Attitudes to CAM subscale was used to assess the convergent validity of all four LBP-TBQ subscales for manual therapy and acupuncture.

**Tampa Scale of Kinesiophobia (TSK-11) Activity Avoidance subscale [48].** Six items assessed beliefs about the relationship between movement, pain and re-injury (TSK-Activity Avoidance, e.g. 'I'm afraid I might injure myself if I exercise',  $\alpha = .81$ ) using 4-point agree-disagree response scales. High scores indicated more intense concerns and beliefs that movement leads to (re)-injury and should therefore be avoided. The TSK-Activity Avoidance subscale was used to assess the convergent validity of the LBP-TBQ concerns subscale for exercise.

**Treatment ranking.** To assess criterion validity, participants were asked to rank the four treatments (pain medication, exercise, manual therapy, acupuncture) in order of preference starting with the treatment they would most like to have. Choices regarding each treatment were coded separately and dichotomised to identify two groups of patients for each treatment: those who did and those who did not select that treatment as their first choice. We hypothesised that scores on the LBP-TBQ subscales would distinguish between these two groups of patients for each treatment.

### **Participants' Demographic and Clinical Characteristics**

We used single items to assess socio-demographic characteristics (age, gender, education level, employment status, ethnicity) and clinical characteristics (duration of LBP, previous/current use of various treatments for LBP and satisfaction with these treatments, self-reported co-morbidities including sciatica, receiving benefits for LBP, perception of general health status).

LBP severity over the past 6 months was assessed with the Chronic Pain Grade Questionnaire [61] which comprises 3 pain intensity items (present, worst, and average) and 4 disability items (number of days kept from usual activities, and pain interference with daily activities, recreational or social activities, and work). Responses were used to compute 5 pain grades, from no pain and disability through to highest disability irrespective of pain intensity [56]. Participants who had experienced LBP in the past year (single item), were asked whether they currently experience pain and whether they had experienced 3 symptoms suggestive of a differential diagnosis of nerve root compression: leg pain worse than back pain, leg pain worse when coughing or sneezing, and numbness or pins and needles in the leg or feet. Participants who reported any of these symptoms were classified as more likely to have nerve root compression.

### **Data Analysis**

Data analysis was performed in SPSS Statistics 21 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.), AMOS 21 [2] and R [mokken package; 3;4;47]. Data entry was checked for accuracy. We identified no systematic pattern of missing data for the selected variables (Little's MCAR test;  $\chi^2(36655) = 36771.69$ ,  $p = .33$ ). Missing data were computed via expectation maximization for the relevant questionnaire items (excluding items where non-completion was expected based on responses to previous items). We computed descriptive statistics for demographics, clinical data, and

validation measures. We performed psychometric analyses separately for the specific and general treatment beliefs items.

**Item selection and structural validity.** We used a confirmatory approach to questionnaire validation and aimed for a final item set that included an optimal, parsimonious, number of items that would permit the use of both sum scores and individual items in latent variable models (i.e. 4 items per construct, of which 2 should be negatively-worded). We examined the structure of the item sets on the first-wave dataset using non-parametric item response theory (NIRT), i.e. Mokken Scale Analysis (MSA) [23;39;40;53;54;59]. According to MSA, items order respondents stochastically on one latent dimension representing the target construct if they meet three criteria: 1. unidimensionality (i.e. respondents that endorse more intense items are also more likely to endorse less intense items, while endorsing less intense items is not related to the probability of endorsing more intense items), 2. local independence (i.e. the statistical relationship between items should be explained solely by the latent construct), and 3. latent monotonicity (i.e. the probability of endorsing an item should not decrease with increasing levels of the latent construct). We investigated these properties by calculating coefficients of homogeneity (H) at item, item-pair, subscale, and scale level. Homogeneity values range from 0 (no association) to 1 (perfect association given differences in item intensity), where .3-.4 indicates weak, .4-.5 medium, and values over .5 good homogeneity. To reduce scale length, items presenting low homogeneity and violations of latent monotonicity were considered for exclusion. We subsequently examined correlations between the resulting subscales.

We further investigated structural validity using confirmatory factor analyses (CFA), since MSA is nonparametric and therefore does not allow modelling the effect of using positively- and negatively-worded items. Models were evaluated in relation to established criteria for the likelihood ratio  $\chi^2$  test, incremental fit and residual-based indices (TLI and

CFI>.95; RMSEA and  $\chi^2$  *p* value <.05) [30;32]. We followed recommendations to judge model fit flexibly within the broader context of model diagnosis and theory and to consider fit criteria alongside overall model tenability and possible sources of misspecification [35;36].

To investigate alternative structures, we used an automated item selection procedure (*aisp*) to group items into Mokken scales in a data-driven manner and identify un-scalable items at increasing homogeneity threshold values (Hemker et al 1995).

**Reliability.** The new subscales were also examined according to classical test theory. Internal consistency was assessed via Cronbach's alpha (above .70), item-total correlations, and Cronbach's alpha if item excluded. Correlations between responses to the same scales at first and second survey administration were used to judge test-retest reliability over 1-2 weeks based on a threshold of  $r=.70$  [45].

**Convergent and discriminant validity.** The subscales were tested for convergent and discriminant validity against existing measures using Pearson's correlations (sensitivity analyses were performed with Spearman's  $\rho$  and Kendall's  $\tau$ , to account for the ordinal level of some measures).

**Criterion validity.** The new LBP-TBQ subscales were examined against treatment ranking reports to assess their ability to predict treatment preferences. We used t-tests to compare scores between participants who did and did not rank each treatment as their first choice.

**Measurement invariance (MI).** Multi-group CFA analyses were performed to investigate whether different subgroups of respondents attribute the same meaning to the target construct (metric invariance) and whether respondents with equal scores on the latent construct also have equal scores on the items (scalar invariance). If a scale has these properties, then group differences in mean scores can be interpreted as substantive

differences as they are not due to participants in different groups attributing different meanings to the scale, or to measurement bias of individual items in these groups [51]. Five comparisons were considered: respondents with nerve root compression likely or unlikely (based on self-reported presence of one or more of three indicative symptoms); respondents that self-reported sciatica diagnosed in relation to LBP or not; respondents with or without experience of each treatment; and the two data collection waves.

**Selection and validation of the short-form LBP-TBQ.** To increase the feasibility of using the LBP-TBQ in multi-measure patient surveys and fast-paced clinical contexts, we developed a short 4-item LBP-TBQ version by selecting one best performing item (i.e. higher homogeneity and lower Cronbach  $\alpha$  if-item-excluded, for all treatments) for each dimension of treatment beliefs. We examined homogeneity, internal consistency, stability, and criterion validity of this scale following the procedures described above.

## Results

### Participants

The survey was completed by 429 participants, of whom 344 (80%) responded to the 1498 invitations mailed to public sector physicians' patients (23% response rate). Participants were aged 18 to 90 years ( $M = 55$ ;  $SD = 15.2$ ); 247 (60%) were women and 393 (91.6%) were of British, Irish or other white ethnic background. The majority (335; 78.1%) completed the paper-version. Participants reported having LBP for between 6 weeks and 52 years (median 6 years; interquartile range 13.18 years); 415 (96%) had LBP in the past year, 400 (93.2%) in the past 6 months, and 308 (71.8%) at the time of the survey; 398 (92.8%) considered their general health to be 'fair' to 'very good', but more than half reported high disability (i.e. chronic pain grade III or IV). Only 61 (15.5%) reported receiving state benefits for LBP. Of the 174 volunteering to complete the survey twice, 115 (66.1%) participants completed and

returned the LBP-TBQ again 1-2 weeks later. There were no differences between respondents and non-respondents to the second survey in age, gender, pain duration, general health levels, pain intensity, disability levels, or chronic pain grade. Additional socio-demographic and clinical data and descriptive statistics for the validating measures are available in Table 1 and Supplementary Digital Content 1 (available online at <http://links.lww.com/PAIN/A79>).

*Insert Table 1 Here*

### **Item selection and structural validity**

Items were selected iteratively based on homogeneity and monotonicity results at item, item-pair, and scale level for all four treatments. Four items were excluded based on item properties and content to achieve four 4-item subscales with two reversed items each and good homogeneity and monotonicity, except three significant violations of monotonicity for Acupuncture items (Table 2; see also Supplementary Digital Content 2 which shows initial homogeneity values for all items and the violations of monotonicity for acupuncture items [available online at <http://links.lww.com/PAIN/A80>]).

*Insert Table 2 Here*

We examined structural validity via CFA for each treatment by comparing the hypothesized 4-factor model (covariance between factors and covariances between error terms of reverse-coded items) with several alternatives (1-factor model, 4-factor model with 1 common higher-order factor, 1-factor model improved via specifying error covariances suggested by modification indices). Although none of the models reached threshold values for all fit indices and all treatment types, the 4-factor model performed slightly better than its

alternatives. As an example, Figure 1 displays model parameter estimates for pain medication items. (Supplementary Digital Content 3 shows parameter estimates and model fit for alternative models and other treatment types [available online at <http://links.lww.com/PAIN/A81>].)

*Insert Figure 1 Here*

**Reliability.** All four subscales showed acceptable internal consistency, with Cronbach's  $\alpha$  values ranging from .73 to .94. They also showed acceptable test-retest validity: over 1-2 weeks pain medication effectiveness exhibited the lowest stability ( $r=.63$ ), and manual therapy concerns exhibited the highest stability ( $r=.83$ ). For details see Table 3.

*Insert Table 3 Here*

### **Convergent and discriminant validity**

As hypothesized, on the whole the LBP-TBQ subscales were at least moderately associated with conceptually-related constructs ( $r>.3$ ), and showed medium to non-significant associations ( $r<.3$ ) with constructs expected to be conceptually different (Table 4). The main exceptions involved the discriminant validity of the exercise subscales which unexpectedly correlated with LBP perceptions (BIPQ): exercise concerns and exercise individual fit were moderately correlated with LBP concerns; and exercise concerns were moderately correlated with perceived emotional impact of LBP. Related to convergent validity, the observed correlations of 0.29 fell just short of the hypothesized 0.3 between attitudes to CAM (HCAMQ) and beliefs regarding effectiveness and individual fit of manual therapy, and



between general beliefs about the harmful effects of medications (BMQ) and LBP-specific medication concerns.

*Insert Table 4 Here*

### **Criterion Validity**

For each treatment, those participants who ranked a treatment as their top choice had more positive beliefs about that treatment than did other participants. This difference was significant for all treatments and all subscales, except beliefs about the Effectiveness of Pain Medication (Table 5). In other words, when participants had stronger beliefs about a treatment's effectiveness, credibility, and individual fit, and had fewer concerns about a treatment, they were more likely to prefer that treatment.

### **An Alternative Structure**

Subscale scores were highly correlated with each other within treatments (shown in Supplementary Digital Content 4, available online at <http://links.lww.com/PAIN/A82>). An exploratory MSA (aisp; results shown in Supplementary Digital Content 5, available online at <http://links.lww.com/PAIN/A83>) revealed that the entire item set could alternatively be considered a single 16-item scale with medium to good homogeneity. For this global scale, homogeneity scores were .46, .59, .60, .67 and Cronbach's  $\alpha$  values were .92, .95, .95, and .96 for pain medication, acupuncture, exercise and manual therapy respectively.

*Insert Table 5 Here*

### **Measurement Invariance**

Multi-group CFA analyses (summarised in Table 6) indicated that all scales showed scalar invariance between the two data collection waves and in most other subgroup comparisons, with some exceptions. Manual Therapy Beliefs displayed metric invariance between subgroups differentiated on probability of nerve root compression and self-reported sciatica diagnoses (results shown in Supplementary Digital Content 6, available online at <http://links.lww.com/PAIN/A84>). Exercise Beliefs displayed metric invariance between treatment experience subgroups (results shown in Supplementary Digital Content 7, available online at <http://links.lww.com/PAIN/A85>). No measurement invariance was observed for Medication Beliefs between patients with and without nerve compression (results shown in Supplementary Digital Content 8, available online at <http://links.lww.com/PAIN/A86>). No measurement invariance was observed for Acupuncture Beliefs in patients with or without sciatica and with or without treatment experience (results shown in Supplementary Digital Content 9, available online at <http://links.lww.com/PAIN/A87>). In addition, no measurement invariance was found between medication, exercise, manual therapy and acupuncture regarding the LBP-TBQ scales (results shown in Supplementary Digital Content 10, available online at <http://links.lww.com/PAIN/A88>).

*Insert Table 6 Here*

### **Short Version of the LBP-TBQ**

One item was selected to represent each subscale based on scores on homogeneity and Cronbach's  $\alpha$  if item excluded (the item with best properties for all treatments). The 4-item version of the questionnaire displayed good homogeneity ( $H=.43-.66$ ) (Table 7), internal consistency ( $\alpha=.70-.86$ ), and stability ( $r=.82-.85$ ) (Table 8). People that ranked a treatment as

their first choice reported significantly more positive treatment beliefs than people that ranked the treatment as a second, third or last option (Table 9).

*Insert Table 7 Here*

*Insert Table 8 Here*

*Insert Table 9 Here*

## **Discussion**

To facilitate the integration of treatment preferences in LBP clinical decision-making in primary care and to stimulate further research on this topic, we developed and validated a 16-item scale to measure core beliefs about four recommended LBP treatments (pain medication, exercise, manual therapy, acupuncture). The newly-developed LBP-TBQ showed good item properties, homogeneity, internal consistency, and stability. Discriminant and convergent validity were confirmed for most treatments, and the 4-factor structure was largely confirmed. Participants were more likely to rank a treatment as their first choice if they had stronger beliefs about its effectiveness, credibility, and individual fit, and fewer concerns about it, thus supporting the LBP-TBQ's criterion validity. A short 4-item LBP-TBQ was also developed with best-performing items and showed good psychometric properties. Both 16-item and 4-item versions (available in Supplementary Digital Content 11 available online at <http://links.lww.com/PAIN/A89>) can be used in future research and clinical practice to assess patients' beliefs about treatments.

The LBP-TBQ has several strengths compared to previous treatment belief questionnaires. First, our scale addresses several LBP beliefs in relatively compact 16-item or 4-item formats applicable to one or more treatments. This allows a comprehensive assessment with relatively low participation burden compared to existing scales, which assess

selected dimensions (e.g. overuse and harm in the 8-item BMQ [29], credibility and outcome expectancies in the 6-item CEQ [14]) and would need to be combined in a longer questionnaire without covering all relevant dimensions (e.g. individual fit). Second, previous scales are treatment specific. For example, deciding on medication use was previously described as involving a cost-benefit analysis comparing perceived benefits to concerns [27], while for other treatments perceived credibility and expected results have been considered more relevant [14]. In the LBP-TBQ, identical items assess beliefs about several treatments, facilitating direct comparisons. Third, the LBP-TBQ takes into account acquiescence bias by including negatively-worded items, which is rarely considered. For example, medication necessity and concerns have previously emerged as separate dimensions, but distinctions may be partly due to item wording [29]. Fourth, using NIRT methods allowed us to account for item difficulty and identify latent constructs under less strict (and more realistic) assumptions than parametric IRT [59]. And fifth, exploring measurement invariance enabled in-depth understanding of possible sources of variance in questionnaire structure and highlighted areas for improvement.

The LBP-TBQ benefits from combining theory and empirical qualitative data. Theoretically, it conceptualizes patients' preferences within the CSM as developing from patients' illness and treatment perceptions [26], and therefore assesses patients' beliefs about treatments to provide information relevant for clinical decision-making. According to the CSM (and confirmed in our qualitative research [16]), patients need to form adequate illness representations (e.g. illness identity and causal representations) to inform their treatment decisions; thus future studies should assess illness representations alongside treatment beliefs. By using empirical qualitative data, we were able to construct items based on the language LBP patients use to describe their beliefs.

Our validation study also revealed unanticipated and interesting differences between how patients perceive LBP treatments. First, contrary to our hypotheses, we identified significant moderate associations between patients' concerns and perceptions of individual fit regarding exercise and their concerns about, and emotional representations of, LBP. These associations suggest that, unlike other (arguably more passive) treatments, patients with fewer LBP concerns and less intense emotions about LBP have fewer concerns about exercise and stronger beliefs that exercise is right for them. Second, effectiveness beliefs showed the weakest stability, particularly for medication ( $r=.63-.76$ ), suggesting that effectiveness beliefs fluctuate more over time than others. It may be that these beliefs are more easily influenced (e.g. by personal or vicarious experience of medications or practitioners) than beliefs about credibility, concerns, and individual fit. Understanding the causes and mechanisms of belief variability may be a promising avenue for further research and may reveal appropriate approaches to influencing the development of treatment beliefs. Third, measurement invariance was not achieved between treatments, suggesting that patients may interpret items and constructs somewhat differently when assessing different treatments. This is expected given differences between treatments (e.g. one's concerns regarding medication can be very different to those about exercise) and suggests that future research using the LBP-TBQ for treatment comparisons should first identify the sources of these differences and establish partial invariance.

The measurement invariance findings prompt specific recommendations regarding future applications and development of the LBP-TBQ. Scalar invariance was achieved for all treatments between data collection waves and pain duration subgroups, supporting the use of the LBP-TBQ in longitudinal studies, or to compare groups of patients experiencing pain for less or more than 3 years. Not all treatments displayed scalar invariance between patients with different probability of nerve root compression, self-reported sciatica, or treatment

experience; patients in these subgroups might attribute different meanings to particular items or constructs. Therefore, the LBP-TBQ structure would benefit from further investigation in relation to different patient characteristics. We recommend researchers using the LBP-TBQ for subgroup comparisons first ascertain whether these subgroups interpret questionnaire items similarly, by performing measurement invariance analyses before examining group differences or using cognitive interviews. Moreover, further psychometric work is needed before using the LBP-TBQ to assess beliefs about other treatments (e.g. cognitive-behavioral therapy, multidisciplinary rehabilitation), or in other populations, languages or settings, and we recommend additional qualitative and psychometric examinations of its relevance and item properties in new contexts [5].

Our research is subject to several limitations. First, our participants were recruited from patients consulting with their primary care GP or CAM practitioner; results might be different in secondary care patients (e.g. those who are undergoing hospital-based treatments or being considered for spinal injections or surgery). However, some of our patients had experienced hospital-based interventions (e.g. 17% had previously attended a pain management clinic, see Supplementary Digital Content 1). Second, our participants were older and more reported unemployment (mostly retired) and chronic pain compared to other primary care LBP samples (e.g. mean age 44 years, 24.6% unemployed (including retirees), 11.4% with pain duration longer than 3 years in another UK-based primary care cohort [17] vs 55 years, 52.5% unemployed (including retirees), and 64.8% with pain for longer than 3 years in our cohort). This is likely due to differences in participant selection (we contacted all LBP patients, not only patients currently consulting their primary care physician) or respondent burden (longer survey in the present study). Thus, the properties of the LBP-TBQ would benefit from confirmation in other samples. Third, although the 4-factor CFA model showed acceptable fit, the subscales were also highly inter-correlated within each treatment

and could alternatively be merged into a single “treatment acceptability” dimension. These strong associations could be seen as contrasting with our previous qualitative findings where they emerged as distinct aspects considered when making treatment decisions [16] and may reflect the different socio-cognitive processes involved in questionnaire response versus group discussion. Further psychometric work should explore the value of a 4-factor versus 1-factor structure, as well as measurement invariance and validity in different populations. Fourth, our study was informed primarily by the NICE care pathway and thus focused on four frontline treatments in the UK; future studies and applications would need to consider local availability and clinical relevance when restricting or expanding this selection.

Our study shows that the LBP-TBQ can be used to assess the beliefs of LBP patients regarding four guideline-recommended treatment options (pain medication, exercise, manual therapy, acupuncture). According to the extended CSM of illness representations [33], such beliefs are likely to be key modifiable determinants of treatment uptake and adherence. Understanding these beliefs and behaviours in LBP could allow us to develop interventions to optimize the use of evidence-based recommended treatments, and thus improve treatment effectiveness and patient outcomes in LBP. Protocols could be developed for clinicians to use the LBP-TBQ to match patients to treatments consistent with their treatment beliefs [e.g. 46] or to identify and address patients’ negative beliefs about treatments. For example, patients could complete the 4-item LBP-TBQ before (e.g. electronically) or during consultations (e.g. as a clinical interview) regarding locally-available treatment options. The patients’ answers could highlight topics for clinicians to explore in the consultation, perhaps after receiving appropriate training: for example, if a patient reports concerns about the clinically-preferred treatment the clinician could discuss these openly before agreeing on a treatment course. While a less focused discussion of treatment options might be inconclusive, completing the LBP-TBQ could help patients crystallise their beliefs and thus prompt a

clinically-appropriate decision that reflects patients' beliefs where possible. Further research is needed to develop and test such clinical applications of the LBP-TBQ.

We recommend the LBP-TBQ for prospective research examining determinants of treatment uptake and adherence in patients with LBP. Because it can assess the same beliefs about different treatments, the LBP-TBQ can also be used to model treatment choices within a more complex decision space than is often considered, i.e. one in which patients are choosing one from among many treatment options. Further research could explore clinical applications of the LBP-TBQ (particularly its 4-item version) to involve patients in treatment decision-making and thus help to more systematically take patients' preferences into account when choosing treatments for LBP [49;50].

ACCEPTED



## Reference List

- [1] Airaksinen O, Brox J, Cedraschi C, Hildebrandt J, Klüber-Moffett J, Kovacs F, Mannion A, Reis S, Staal J, Ursin H, Zanoli G, Obot C. Chapter 4 European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15:s192-s300.
- [2] Arbuckle JL. Amos (Version 21.0) [Computer Program]. Chicago: IBM SPSS, 2012.
- [3] Ark LAvd. Mokken scale analysis in R. *Journal of Statistical Software* 2007;20:1-19.
- [4] Ark LAvd. New developments in Mokken scale analysis in R. *Journal of Statistical Software* 2012;48:1-27.
- [5] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–3191
- [6] Bishop FL, Dima A, Ngui J, Little P, Moss-Morris R, Foster NE, Lewith GT. Clinical guidelines for low back pain in primary care: a qualitative study with UK clinicians. 2015. In Preparation.
- [7] Bishop FL, Yardley L, Lewith GT. Treatment appraisals and beliefs predict adherence to complementary therapies: A prospective study using a dynamic extended self-regulation model. *Br J Health Psychol* 2008;13:701-718.
- [8] Bishop FL, Yardley LF, Prescott PF, Cooper CF, Little PF, Lewith GT. Psychological covariates of longitudinal changes in back-related disability in patients undergoing acupuncture. *Clin J Pain* 2015;31:254-264.

- [9] Bishop FL, Yardley L, Lewith G. Developing a measure of treatment beliefs: The complementary and alternative medicine beliefs inventory. *Complement Ther Med* 2005;13:144-149.
- [10] Broadbent E, Petrie KJ, Main J, Weinman J. The Brief Illness Perception Questionnaire. *J Psychosom Res* 2006;60:631-637.
- [11] Chou R, Qaseem A, Snow V, Casey D, Cross JT, Jr., Shekelle P, Owens DK, for the Clinical Efficacy Assessment Subcommittee of the American College of Physicians and the American College of Physicians/American Pain Society Low Back Pain Guidelines Panel\*. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478-491.
- [12] Cohen J. A power primer. *Psychol Bull* 1992;112:155-159.
- [13] Croft PR, Macfarlane GJ, Papageorgiou AC, Thomas E, Silman AJ. Outcome of low back pain in general practice: a prospective study. *BMJ* 1998;316:1356.
- [14] Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry* 2000;31:73-86.
- [15] Deyo RA, Weinstein JN. Low Back Pain. *N Engl J Med* 2001;344:363-370.
- [16] Dima A, Lewith GT, Little P, Moss-Morris R, Foster NE, Bishop FL. Identifying patients' beliefs about treatments for chronic low back pain in primary care: a focus group study. *Br J Gen Pract.* 2013;63:e490-8.

- [17] Foster NE, Bishop A, Thomas E, Main C, Horne R, Weinman J, Hay E. Illness perceptions of low back pain patients in primary care: What are they, do they change and are they associated with outcome? *Pain* 2008;136:177-187.
- [18] French DJ, France CR, Vigneau F, French JA, Evans RT. Fear of movement/(re)injury in chronic pain: A psychometric assessment of the original English version of the Tampa scale for kinesiophobia (TSK). *Pain* 2007;127:42-51.
- [19] George SZ, Robinson ME. Preference, expectation, and satisfaction in a clinical trial of behavioral interventions for acute and sub-acute low back pain. *J Pain* 2010;11:1074-1082.
- [20] Glattacker M, Heyduck K, Meffert C. Illness beliefs, treatment beliefs and information needs as starting points for patient information - Evaluation of an intervention for patients with chronic back pain. *Patient Educ Couns* 2012;86:378-389.
- [21] Haanstra TM, Hanson L, Evans R, Nes FA, Vet HCW, Cuijpers P, Ostelo RWJG. How do low back pain patients conceptualize their expectations regarding treatment? Content analysis of interviews. *Eur Spine J* 2013;22:1986-1995.
- [22] Hagger MS, Orbell S. A meta-analytic review of the common-sense model of illness representations. *Psychol Health* 2003;18:141-184.
- [23] Hemker BT, Sijtsma K, Molenaar IW. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Appl Psychol Meas* 1995;19:337-352.

- [24] Henderson CJ, Orbell S, Hagger MS. Illness schema activation and attentional bias to coping procedures. *Health Psychol* 2009;28:101-107.
- [25] Hoffmann T, Del Mar C, Strong J, Mai J. Patients' expectations of acute low back pain management: implications for evidence uptake. *BMC Fam Pract* 2013;14:7.
- [26] Horne R. Representations of medication and treatment: advances in theory and measurement. In: Petrie KJ, Weinman JA, editors. *Perceptions of Health and Illness*. Amsterdam: Harwood Academic Publishers, 1997. pp. 155-188.
- [27] Horne R, Weinman J. Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. *J Psychosom Res* 1999;47:555-567.
- [28] Horne R, Weinman J. Self-regulation and self-management in asthma: exploring the role of illness perceptions and treatment beliefs in explaining non-adherence to preventer medication. *Psychol Health* 2002;17:17-32.
- [29] Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. *Psychol Health* 1999;14:1-24.
- [30] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1-55.
- [31] Hyland ME, Lewith GT, Westoby C. Developing a measure of attitudes: the holistic complementary and alternative medicine questionnaire. *Complement Ther Med* 2003;11:33-38.

- [32] Jackson DL, Gillaspay JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychol Methods* 2009;14:6-23.
- [33] Leventhal HA, Brissette I, Leventhal EA. The common-sense model of self-regulation of health and illness. In: Cameron LD, Leventhal H, editors. *The self-regulation of health and illness behaviour*. London: Routledge, 2003. pp. 42-65.
- [34] Mao JJ, Xie SX, Bowman MA. Uncovering the expectancy effect: the validation of the acupuncture expectancy scale. *Altern Ther Health Med* 2010;16:22-27.
- [35] Markland D. The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Pers Individ Dif* 2007;42:851-858.
- [36] Marsh HW, Hau K-T, Wen Z. In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct Equ Modeling* 2004;11:320-341.
- [37] McCracken LM, Hoskins J, Eccleston C. Concerns about medication and medication use in chronic pain. *J Pain* 2006;7:726-734.
- [38] Meade AW, Bauer DJ. Power and precision in confirmatory factor analytic tests of measurement invariance. *Struct Equ Modeling* 2007;14:611-635.
- [39] Meijer RR, Beneke JJ. Analyzing psychopathology items: a case for nonparametric item response theory modeling. *Psychol Methods* 2004;9:354-368.

- [40] Molenaar IW, Sijtsma K. Internal consistency and reliability in Mokken's nonparametric item response model. *Tijdschrift voor Onderwijsresearch* 1984;9:257-268.
- [41] Moss-Morris R, Weinman J, Petrie KJ, Horne R, Cameron LD, Buick D. The revised illness perception questionnaire (IPQ-R). *Psychol Health* 2002;17:1-16.
- [42] Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, AlMazroa MA, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Abdulhak AB, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT-A, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, Leo DD, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne C, Dolk H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewoigbokhan SE, Farzadfar F, Feigin V, FELSON DT, Ferrari A, Ferri CP, Fèvre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG,

Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gonzalez-Medina D, Gosselin R, Grainger R, Grant B, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Laden F, Lalloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Levinson D, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Mabweijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Memish ZA, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mock C, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KMV, Nelson PK, Nelson RG, NEVITT MC, Newton CR, Nolte S, Norman P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 2012;380:2197-2223.

- [43] Myers SS, Phillips RS, Davis RB, Cherkin DC, Legedza A, Kaptchuk TJ, Hrbek A, Buring JE, Post D, Connelly MT, Eisenberg DM. Patient expectations as predictors of outcome in patients with acute low back pain. *J Gen Intern Med* 2008;23:148-153.

- [44] Nicklas LB, Dunbar M, Wild M. Adherence to pharmacological treatment of non-malignant chronic pain: The role of illness perceptions and medication beliefs. *Psychol Health* 2009;25:601-615.
- [45] Nunnally J, Bernstein I. *Psychometric Theory*. New York: McGraw Hill, 1994.
- [46] Patel S, Ngunjiri A, Sandhu H, Griffiths F, Thistlewaite J, Brown S, Friede T, Lord J, Tysall C, Woolvine M, Underwood M. Design and development of a decision support package for low back pain. *Arthritis Care Res* 2014;66:925-933.
- [47] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: Retrieved from <http://www.R-project.org>, 2013.
- [48] Roelofs J, Sluiter JK, Frings-Dresen MHW, Goossens MI, Thibault P, Boersma K, Vlaeyen JWS. Fear of movement and (re)injury in chronic musculoskeletal pain: Evidence for an invariant two-factor model of the Tampa Scale for Kinesiophobia across pain diagnoses and Dutch, Swedish, and Canadian samples. *Pain* 2007;131:181-190.
- [49] Savigny P, Kuntze S, Watson P, Underwood M, Ritchie G, Cotterell M, Hill D, Browne N, Buchanan E, Coffey P, Dixon P, Drummond C, Flanagan M, Greenough C, Griffiths M, Halliday-Bell J, Hettinga D, Vogel S, Walsh D. Low back pain. Early management of persistent non-specific low back pain., Vol. NICE clinical guideline 88. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners, 2009. p. [www.nice.org.uk/CG88](http://www.nice.org.uk/CG88).
- [50] Savigny P, Watson P, Underwood M, on behalf of the Guideline Development Group. Early management of persistent non-specific low back pain: summary of NICE guidance. *BMJ* 2009;338:b1805.



- [51] Schoot Rvd, Lugtig P, Hox J. A checklist for testing measurement invariance. *Eur J Dev Psychol* 2012;9:486-492.
- [52] Sherman KJ, Cherkin DC, Ichikawa L, Avins AL, Delaney K, Barlow WE, Khalsa PS, Deyo RA. Treatment expectations and preferences as predictors of outcome of acupuncture for chronic back pain. *Spine* 2010;35:1471-1477.
- [53] Sijtsma K. Methodology Review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Appl Psychol Meas* 1998;22:3-31.
- [54] Sijtsma K, Molenaar IW. Reliability of test scores in nonparametric item response theory. *Psychometrika* 1987;52:79-97.
- [55] Smeets RJEM, Beelen S, Goossens MEJB, Schouten EGW, Knottnerus JA, Vlaeyen JWS. Treatment expectancy and credibility are associated with the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *Clin J Pain* 2008;24:305-315.
- [56] Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, Chambers WA, Smith WC. The Chronic Pain Grade questionnaire: validation and reliability in postal research. *Pain* 1997;71:141-147.
- [57] Straat JH, van der Ark LA, Sijtsma K. Minimum sample size requirements for Mokken scale analysis. *educational and psychological measurement* 2014;74:809-822.
- [58] van Hartingsveld FM, Ostelo RWJG, Cuijpers PP, de Vos RP, Riphagen IIM, de Vet HCW. Treatment-related and patient-related expectations of patients with

- musculoskeletal disorders: a systematic review of published measurement tools. *Clin J Pain* 2010;26:470-488.
- [59] van Schuur WH. Mokken scale analysis: between the guttman scale and parametric item response theory. *Polit Anal* 2003;11:139-163.
- [60] Vlaeyen JWS, Kole-Snijders AMJ, Boeren RGB, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain* 1995;62:363-372.
- [61] Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain* 1992;50:133-149.
- [62] Waddell G, Newton M, Henderson I, Somerville D, Main C. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157-168.
- [63] Weinman J, Petrie KJ, Moss-Morris R, Horne R. The illness perception questionnaire: A new method for assessing the cognitive representation of illness. *Psychol Health* 1996;11:431-445.
- [64] Willis GB. *Cognitive Interviewing: A "How To" Guide*. Research Triangle Institute, 1999.
- [65] Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. *Couns Psychol* 2006;34:806-838.

## **Acknowledgements**

This paper presents independent research funded by the National Institute for Health Research (NIHR) School of Primary Care Research (Grant Reference Number 75) and an NIHR Research Professorship for NE Foster (NIHR-RP-011-015). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. FLB's post was funded by Arthritis Research UK (Career Development Fellowship 18099). The study sponsors and funders had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

We thank the participants for their time in completing the survey and for sharing their views, and the Primary Care Research Network and the clinicians who helped to recruit the participants. Cathy Salmon and Sara Oroz assisted in preparing questionnaire packs.

## **Conflicts of Interest**

AD declares consultancy work for the Baylor Black Sea Foundation Respiratory Effectiveness Group. All other authors declare they have no conflicts of interest.

## Figure Captions

Figure 1. CFA Model of the Final LBP-TBQ (Pain Medication items, wave 1, maximum likelihood estimation).

ACCEPTED

Table 1. Sample characteristics (N=429)

Characteristic	N (%) <sup>a</sup>
<u>Education</u>	
Secondary school or less	183 (42.7%)
Sixth form-college	107 (24.9%)
Undergraduate study	75 (17.5%)
Postgraduate study	34 (7.9%)
<u>Work status</u>	
Employed	179 (41.7%)
...at usual job	152 (35.4%)
...on light duty	16 (3.7%)
...paid leave or sick leave	8 (1.9%)
...unpaid leave	3 (0.7%)
Retired	134 (31.2%)
Unemployed	230 (21.3%)
...because of LBP	25 (5.8%)
...on disability	22 (5.1%)
...homemaker	27 (6.3%)
...student	9 (2.1%)
...for other reasons	13 (3.0%)
<u>Pain duration (3 categories)</u>	
Persistent LBP (6 wks – 12 months)	88 (20.5%)
Chronic/recurrent LBP (12 months – 3 years)	63 (14.7%)
Chronic/recurrent LBP (>3yrs)	278 (64.8%)
<u>Chronic Pain Grade</u>	
Grade I – low disability, low intensity	82 (19.1%)
Grade II – low disability, high intensity	90 (21%)
Grade III – high disability, moderately limiting	81 (18.9%)
Grade IV – high disability, severely limiting	147 (34.3%)
Reporting one or more comorbidities	282 (65.7%)
<u>Subgroups Examined for Measurement Invariance</u>	
Pain duration <3 years	151 (35.2%)
Self-reported Sciatica	192 (44.8%)
At least 1 symptom of nerve root compression	144 (33.6%)
Past Experience of Medication	208 (48.5)
Past Experience of Manual therapy	273 (63.6)
Past Experience of Exercise	187 (43.6)
Past Experience of Acupuncture	127 (29.6)

<sup>a</sup> Note: Percentages reported without including missing values

Table 2. LBP-TBQ Scale and Item Homogeneity for 4-Item Scales

Content	H (SE)			
	Pain medication	Exercise	Manual therapy	Acupuncture
Credibility	0.51 (0.04)	0.67 (0.03)	0.69 (0.03)	0.67 (0.04)
Taking/Having [...] for back pain makes a lot of sense	0.53 (0.04)	0.70 (0.03)	0.71 (0.03)	0.68 (0.04)
Generally, [...] is a believable therapy for back pain	0.53 (0.04)	0.67 (0.04)	0.72 (0.04)	0.71 (0.04)
I am sceptical about [...] as a treatment for back pain in general (r)	0.52 (0.05)	0.66 (0.04)	0.68 (0.04)	<b>0.62 (0.05)</b>
I do not understand how [...] could help people with back pain (r)	0.46 (0.05)	0.66 (0.04)	0.65 (0.04)	0.67 (0.04)
Effectiveness	0.57 (0.04)	0.63 (0.03)	0.74 (0.03)	0.72 (0.04)
[...] cannot help people with back pain (r)	0.58 (0.05)	0.61 (0.05)	0.72 (0.04)	0.75 (0.04)
[...] can work well for people with back pain	0.56 (0.05)	0.61 (0.05)	0.79 (0.03)	0.70 (0.05)
I think [...] is pretty useless for people with back pain (r)	0.60 (0.04)	0.65 (0.04)	0.76 (0.03)	0.72 (0.05)
[...] can make it easier for people to cope with back pain	0.54 (0.05)	0.66 (0.04)	0.70 (0.04)	0.73 (0.05)
Concerns	0.46 (0.04)	0.62 (0.03)	0.48 (0.03)	0.64 (0.04)
I worry that [...] could make my back worse (r)	0.35 (0.04)	0.61 (0.04)	0.60 (0.03)	0.62 (0.04)
I have concerns about taking/having [...] for my back pain (r)	0.51 (0.04)	0.68 (0.03)	0.59 (0.03)	0.69 (0.03)
I would feel at ease about taking/having [...] for my back pain	0.50 (0.04)	0.64 (0.04)	0.58 (0.03)	0.62 (0.04)
I feel that [...] would not harm me	0.48 (0.04)	0.55 (0.05)	0.52 (0.03)	0.60 (0.05)
Individual fit	0.69 (0.03)	0.76 (0.03)	0.85 (0.02)	0.71 (0.04)
I think [...] could suit me as a treatment for my back pain	0.68 (0.03)	0.74 (0.03)	0.84 (0.03)	<b>0.72 (0.04)</b>
For me, taking/having [...] would be a waste of time (r)	0.70 (0.04)	0.77 (0.03)	0.87 (0.02)	0.73 (0.04)
I am confident [...] would be a suitable treatment for my back pain	0.69 (0.03)	0.77 (0.03)	0.86 (0.02)	0.74 (0.04)
Given what I know about my back pain, I doubt [...] would be right for me (r)	0.68 (0.03)	0.75 (0.03)	0.83 (0.03)	<b>0.66 (0.05)</b>

Note: Higher scores represent better homogeneity. Italic font is used for items with violations of monotonicity, bold font represents significant violations (at default rest group minsize). Items excluded during these analyses: [...] can help people with back pain to get on with their lives; I think I would find it unpleasant to take/have [...] for my back pain (r); I am worried that I cannot afford to pay for [...] (r); I think [...] would not work for my back pain (r)

ACCEPTED

Table 3. LBP-TBQ Subscales Means (M), Standard Deviations (SD), Stability and Cronbach  $\alpha$

Scale/Item	Pain Medication			Exercise			Manual Therapy			Acupuncture		
	M(SD)	$\alpha^a$	Test-retest	M(SD)	$\alpha^a$	Test-retest	M(SD)	$\alpha^a$	Test-retest	M(SD)	$\alpha^a$	Test-retest
Credibility	3.60 (0.74)	.77	.72	3.81 (0.74)	.87	.71	3.62 (0.83)	.88	.79	3.17 (0.81)	.87	.78
Taking/Having [...] for back pain makes a lot of sense	3.75 (0.91)	.70		3.86 (0.88)	.81		3.67 (0.92)	.83		3.21 (0.90)	.82	
Generally, [...] is a believable therapy for back pain	3.53 (0.90)	.71		3.82 (0.85)	.83		3.70 (0.89)	.83		3.30 (0.87)	.81	
I am sceptical about [...] as a treatment for back pain in general (r)	3.35 (1.04)	.71		3.68 (0.98)	.85		3.50 (1.02)	.86		3.10 (1.04)	.85	
I do not understand how [...] could help people with back pain (r)	3.77 (0.97)	.75		3.87 (0.97)	.83		3.59 (1.02)	.87		3.08 (1.03)	.82	
Effectiveness	3.94 (0.71)	.81	.63	3.89 (0.72)	.85	.69	3.80 (0.75)	.90	.70	3.42 (0.68)	.89	.76
[...] cannot help people with back pain (r)	4.01 (0.93)	.74		3.99 (0.91)	.82		3.82 (0.88)	.89		3.50 (0.80)	.85	
[...] can work well for people with back pain	3.74 (0.90)	.77		3.80 (0.84)	.82		3.76 (0.83)	.86		3.36 (0.75)	.87	
I think [...] is pretty useless for people with back pain (r)	3.98 (0.96)	.72		3.99 (0.91)	.80		3.86 (0.87)	.86		3.43 (0.82)	.87	
[...] can make it easier for people to cope with back pain	4.03 (0.79)	.79		3.77 (0.82)	.79		3.77 (0.82)	.90		3.39 (0.76)	.86	
Concerns	3.48(0.79)	.73	.80	3.32 (0.91)	.85	.77	3.33 (0.94)	.89	.83	3.36 (0.81)	.85	.80
I worry that [...] could make my back worse (r)	3.86 (0.96)	.75		3.15 (1.16)	.81		3.15 (1.13)	.85		3.40 (0.95)	.81	



I have concerns about taking/having [...] for my back pain (r)	3.25 (1.21)	.62		3.22 (1.15)	.76		3.23 (1.17)	.85		3.20 (1.06)	.76	
I would feel at ease about taking/having [...] for my back pain	3.52 (1.02)	.64		3.48 (1.06)	.79		3.49 (1.04)	.85		3.26 (1.05)	.81	
I feel that [...] would not harm me	3.28 (0.99)	.66		3.44 (1.01)	.84		3.47 (0.98)	.88		3.58 (0.83)	.83	
Individual Fit	3.51 (0.85)	.87	.77	3.56 (0.91)	.91	.82	3.40 (0.98)	.94	.82	3.05 (0.85)	.89	.80
I think [...] could suit me as a treatment for my back pain	3.33 (1.01)	.83		3.51 (0.98)	.88		3.43 (0.99)	.93		3.07 (0.92)	.85	
For me, taking/having [...] would be a waste of time (r)	3.85 (0.96)	.84		3.77 (0.99)	.88		3.51 (1.05)	.92		3.17 (1.02)	.85	
I am confident [...] would be a suitable treatment for my back pain	3.34 (0.98)	.83		3.45 (1.02)	.87		3.30 (1.06)	.92		2.94 (0.95)	.84	
Given what I know about my back pain, I doubt [...] would be right for me (r)	3.52 (1.04)	.84		3.49 (1.11)	.88		3.37 (1.13)	.93		3.02 (1.03)	.88	

<sup>a</sup> Cronbach  $\alpha$  reported at scale level and item level (if item excluded); higher scores represent better internal consistency.

Table 4. Convergent and Discriminant Validity for each Treatment Version of the LBP-TBQ Subscales

Subscale	Treatment Version							
	Pain Medication		Exercise		Manual Therapy		Acupuncture	
	Hypothesis	Pearson's r	Hypothesis	Pearson's r	Hypothesis	Pearson's r	Hypothesis	Pearson's r
Credibility	↑ CEQ	.44**	↑ CEQ	.66**	↑ CEQ	.67**	↑ CEQ	.68**
	Credibility = BIPQ	-.05	Credibility = BIPQ	-.29**	Credibility = BIPQ	-.22**	Credibility = BIPQ	.02
	Concerns = BIPQ	-.06	Concerns = BIPQ	-.24**	Concerns = BIPQ	-.18**	Concerns = BIPQ	-.02
	Emotions		Emotions		Emotions		Emotions	
					↓ HCAMQ	-.32**	↓ HCAMQ	-.38**
Effectiveness	↑ CEQ	.48**	↑ CEQ	.59**	↑ CEQ	.53**	↑ CEQ	.50**
	Expectancy = BIPQ	-.15**	Expectancy = BIPQ	-.27**	Expectancy = BIPQ	-.24**	Expectancy = BIPQ	-.03
	Concerns = BIPQ	-.13**	Concerns = BIPQ	-.22**	Concerns = BIPQ	-.19**	Concerns = BIPQ	-.05
	Emotions		Emotions		Emotions		Emotions	
					↓ HCAMQ	-.29**	↓ HCAMQ	-.36**
(Fewer) Concerns	↓ BMQ	-.29**	↓ TSK	-.60**	↓ HCAMQ	-.32**	↓ HCAMQ	-.35**
	Harm		Activity Avoidance = BIPQ	-.39**	Attitudes = BIPQ	-.24**	Attitudes = BIPQ	-.06
	↓ BMQ	-.39**	Concerns = BIPQ	-.38**	Concerns = BIPQ	-.27**	Concerns = BIPQ	-.10*
	Overuse = BIPQ	-.10*						

	Concerns = BIPQ Emotions	-.11*	Emotions		Emotions		Emotions	
Individual Fit	= BIPQ Concerns	-.04	= BIPQ Concerns	-.32**	↓ HCAMQ Attitudes	-.29**	↓ HCAMQ Attitudes	-.31**
	= BIPQ Emotions	-.06	= BIPQ Emotions	-.27**	= BIPQ Concerns	-.23**	= BIPQ Concerns	-.04
					= BIPQ Emotions	-.23**	= BIPQ Emotions	-.05

Note. ↑ at least moderate positive correlation hypothesised (convergent validity;  $r \geq 0.3$ ). ↓ at least moderate negative correlation hypothesised (convergent validity;  $r \geq 0.3$ ). = at most moderate correlation hypothesised (divergent validity;  $r < 0.3$ ).

\*  $p < .05$  \*\*  $p < .01$ .

Abbreviations: BIPQ = Brief Illness Perceptions Questionnaire; BMQ = Beliefs about Medicine Questionnaire; CEQ = Credibility Expectancy Questionnaire; HCAMQ = Holistic Complementary and Alternative Medicine Questionnaire; TSK = Tampa Scale of Kinesiophobia.

ACCEPTED

Table 5. Criterion Validity of LBP-TBQ Subscales

Scale	Group 1: Treatment Ranked 1			Group 2: Treatment Ranked 2-4			Between-Group Comparison		
	M	SD	n	M	SD	n	t	df	p
<b>Pain Medication</b>									
Credibility	3.88	.60	152	3.47	.79	189	-5.44 <sup>a</sup>	337.8	<.001
Effectiveness	4.05	.70	152	3.94	.72	189	-1.35	339	.177
(Fewer) Concerns	3.74	.68	152	3.30	.82	189	-5.29	339	<.001
Individual Fit	3.82	.70	152	3.32	.89	189	-5.87 <sup>a</sup>	338.8	<.001
<b>Exercise</b>									
Credibility	4.24	.56	88	3.75	.81	239	-6.22 <sup>a</sup>	223.3	<.001
Effectiveness	4.27	.58	88	3.84	.71	239	-6.05	325	<.001
(Fewer) Concerns	3.73	.82	88	3.19	.90	239	-4.94	325	<.001
Individual Fit	4.07	.62	88	3.46	.94	239	-6.84 <sup>a</sup>	233.1	<.001
<b>Manual Therapy</b>									
Credibility	4.14	.72	89	3.44	.78	240	-7.34	327	<.001
Effectiveness	4.24	.63	89	3.68	.72	240	-6.50	327	<.001
(Fewer) Concerns	3.98	.79	89	3.09	.89	240	-8.34	327	<.001
Individual Fit	4.14	.73	89	3.17	.94	240	-9.80 <sup>a</sup>	202.0	<.001
<b>Acupuncture</b>									
Credibility	3.99	.71	24	3.12	.81	299	-5.15	321	<.001
Effectiveness	3.94	.80	24	3.41	.65	299	-3.78	321	<.001
(Fewer) Concerns	3.878	.88	24	3.34	.80	299	-3.11	321	.002
Individual Fit	3.71	1.10	24	2.98	.83	299	-4.02	321	<.001

<sup>a</sup> Equal variances not assumed

Table 6. Measurement invariance of the LBP-TBQ for each treatment

Treatment	Nerve compression (present or absent)	Sciatica (present or absent)	Pain duration (<3yrs or $\geq 3$ yrs)	Treatment experience (present or absent)	Time (1 <sup>st</sup> or 2 <sup>nd</sup> survey)
Medication	None	Scalar	Scalar	NA <sup>a</sup>	Scalar
Exercise	Scalar	Scalar	Scalar	Metric	Scalar
Manual therapy	Metric	Metric	Scalar	NA <sup>a</sup>	Scalar
Acupuncture	Scalar	None	Scalar	None	Scalar

<sup>a</sup> NA = data not available due to low numbers of treatment naive patients.

ACCEPTED

Table 7. 4-item LBP-TBQ Scale and Item Homogeneity

Content	H (SE)			
	Pain medication	Exercise	Manual therapy	Acupuncture
Short form	0.43 (0.04)	0.58 (0.04)	0.66 (0.03)	0.61 (0.03)
Taking/Having [...] for back pain makes a lot of sense	0.48 (0.04)	0.62 (0.04)	0.69 (0.03)	0.64 (0.04)
I think [...] is pretty useless for people with back pain (r)	0.39 (0.04)	0.53 (0.05)	0.64 (0.04)	0.61 (0.04)
I have concerns about taking/having [...] for my back pain (r)	0.37 (0.04)	0.56 (0.04)	0.63 (0.03)	0.56 (0.04)
I am confident [...] would be a suitable treatment for my back pain	0.49 (0.04)	0.62 (0.04)	0.68 (0.03)	<b>0.64 (0.03)</b>

ACCEPTED

Table 8. 4-item LBP-TBQ: Means (M), Standard Deviations (SD), Stability and Cronbach  $\alpha$

Scale/Item	Pain Medication			Exercise			Manual Therapy			Acupuncture		
	M(SD)	$\alpha^\dagger$	Test-retest	M(SD)	$\alpha^\dagger$	Test-retest	M(SD)	$\alpha^\dagger$	Test-retest	M(SD)	$\alpha^\dagger$	Test-retest
Short form	3.58 (0.75)	.70	.82	3.63 (0.79)	.81	.84	3.51 (0.84)	.86	.85	3.19 (0.76)	.83	.82
Taking/Having [...] for back pain makes a lot of sense	3.75 (0.91)	.60		3.86 (0.88)	.74		3.67 (0.92)	.80		3.21 (0.90)	.76	
I think [...] is pretty useless for people with back pain (r)	3.98 (0.96)	.67		3.99 (0.91)	.79		3.86 (0.87)	.83		3.43 (0.82)	.78	
I have concerns about taking/having [...] for my back pain (r)	3.25 (1.21)	.71		3.22 (1.15)	.77		3.23 (1.17)	.83		3.20 (1.06)	.81	
I am confident [...] would be a suitable treatment for my back pain	3.34 (0.98)	.59		3.45 (1.02)	.73		3.30 (1.06)	.80		2.94 (0.95)	.76	

ACCEPTED

Table 9. 4-item LBP-TBQ Criterion Validity

	Group 1: Treatment Ranked 1			Group 2: Treatment Ranked 2-4			Between-Group Comparison		
	M	SD	n	M	SD	n	t	df	p
Medication	3.86	.62	152	3.41	.78	189	-5.78	339	<.001
Exercise	4.07	.58	88	3.55	.81	239	-5.58	325	<.001
Manual therapy	4.14	.71	89	3.30	.78	240	-9.25 <sup>a</sup>	171.02	<.001
Acupuncture	3.83	.93	24	3.14	.73	299	-4.33	321	<.001

<sup>a</sup> Equal variances not assumed

ACCEPTED



