Barons, MJ; Griffiths, FE; Parsons, N; Alba, A; Thorogood, M; Medley, GF; Lamb, SE (2014) Matching patients to an intervention for back pain: classifying patients using a latent class approach. Journal of evaluation in clinical practice, 20 (4). pp. 544-50. ISSN 1356-1294 DOI: https://doi.org/10.1111/jep.12115

Downloaded from: http://researchonline.lshtm.ac.uk/2814511/

DOI: 10.1111/jep.12115

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
Matching patients to an intervention for back pain: classifying patients using a latent class approach

Martine J. Barons PhD,1 Frances E. Griffiths MRCGP PhD,2 Nick Parsons PhD,3 Anca Alba PhD,4* Margaret Thorogood PhD,5 Graham F. Medley PhD6 and Sarah E. Lamb PhD7,8

1Research Fellow, Complexity Science Centre, University of Warwick, Coventry, UK
2Professor of Medicine in Society, 3Statistician, 4Research Fellow, 5Professor of Epidemiology, Warwick Medical School, University of Warwick, Coventry, UK
6Professor of Infectious Disease Epidemiology, School of Life Sciences, University of Warwick, Coventry, UK
7Co-Director, Oxford Clinical Trials Research Unit, University of Oxford, Oxford, UK
8Professor of Rehabilitation, Clinical Trials Unit, University of Warwick, Coventry, UK

Keywords
medical informatics, patient-centred care, person-centred medicine

Correspondence
Dr Martine Jayne Barons
Centre for Complexity Science
University of Warwick
Coventry CV4 7AL
UK
E-mail: martine.barons@warwick.ac.uk

*Now Clinical Psychiatry, Institute of Psychiatry, King’s College, London.

Accepted for publication: 22 January 2014
doi:10.1111/jep.12115

Introduction
Almost everyone experiences non-specific low back pain sometime during their lifetime [1]. Each year in the UK, about one-third of the population experiences back pain and of this 20% consult their GPs about it [2]. There are a number of interventions known to be effective for non-specific low back pain including exercise programmes, manual therapy, acupuncture [2] and a cognitive behavioural approach [3]. However, these interventions have small effects when averaging cross the full population. This has prompted researchers to try to identify ways of classifying those seeking treatment so that patients can be matched to interventions in order to maximize treatment effects [4]. Many different classifications have been developed [5] but there is little consensus on their use [6]. For non-specific low back pain of greater than 6 weeks duration, the UK National Institute of Health and Clinical Excellence has suggested that patient preference should guide the choice of treatment from a range of effective interventions [2]. This study aimed to investigate LCA to classify patients and consider whether different groups responded differently to a cognitive
behavioural intervention. The results could provide additional
guidance for patients and clinicians as whether or not to consider
the use of a cognitive behavioural approach [3].

We undertook secondary analysis of data from a multi-centred
randomized controlled trial (RCT) of a primary care-based cogni-
tive behavioural program for low back pain (the Back Skills Train-
ing Trial – ‘BeST’) [3]. The cognitive behavioural approach [3]
targeted modifiable health behaviours and beliefs, for which evi-
dence exists that they contribute to low back pain becoming chronic
and disabling. The intervention targeted: activity levels, catastro-
phizing, fear avoidance, and coping skills. It comprised an assess-
ment followed by six group sessions, which tackled these
behaviours and beliefs, but did not target social factors such as
educational level, work-related risk factors and job satisfaction.
The control group received the internationally accepted best practice
recommended for primary care which is to promote physical activity,
prescribe analgesia and encourage a positive outlook. The trial
found a positive effect of the intervention [3]. Pre-specified sub-
group analysis of trial data found an interaction between baseline
fear avoidance [7] and outcome (reduced disability) [8] with a larger
treatment effect for those who were not fear avoidant at baseline,
those whose back pain was at least very troublesome and those with
a longer duration of back pain. This finding could be explained by
the lack of effect on this group of patients of the treatment received
in the control arm group of the trial [3]. In post hoc analysis by the
trial team, no reliable evidence was found of a moderating effect of
treatment outcome by baseline variables, although there was a
suggestion that being younger and currently working moderated the
treatment effect. In that post hoc analysis, back pain troublesome-
ess and fear avoidance had no moderating effects [9].

For our analyses, we used a latent class model to explore whether
measures of a combination of modifiable beliefs and behaviours,
amenable to clinical assessment and targeted by the cognitive
behavioural approach, can be used to classify individuals with low
back pain into different classes. LCA is a statistical technique that
uses multivariate categorical data, to identify classes of similar
individuals, the characteristics of the group and the probability of
the classification of the intervention. LCA is a probabilistic model because each class
it identifies is characterized by a pattern of conditional probabilities
that indicate the likelihood that variables take on certain values. The
aim of latent class modelling is to obtain the number and definition
of groups that best account for the dependencies between observed
variables. We then assessed the impact on back pain-related dis-
ability, of class membership, and of the interaction between class
membership and the intervention. There are many other widely used
methods of classification, such as decision trees, random forests and
support vector machines, which could be used for classification.
The results of using artificial neural networks are reported else-
where [10]. However, the purpose of this paper is not to conduct a
comprehensive evaluation of analysis methods, but to try a proba-
bilistic method that retains the clear interpretation of relationship
between measured patient characteristics and prediction so vital in
medical practice.

For comparison, the latent class approach to classifying patients
was compared with the Subgroups for Targeted Treatment (STarT)
back screening tool [11] which has been developed more recently.
The STarT tool comprises a set of nine questions based on referred
leg pain, co-morbid pain, disability, bothersomeness, catastro-
phizing, fear, anxiety and depression [12]. This tool classifies
patients as low, medium or high risk of persistent disability, and
treatment is tailored to each group: all patients were given advice
and were shown a 15-minute educational video entitled Get Back
Active20 and given the Back Book. Low-risk patients were only
given this clinic session, medium-risk patients were referred for
standardized physiotherapy to address symptoms and function, and
high-risk patients were referred for psychologically informed
physiotherapy to address physical symptoms and function, and also
psychosocial obstacles to recovery. Our aim is to contribute to the
development of evidence-based guidance for patients and clinicians
when choosing between interventions known to be effective for low
back pain.

Methods

This secondary analysis was undertaken independently of the team
that undertook the clinical trial. SL was the chief investigator of the
trial and was involved in initial discussions about undertaking the
analysis reported here and discussed the implications of the find-
ings. During the trial, FG supervised the nested qualitative study
only.

Clinical trial setting and participants

Between April 2005 and April 2007, 701 adults with at least
moderately troublesome low back pain of a minimum of 6 weeks
duration [13] were recruited to the trial from 56 general practices
in seven localities across England. Exclusion criteria were: aged
<18 years, having factors associated with severe pathology, severe
psychiatric or psychological disorders, and those who had had
been managed previously in a cognitive behaviour programme [3].
Baseline measures were collected pre-randomization face to face.
Outcome measures were collected by postal questionnaire.
Random allocation was 2:1 for treatment: control arm of the trial.

Data used in secondary analysis

We limited our analysis to the 407 participants of the 701 in the
trial with complete data at baseline for the variables used for LCA
and at 12 months for measuring outcome. We included those
variables that were potentially modifiable by the intervention, and
that could be assessed by a clinician in daily practice. These had
been included in the trial data collection to provide indications as
to why the treatment had or had not been effective [3]. The vari-
ables we chose included the fear avoidance beliefs questionnaire
(items 2–5) [7] and the pain self-efficacy score [14] as they assess
factors directly addressed by the intervention. We included the
Hospital Anxiety and Depression Scores (HADS) [15] because
challenging unhelpful thoughts, pacing, goal setting and relaxation
are approaches used for the treatment of depression and anxiety
[16], and back pain can become worse in the presence of depres-
sion [17]. We included question 7 of the SF12 [18] ‘During the
past 4 weeks, how much of the time has your physical health or
emotional problems interfered with your social activities (like
visiting with friends, relatives, etc.)?’ to provide a single measure of
the social impact of back pain as the intervention encouraged
increased activity, and also included troublesomeness of back pain
(moderately; very; extremely) as the intervention taught coping
skills which might reduce troublesomeness. Scores were calculated according to questionnaire manuals.

HADS were split according to their standard interpretation; scores less than 7 were classed as not depressed or anxious, 7–10 borderline, and greater than 10 depressed or anxious [15]. The trial interpreted fear avoidance as follows: less than 14 not fear avoidant and 14 or over was fear avoidant, so we adopted these categories. No standard categorization exists for the pain self-efficacy score; we categorized into the following groups (0–20, 21–30 and 31–60) to differentiate between very low self-efficacy and low self-efficacy which could be clinically important, but disregarded the difference between high and very high self-efficacy.

We used the categorical outcome of a three-point change on the Roland–Morris Disability Questionnaire (RMQ) [19,20], and calculated the change between baseline and 12 months (a decrease in score indicates decreased disability). Participants who scored 3 or less on the RMQ at baseline and also at 12 months were included in the improved group. There were no patients with a score of 22 or above (maximum disability score 24) at baseline or 12 months. Our adjusted models used age, gender and employment status.

**Statistical analysis**

Polytomous variable LCA was implemented in R statistical software package poLCA [21] using both expectation-maximization and Newton–Raphson algorithms to fit finite mixture models. Two 3, 4, 5 and 6 class models were fitted separately and compared using Akaiake Information Criterion (AIC) [22] and the Bayesian Information Criterion (BIC) [23].

Data from participants in both intervention and control arms of the trial were included in the analysis. Subjects were allocated to classes based on the maximum posterior probabilities of belonging to each class. Fisher’s exact test was used to assess any association between class membership and outcome for the intervention group and for the control group. Logistic regression was used to explore whether class membership predicted improvement in back pain disability as defined.

The predictive ability of the logistic regression using only class membership to predict outcome was tested using 10-fold cross-validation [24]. This was compared with a simple logistic regression using all the variables used to construct the classes as predictors which was also tested using 10-fold cross-validation.

The patients in the BeST trial did not complete the STarT questionnaire, but there were sufficient data collected for us to construct STarT scores retrospectively. The BeST trial exclusion criteria would have eliminated anyone who responded positively to any of the first two questions on STarT. We selected available data to complete the remaining items in the STarT questionnaire as follows: STarT item 3 and RMQ question 17, STarT item 4 and RMQ question 9, Start item 5 and fear avoidance STarT item 6 and HADS question 5, STarT item 7 and pain self-efficacy, STarT item 8 and HADS question 2, STarT item 9 and troublesomeness. We calculated the STarT scores for each BeST patient according to the STarT tool instructions, and so obtained a second categorization of the BeST patients into three groups for comparison with the LCA classification. We tested the STarT group allocation for association with outcome using logistic regression and analysis of deviance as for the classes derived from LCA.

### Results

#### Missing data

Of the 701 participants in the trial, 407 provided complete data for the variables used for determining class membership and outcome. Table 1 provides details of the missing. There were no missing values for age and gender but two cases of the 407 did not have data on employment status and were omitted when fitting the adjusted model.

Of the 407 cases, 281 (69%) were in the intervention arm (60.0% of total in intervention arm of trial) and 126 (31%) in the control arm of the trial (54.1% of total in control arm of trial), which is in line with the 2:1 allocation to treatment.

#### Latent class models and association with outcome for intervention and control patients

Models were fitted with 2, 3, 4, 5 and 6 classes. The model with the lowest BIC was the 3 class model and the model with the lowest AIC was the 4 class model (see Table 2).

There was an association between outcome and class for the 3 class model (Fisher’s exact test $P = 0.05$) for those receiving the intervention, but not for the 2, 4, 5 or 6 class model. There was no association between outcome and class for those in the control arm of the trial for any model. Further analysis focused on the three-class model.

#### Characteristics of classes in three-class model

Table 3 presents the characteristics of the members of each of the three classes and for all cases included in the LCA.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Complete cases (%)</th>
<th>Missing (%)</th>
<th>Accumulated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fear avoidance</td>
<td>662 (94.4)</td>
<td>5 (0.7)</td>
<td>34 (4.9)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>686 (97.9)</td>
<td>2 (0.3)</td>
<td>13 (1.8)</td>
</tr>
<tr>
<td>Depression</td>
<td>693 (98.9)</td>
<td>2 (0.3)</td>
<td>6 (0.8)</td>
</tr>
<tr>
<td>SF12 question 7*</td>
<td>689 (98.3)</td>
<td>12 (1.7)</td>
<td>–</td>
</tr>
<tr>
<td>Pain self-efficacy</td>
<td>676 (96.4)</td>
<td>3 (0.4)</td>
<td>22 (3.1)</td>
</tr>
<tr>
<td>Troublesomeness of back pain*</td>
<td>638 (91.0)</td>
<td>63 (9.0)</td>
<td>–</td>
</tr>
<tr>
<td>RMQ at baseline</td>
<td>700 (99.9)</td>
<td>1 (0.1)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>RMQ 12 months</td>
<td>498 (71.0)</td>
<td>203 (29.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

*Single-item score.

Table 2: BIC and AIC for latent class models with 2–6 classes

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>4336.3</td>
<td>4304.6</td>
<td>4357.0</td>
<td>4426.3</td>
<td>4503.2</td>
</tr>
<tr>
<td>AIC</td>
<td>4211.9</td>
<td>4116.2</td>
<td>4104.5</td>
<td>4109.5</td>
<td>4122.3</td>
</tr>
</tbody>
</table>

AIC, Akaiake Information Criterion; BIC, Bayesian Information Criterion.
Most members of class I found their back pain did not interfere with social activities, they were not anxious or depressed, their back pain was moderately troublesome, they were not fear avoidant, and they had moderate or higher self-efficacy.

Most members of class II found their back pain interfered with social activities, they were anxious and depressed, their back pain was very or extremely troublesome, they were fear avoidant, and they had low or very low self-efficacy.

Most members of class III found their back pain interfered with social activities some of the time, they were borderline anxious and borderline depressed, their back pain was moderately or very troublesome, they were fear avoidant, and they had moderate or higher self-efficacy.

Table 4 shows the percentage of each class in each arm of the trial and the number and percentage of patients in each arm of the trial who did and did not improve, divided into the three classes. Of those receiving the intervention, in class III a higher proportion improved than did not improve. In class I a higher proportion improved but the difference was not as marked as in class III, and in class II a higher proportion did not improve than improved.

Logistic regression adjusted for age, gender and employment [9] indicated a predictive effect of improving with treatment for those in work (see Table 5).
The error rate for the logistic regression using only class to predict outcome was 39.8% (sensitivity 0.59, specificity 0.61), and for the ordinary logistic regression the error rate was 41.5% (sensitivity 0.54, specificity 0.61).

We tested the STarT group membership for association with outcome in an identical fashion as for the classes derived from LCA. Both sets of results were remarkably similar. The medium-risk group was the smallest group, containing 20% of the patients. There were no differences between STarT groups in respect of the spread patient ages or gender. The low-risk group contained twice as many working patients as not working patients and this proportion was reversed for the high-risk group. In the medium-risk group, the numbers were even.

Logistic regression showed a strong association between treatment and outcome, but no association between membership of any of the STarT groups and outcome. The analysis of deviance showed a strong association between treatment and outcome and STarT group and outcome, but no significant interactions. A chi-square test between STarT group and outcome showed a strong association for the intervention patients but no association for control patients. Unlike the LCA, using the STarT groups did not pick out work status as significant; work is known to be an effect modifier [9].

**Discussion**

We were able to classify the trial participants using a probabilistic model based on psychosocial baseline scores relevant to the intervention. Of the parsimonious models tested for association between class membership and outcome, an association was only found with one model which had three classes. For the trial participants who received the intervention, there was an association between class membership and outcome, but not for those who did not receive the intervention. However, we were unable to detect an effect on outcome from interaction between class membership and receiving the intervention or not.

The STarT tool, a popular, alternative, tree-based method, was employed for comparison and was found to perform equally well in most respects, but not better on this cohort of patients. In fact, our analysis using the STarT groups did not identify work status as an effect modifier, whereas the same analysis using the latent classes did. There remains a need for decision support for clinicians in allocating individual patients with non-specific low back pain to the most appropriate of the array of interventions that have been shown to be effective when outcome is averaged over all patients.

**Strengths and limitations of the study**

By limiting our choice of data for analysis to that with direct relevance to the intervention mechanisms, we reduced the danger of identifying classes with no relevance to the question of deciding who might benefit from the intervention [4]. LCA can only be used with a complete data set. For linear models, there are methods for

<p>| Table 4 Participants who experienced improvement in back pain in each class of the three-class model |
|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Received intervention</th>
<th>Number that improved (%)</th>
<th>% of class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excluded</td>
<td>167 (89.3)</td>
<td>20 (10.7)</td>
</tr>
<tr>
<td>Class I</td>
<td>73 (52.5)</td>
<td>66 (47.5)</td>
</tr>
<tr>
<td>Class II</td>
<td>25 (61.0)</td>
<td>16 (39.0)</td>
</tr>
<tr>
<td>Class III</td>
<td>41 (40.6)</td>
<td>60 (59.4)</td>
</tr>
<tr>
<td>Total</td>
<td>139 (49.4)</td>
<td>142 (50.5)</td>
</tr>
</tbody>
</table>

| Table 5 Logistic regression models to predict recovery from back pain: unadjusted and adjusted models |
|-----------------|-----------------|-----------------|
| Variable        | Class and treatment model | Adjusted model† |
|                 | Coefficient     | Standard error | P-value | Coefficient     | Standard error | P-value |
| Intercept       | −1.39           | 0.79           | 0.08#   | −1.46           | 0.98           | 0.14    |
| Cognitive behaviour approach (treatment) | 0.94 | 0.85 | 0.27 | 0.78 | 1.09 | 0.47 |
| Class I         | 0.65            | 0.82           | 0.43    | 1.08            | 0.91           | 0.23    |
| Class II        | 0.58            | 0.86           | 0.49    | 0.84            | 0.89           | 0.35    |
| Class III       | −0.31           | 0.90           | 0.74    | −0.82           | 0.99           | 0.41    |
| Treatment * Class I | 0.0009 | 0.98 | 0.99 |
| Treatment * Class II | 0.24 | 0.93 | 0.79 |
| Treatment * Class III | 1.48 | 0.62 | 0.02* |

†Adjusted for age, gender and employment status.
Significance codes: # significant at α = 0.1, *significant at α = 0.05.
multiple imputation of missing values and pooling of model coefficients based on robust theory and which can quantify the uncertainty in the model coefficients, but these are not readily available for LCA. A clinical trials data set is relatively small for use with subgrouping [25]. A further limitation is that sensitivity to change is known to vary across the RMQ scale with better sensitivity in the middle range compared with the low and high ranges [26,27]. Our analysis confirms the well-known association between employment status and back pain outcome [28].

Classifying back pain to improve treatment outcome

A systematic review of the role of classification in low back pain published in 2011 [29] identified three types of classification: to classify by diagnosis, to describe prognosis, and considering treatment response, as in our study. The review identified 28 classification systems of which five were in the treatment-based category where treatment was tailored to the classification of the individual patient. In all five, classification was done by the clinician and based wholly or in part on observations about pain location and change with movement. With movement, two had been subject to a clinical trial and for only one of these was there some evidence of effectiveness (rated as insufficient).

Our study differs by including variables based on patient self-report and capturing a range of variables that influence the experience of back pain, rather than focusing only on the pain. In the RCT, high-risk patients received support for psychosocial barriers to recovery in addition to the physiotherapy received by the medium-risk group. All groups in the intervention arm received advice. There was a larger improvement in RMQ scores in the intervention group at 4 and 12 months, and the cost was lower [30]. The STarT tool was evaluated with an RCT involving 329 patients and a cohort study of 410. Our study is similar to the STarT study although without the rigour of being an RCT.

Unlike the pre-specified subgroup analysis and the post hoc analysis on the same data set as our study [3,9], our LCA suggests there are classes of patients which benefit most from the intervention. This suggests a benefit from the use of LCA that is able to capture the non-linear relationships between variables. We have shown that the error rate using the classes as predictors of treatment response in a logistic regression is lower than using the variables on which the classes were built in a logistic regression, showing that classes add value to predictions of outcome. Moreover, while the specificity was the same for both regression models, the sensitivity was superior in the case of the model using the classes as predictors. Although we have shown a clear association between class and outcome for those who received the intervention, we have not been able to show a particular class that had had a statistically significant difference to treatment response. During the data cleaning, the regression was run on several versions of the data which varied by a few cases showing the optimal classes, class assignment, and association between class membership and outcome in the intervention group were robust. However, change in a few cases made a difference as to whether outcome was influenced by interaction between class membership and intervention. This suggests that the sample size, after excluding the missing data, was not large enough to be robust to individual patient effects, reducing statistical power of the study; there might be an effect that is too small to detect. Conversely, if the size of the effect of class membership exists but is small, then it is unlikely to have an important clinical use.

Future use of LCA and the classification of back pain

A larger data set would enable greater use of methods such as LCA for classification of back pain such as the repository of individual data from back pain clinical trials [9]. Confidence in the finding that class membership influences the likely effect of the intervention would be increased if it was possible to detect the effect of interaction between class membership and treatment. Missing data remains a problem.

Combining variables relevant to back pain treatment for identifying classes results in a new qualitative assessment of the patient-class membership. This is somewhat similar to the patient-centred approach used by clinicians, which takes account of the whole person including their context and relationships [31]. However, our approach is restricted to considering factors which the intervention was designed to tackle. There is evidence that a clinician’s assessment of a patient is influenced by the potential management routes available [32]. In contrast to this multidimensional classification used at baseline in our study, the outcome assessment consisted of a single measure of back pain disability. There is also evidence that patients can change qualitatively even if their scores on an outcome measure do not change [3,33]. If a range of data had been collected at follow-up, the outcome assessment could also be qualitative and multidimensional.

Using LCA increases the complexity of assessments, as non-linear relationships between variables are captured. Using the same classification as at baseline, it might be possible to identify individuals that change class over time. This idea is supported by a small longitudinal interview study that included people with back pain, recruited in a similar way to the study reported here [34]. The categories they identified captured the emergent dynamic of individuals in relation to their back pain – emerging from the interaction of many different, not necessarily all acknowledged, factors [35]. Categorizing using a small range of variables, as here, gives a simpler qualitative assessment in which all components are measurable.

Classification post hoc to an intervention trial to seek categorizations that influenced outcome of intervention is an extremely attractive technique, but will always require RCTs to confirm that the associations found are genuine causal relationships rather than spurious, random associations. Nonetheless, the application of the same techniques to multiple data sets (i.e. multiple trials) would impart greater confidence that such associations are genuine. The main outcome from such techniques at the moment is to encourage consideration of variables (and meta-variables) that cannot be directly or easily be measured, and their potential for predicting outcomes.

Acknowledgements

We would like to acknowledge Ranjit Lall, trial statistician for assistance with the data, and the trial team: Zara Hansen, Emanuela Castelnuovo, Emma J. Withers, Vivien Nichols, Rachel Potter and Martin R. Underwood.
References


