Modelling and classifying joint trajectories of self-reported mood and pain in a large cohort study

Rajenki Das¹,*, Mark Muldoon¹, Mark Lunt², John McBeth², Belay Birlie Yimer², and Thomas House¹

¹Department of Mathematics, University of Manchester, Manchester, UK
²Centre for Epidemiology Versus Arthritis, University of Manchester, Manchester, UK
*Corresponding Author: rajenki.das@manchester.ac.uk

Abstract

It is well-known that mood and pain interact with each other, however individual-level variability in this relationship has been less well quantified than overall associations between low mood and pain. Here, we leverage the possibilities presented by mobile health data, in particular the “Cloudy with a Chance of Pain” study, which collected longitudinal data from the residents of the UK with chronic pain conditions. Participants used an App to record self-reported measures of factors including mood, pain and sleep quality. The richness of these data allows us to perform model-based clustering of the data as a mixture of Markov processes. Through this analysis we discover four endotypes with distinct patterns of co-evolution of mood and pain over time. The differences between endotypes are sufficiently large to play a role in clinical hypothesis generation for personalised treatments of comorbid pain and low mood.

Author summary

Mood and pain are known to interact, and a mobile-phone application recorded information on the variations of mood and pain amongst people in the UK. Using this data, we observed that people have a general tendency of feeling the same mood and pain the next day. Studying further, we were able to separate the people into four groups- three of which were quite different from the general pattern of mood pain. The additional patterns we saw were 1) their mood and pain deteriorating the next day, 2) their mood and pain improving the next day and 3) mood is improving but pain deteriorates the next day. These additional characteristics tell us that there is no definite way that mood and pain are associated for everyone, and personalised treatment to tackle challenges in mood and pain can deliver better results.

1 Introduction

Mental disorder has been associated with a substantial excess in all-cause mortality risk (Prince et al., 2007). It is often accompanied by mood disorders which, according to the World Health Organisation (WHO) (Organization and Others, 2017), are one of the leading causes of disability. Mental health can suffer due to many social, physical and other factors, and mathematical approaches are uniquely placed to disentangle these complex issues. In view of the difficulty in clearly defining “mental illness” itself, simply linking its absence with positive mental health is not enough (Jahoda, 1958; Galderisi et al., 2015). One may not suffer from any “mental
illness”, yet not be mentally fit. So, identifying markers of mental health disorders remains a vital challenge.

Chronic pain is a persistent or intermittent pain that lasts for more than 3 months (Sheng et al., 2017), and approximately one fifth of the population in the USA and Europe are affected by it (Breivik et al., 2006). Chronic pain can cause a lot of emotional distress and affect lifestyle by interrupting activities (van den Berg-Emons et al., 2007) thereby it can potentially lower a person’s mood. Low mood and low self esteem often give birth to mental disorders like depression. Fordyce (1976) and Sternbach (1974) noted that depression is a frequent accompaniment to chronic pain while, Von Knorring et al. (1983) observed that those who suffer from depression often complain of pain. Depression, which is commonly associated with chronic pain (Fishbain et al., 1997; Zis et al., 2017), is one of the leading contributors to global disease burden (Whiteford et al., 2013; Collins et al., 2011). It has been seen that chronic pain and depression tend to coexist (Romano and Turner, 1985) and the relationship between the two is widely studied. Tang et al. (2008) showed that when a depressed mood was induced in patients with chronic back pain, their pain ratings increased, while participants with a happy mood had lower pain ratings. Fishbain et al. (1997) observed evidence against the hypothesis of depression preceding the development of pain and indicated that pain may play a causal role for depression. Chronic pain could be due to presence of inflammatory diseases (Ji et al., 2016), which cause inflammation in the body that can produce cytokines which can lower mood (Wright et al., 2005), and according to Irwin (2002), higher levels of biomarkers associated with inflammation are linked with depression. So today, the causal relationship of these associations between inflammation and mood disorders is said to be bi-directional (Rosenblat et al., 2014; Jones et al., 2020; Lwin et al., 2020).

It is widely recognised that healthcare increasingly involves dealing with comorbidities (Gijsen et al., 2001), and also personalisation of treatment plans (Vicente et al., 2020). Health issues such as mood disorders and conditions associated with chronic pain are often comorbid (Tunks et al., 2008; Agiera et al., 2010), but the manner in which these conditions influence each other varying from person to person is still considerably uncertain. Mood disorders or depression can be treated in three ways: antidepressants, psychotherapy and electro-convulsive therapy (ECT) (Nemeroff and Owens, 2002). Chronic pain treatments can be based on multiple aspects of pain experience like the intensity and quality of pain, and use of rescue analgesic medications (Patel et al., 2021). For certain types of chronic pain, drug therapy including intake of analgesics like non-steroidal anti-inflammatory drugs (NSAIDs) could be the option, while for others, a multimodal approach may be required (Portenoy, 2000) eg: a pharmacotherapy consisting analgesics and Cognitive–behavioural therapy (CBT) together can be effective when chronic pain and anxiety disorders co-occur (Asmundson and Katz, 2009). But when dealing with both mood disorders and chronic pain, especially when considering only pharmaceutical interventions, it must be noted that the combined usage of anti-depressants and NSAIDs can have negative effect, as shown in Shin et al. (2015); Hou et al. (2021) where there observed a risk of intracranial haemorrhage although there was no such association found in independent use.

Nowadays, technology is making its presence felt in several sectors, one of which is the health sector. It is only in the early 21st century that eHealth, a broad term for the combined usage of electronic and communication technologies in the health sector, emerged (Harrison and Lee, 2006). Many novel ways have developed to tackle healthcare issues and provide support. From wearable accessories to smartphone applications, all of these are aiding healthcare. From a global perspective, e-health is useful in dissemination of health information as well as ensuring that the most updated information is used to improve the health (Kwankam, 2004; Kendall et al., 2020). The WHO’s Global Observatory for eHealth defines mobile health (mHealth) as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices”. mHealth is a powerful
way to cater to individual requirements. Few of the benefits of the mHealth tools, especially for
the purpose of research, are: (i) cost-effectiveness while collecting voluminous amount of data;
(ii) more honesty in answers received as there is no direct human intervention in collection of
data; and (iii) convenience of easily linking mHealth apps to other link to other sensing tools.
More than one in four people are affected by mental health disorders like depression, anxiety etc.
worldwide (Ginn and Horder, 2012), and digital technology interventions show the potential to
extend support to those who suffer from mental health problems. There is a growing need to
make digital based mental health care aid accessible to as many people as possible (Naslund
et al., 2017), and in this study, we make use of digital health data to analyse mood-pain patterns
in a cohort of residents of the UK with chronic pain conditions.

We explore the association between pain and mood by analysing long records of self-reported,
daily data collected using a mobile phone application. We perform clustering on the basis of
the transitions of mood-pain and show how an intervention to improve low mood or high pain
symptoms can affect the clusters differently.

2 Methods

2.1 Data

We use data from the Cloudy with a Chance of Pain study (Reade et al., 2017; Dixon et al.,
2019), which was conducted to investigate the relationship between weather and pain, but in
doing so created an extremely rich dataset suitable to answer a diversity of research questions.
Data were collected from January 2016 to April 2017 from participants resident in the UK who
were aged 17 or above and had experienced chronic pain for at least 3 months preceding the
survey (Druce et al., 2017). These participants were recruited from the general public, with
particular support from the charity Versus Arthritis (then Arthritis Research UK) rather than
via the healthcare system, meaning that it is important to interpret these results as applying to
people’s experiences of chronic pain in the wider community rather than purely under clinical
management.

The cohort had 10,584 survey participants, each of whom was asked to rate their symptoms
and other variables on a mobile application in five ordinal categories (e.g. pain scores ranged
from 1 for no pain to 5 for very severe pain). Data were recorded for pain interference, sleep
quality, time spent outside, tiredness, activity, mood, well-being, pain severity, fatigue severity
and stiffness on a daily basis. However, participants did not always report all the data daily so
we considered only those (Mood, Pain) states where both the values are available, leaving us
with $N = 9990$ participants for our analysis. The average (rounded off) length of trajectories
is 44 observations.

In this paper we analyse trajectories of pairs of self-reported pain severity and mood scores.
Participants were asked to provide information on these on a five-point Likert scale, with accom-
panying text for each of the ordinal levels. For mood, a score of 1 represents worst mood and 5
represents best, whereas for pain a score of 1 represents least pain and 5 represents most.

For easier analysis of the data and interpretation of results, we regrouped the severity of mood
and pain into two categories each on the basis of the descriptions associated with each ordinal
value. Mood scores of 1–3 and 4–5 were labelled Bad (B) and Good (G) respectively, while pain
levels of 1–2 and 3–5 were, respectively, labelled Low (L) and High (H). Thus, at a given time,
a participant’s mood and pain scores fall into one of four states: GL; GH; BL; and BH. Full
details are shown in Table 1.

Participants self-reported diagnoses, and also provided information on age, sex, pain condition
diagnosed and the site of pain. They might have more than one condition and site of pain. The
list of conditions includes Rheumatoid arthritis, Osteoarthritis, Spondyloarthropathy, Gout, Unspecific arthritis, Fibromyalgia, Chronic headache and Neuropathic pain. The list of sites of pain taken in this analysis includes mouth or jaw, neck or shoulder, back pain, stomach or abdominal, hip pain, knee pain, and hands.

Code for this study is made available at: https://github.com/rajenkidas/EM-clustering-on-Markov-Chains. The data is scheduled to be made available to the wider research community via a trusted research environment in 2023.

Residual analysis

We performed an initial data analysis based on Pearson residuals, looking for notable patterns in the co-evolution of mood and pain over time using standard methodology as outlined by e.g. Bishop et al. (1975). Such an analysis particularly helps to visualise the ways in which observed patterns deviate from a simple ‘null’ model.

We begin by visualising a matrix of transitions observed in the data. Let $Y$ be the count matrix whose element $Y_{ij}$ denotes the total number of observed transitions—across all participants—from state $i$ one reporting day to state $j$ the next reporting day. We then perform Pearson residual analysis to compare observed transition probabilities with the expected values given a specified ‘null’ model assumption, which we fit by maximum likelihood estimation. Throughout this work we will use the standard result that the maximum likelihood estimator for a probability of an outcome is the observed number of such outcomes divided by the number of observations under binomial and Poisson sampling (which we also assume throughout as appropriate).

We have seen that participants are most likely to remain in their current state rather than move to another one. That is, their mood and pain scores do not usually change from one day to the next, as shown in Fig 1A. These observations allow us to define a simple first model for their behaviour and perform residual analyses as described below. In this exploratory analysis we work with the original data, so there are $n = 5 \times 5 = 25$ states.

We therefore define a null model in which the number of participants starting in state $i$ is $N_i$, the probability of staying in state $i$ is $\pi_i$ and when a person does change state, the probabilities $P_{ij}$ of a transition from state $i$ to state $j \neq i$ are uniform. The model parameters can then have maximum likelihood estimators (indicated with hats) as follows. For $i, j \in \{1, 2, \ldots, n\}$,

$$
\hat{N}_i = \sum_{k=1}^{n} Y_{ik}, \quad \hat{\pi}_i = \frac{Y_{ii}}{\sum_{k=1}^{n} Y_{ik}}, \quad \hat{P}_{ij} = \begin{cases} \hat{\pi}_i & \text{if } i = j, \\ \frac{1 - \hat{\pi}_i}{n-1} & \text{otherwise,} \end{cases} \quad E_{ij} = \hat{N}_i \hat{P}_{ij}, \quad (1)
$$

where $E_{ij}$ is the $(i, j)$-th element of the matrix of expected counts, $E$. The associated entry in the Pearson residual matrix $R$ is then given by

$$
R_{ij} = \frac{Y_{ij} - E_{ij}}{\sqrt{E_{ij}}}. \quad (2)
$$

Since we expect such residuals to be asymptotically standard normal under the null (Bishop et al., 1975), we will interpret these as values over 2 indicating significantly more events than expected under the null, and values under −2 indicating significantly fewer.

Clustering analysis

In this section we outline methods used to classify the participants using unsupervised learning, organising the participants into clusters on the basis of their sequences of reduced (Mood, Pain) states: GL, GH, BL, BH.
Model setup

We assume the sequence of self-reported mood-pain states $X = (X_t; t \geq 0)$ is generated by a Markov chain:

$$\Pr(X_{t+1} = j \mid X_0 = k_0, \ldots, X_t = i) = \Pr(X_{t+1} = j \mid X_t = i) =: P_{ij},$$

where the $P_{ij}$ are called the chain’s transition probabilities.

Our data consists of trajectories of mood-pain pairs that we reduce to matrices tabulating numbers of transitions observed for each participant individually. We then cluster these count-matrices by using the EM algorithm to fit a mixture of Markov chains with a distinct matrix of transition probabilities for each component of the cluster.

Let the number of states be $n$ and the number of participants be $S$. We write $C$ for the matrix of total count of transitions from one state to another, and use $C_s$ for the matrix of counts of transitions that appear in the trajectory of states of mood-pain of participant $s$. We note that $C$ is distinguished from the count matrix $Y$ introduced before in Residual Analysis section since now it involves only the four reduced states.

The expectation-maximisation algorithm

The classical Expectation-Maximisation (EM) algorithm (Dempster et al., 1977) provides a way to do maximum-likelihood estimation of parameters in a setting where some variables are unobserved or unknown. In our case, the missing data are the classes to which the participants belong. The algorithm involves iteration of two alternating steps: the E, or expectation step, during which one computes the expected value of the log likelihood for the observed data, given the current estimates of the parameters, and the M, or maximisation, step during which one re-estimates the parameters is maximising the expected value as calculated in the E-step.

The details of this algorithm are given in Supplementary Material §S.1.3. Its outputs are a number of clusters $K$, and an $S \times K$ matrix $\Gamma$ such that its $(s, c)$-th element $\Gamma_{sc}$ is the probability that participant $s$ belongs to cluster $c$. Finally, cluster assignments are then made on the basis of the class membership probabilities: participants are assigned to whichever cluster they have the highest probability of belonging to.

Associated stationary distribution

The stationary distribution for a Markov chain with $n \times n$ transition matrix $M$ has probability $x_i$ associated with state $i$, where $\mathbf{x} = (x_i)$ solves the left Eigenvalue equation

$$x_k = \sum_{i=1}^{n} x_i M_{ik},$$

where $k \in \{1, \ldots, n\}$, and we impose conditions ensuring that $\mathbf{x}$ is a probability vector: $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 1$.

The solution to Eqn (4) need not be unique, but as the transition matrices of our problem are regular, we do get a unique stationary distribution for each component of the mixture (Stirzaker, 2003). That is, for each cluster, we get a distribution over the states BH, BL, GH and GL. Further, as the Markov chains are ergodic, the modelled expected fraction of time an individual participant spends in state $k$ is given by $x_k$. 
Intervention

In this section, we explore the prospect of alleviating low mood or high pain, which can be done by taking the appropriate treatment targeting mood or pain. We naively examine how the interventions could work by altering the transition probabilities associated with the clusters and see what effect this has on the cluster’s stationary distribution. Throughout, we will let the transition probability matrix before intervention be represented as:

\[
\begin{bmatrix}
GL & GH & BL & BH \\
BH & M_{11} & M_{12} & M_{13} & M_{14} \\
BL & M_{21} & M_{22} & M_{23} & M_{24} \\
GH & M_{31} & M_{32} & M_{33} & M_{34} \\
GL & M_{41} & M_{42} & M_{43} & M_{44}
\end{bmatrix}.
\] (5)

Improving mood

To model an improvement in mood, we increase the probabilities of transitions from states of bad mood to those with good mood. We get an updated transition matrix \( M'_c \) for every cluster \( c \) in the following way:

\[
\begin{bmatrix}
GL & GH & BL & BH \\
BH & M_{11} + \beta_M & M_{12} + \beta_M & 0.8 \times (M_{13} + M_{14} - 2\beta_M) & 0.2 \times (M_{13} + M_{14} - 2\beta_M) \\
BL & M_{21} + \beta_M & M_{22} + \beta_M & 0.8 \times (M_{23} + M_{24} - 2\beta_M) & 0.2 \times (M_{23} + M_{24} - 2\beta_M) \\
GH & M_{31} & M_{32} & M_{33} & M_{34} \\
GL & M_{41} & M_{42} & M_{43} & M_{44}
\end{bmatrix},
\] (6)

where the rows are labelled by the (Mood, Pain) states from which the transition starts, while the columns are labelled by the states to which it goes. Here \( \beta_M \) must be chosen so that all transition probabilities remain in the range \( 0 \leq M'_{cij} \leq 1 \). For our fitted transition matrices, these constraints mean that \( 0 \leq \beta_M \leq 0.15 \).

One can see that we distribute the probabilities disproportionately between transitions to BH and BL from BH and BL. This has been done to reduce the probability of moving to BL, which we wish to model as less likely under an intervention assumed to be beneficial. In fact, in general the probability of moving to good mood from bad mood could have been achieved in numerous other ways through changes to the full matrices. The choice used here permits a more substantial increase in the probabilities of improved mood than simpler formulæ, many of which are strongly constrained by the necessity of keeping all probabilities to the laws of probability.

Improving pain

Similar to improvement of mood, we considered altering the transition probabilities to improve pain, which means increasing probability of transitioning to low pain through adding and subtracting \( \beta_P \) as shown below for the updated transition probability matrix \( M'_c \) for every cluster \( c \) in the following way:

\[
\begin{bmatrix}
GL & GH & BL & BH \\
BH & M_{11} + \beta_P & 0.8 \times (M_{12} + M_{14} - 2\beta_P) & M_{13} + \beta_P & 0.2 \times (M_{13} + M_{14} - 2\beta_M) \\
BL & M_{21} & M_{22} & M_{23} & M_{24} \\
GH & M_{31} + \beta_P & 0.8 \times (M_{32} + M_{34} - 2\beta_M) & M_{33} + \beta_P & 0.8 \times (M_{32} + M_{34} - 2\beta_M) \\
GL & M_{41} & M_{42} & M_{43} & M_{44}
\end{bmatrix}.
\] (7)

Here \( 0 \leq \beta_P \leq 0.2 \). In both cases, we then examine the resulting changes in the stationary distributions to see the consequences of the intervention for each cluster individually.
Results

Residual Analysis

The resulting transition probability matrix is illustrated in Fig 1A, which is a heatmap illustrating the probabilities with which participants switch from one pair of mood-pain scores to another. It is based on the original data and so has $5 \times 5 = 25$ possible states and $25 \times 25 = 625$ possible transitions. It has rows labelled by a current mood-pain pair and columns labelled by the mood-pain pair on the following day.

Note that the diagonal elements—those that correspond to remaining in the same state on successive days—have high probabilities. The entries at upper right and lower left, which correspond, respectively, to the worst and best mood-pain scores, are especially large (near their maximum value, 1) indicating that participants at the extremes of the scale have a strong tendency to remain there.

In Fig 1B and C which illustrate the distribution of residuals for this model as computed with Eqn (2), clearly show that the residuals do not appear to be normally distributed. Looking at the residual heatmap in Fig 1D, we can say that the naïve model specified by Eqn (1) does not describe the data well.

This suggests we try another model or check for latent variables or clusters. We try another model in the Supplementary (S5) which showed an improvement in fitting since the residual range decreases in Fig S8, but it still did not fit the data well as we see in Fig S9. So we move on to clustering the data, as explained in the next section.

Clustering

We found four clusters using the EM algorithm to do model-based clustering using a mixture of Markov chains, as illustrated in Fig 2, where the clusters are represented by heatmaps of their transition matrices.

Before describing the clusters, it should be noted that GL is the best state as both mood and pain are good, while BH is the least preferable state to be in as both mood and pain are bad here. Based on the transition probabilities, the four clusters for mood-pain dynamics can be broadly characterised as:

**Cluster 1:** Movement to the least preferable state. 1783 members.

Here, we see that there are high probabilities of moving to the state where there is bad mood and high pain.

**Cluster 2:** Movement to the ideal state. 1558 members.

In this cluster, we observe, irrespective of the current state, a participant is most likely to be in good mood and low pain the next day.

**Cluster 3:** Good mood, high pain. 2019 members.

In this cluster, the dominant movement is to the state with good mood and high pain.

**Cluster 4:** Remain in the same state. 4630 members.

Most of the participants tend to stay in the same state.

Given the total of 9990 participants, we see that it is most common for participants (46%) to be members of Cluster 4 involving staying in the same state, which is consistent with our exploratory analysis of transitions. The smallest cluster (number 2) with 16% of participants, consists of those who tend to the ideal state, but at the same time, not many (18%) are in Cluster number 1 that tends to the worst state. The remainder (20%) belong to the third cluster: good mood, high pain.
In Fig 3, we present a set of comparisons of properties of the clusters. The stationary distributions as defined by Eqn (4) are shown in Fig 3A, and as would be expected from the full estimated transition probability estimates they are derived from: Cluster 1 has most probability mass on BH; Cluster 2 has most probability mass on GL; Cluster 3 has most probability mass on GH; and Cluster 4 has evenly distributed probability masses.

In Fig 3B, we compare age distributions by sex and cluster, seeing that Clusters 1 and 4 have comparable age distributions, but Cluster 3 is associated with older ages than these two and Cluster 2 is associated with older ages than all three other clusters. Males are typically older than females in all clusters.

Participants had one or more conditions and sites of pain and the log odds ratios for these per cluster are shown in Figs 3C and 3D. These show that while some conditions and sites such as gout and hands are not strongly associated with any cluster, for others this is not the case. Fibromyalgia and stomach pain are particularly strongly associated with Cluster 2, for example.

**Intervention**

We look at how interventions could work help alleviate the symptoms of bad mood and high pain.

In Fig 4A, Cluster 2 shows least improvement in mood, while Cluster 1 shows the most followed by Cluster 4. Decrease in state BH is the highest for Cluster 1, followed by Cluster 4 and least for Cluster 2. Overall, Cluster 1 shoes the most drastic changes in probability distribution while Cluster 2 is the least. We also note that in the case of improving mood from bad mood, state BL probability drops for Clusters 2 and 4, while it increases for 1 and 3.

In Fig 4B, we again find Cluster 1 with maximum changes. When intervened to improve pain by lessening the intensity of pain, probability of GH state improves only for Cluster 1.

**Discussion**

In this work, we have performed an analysis of joint trajectories of mood and pain of participants in the large mobile health cohort, “Cloudy with a Chance of Pain”. In addition to analysis of the full set of transitions using residuals, we performed clustering on transitions between a simplified set of variables and in doing so found four digital behavioural phenotypes on the basis of people’s past trajectories of their mood-pain states. This suggests that even though mood and pain have been known to be correlated, the association may not be generalised in one single way for an entire population.

Previous studies on mood-pain relationships have tended to reach the conclusions on universal associations between mood and pain – i.e. generalising the result for everyone. The clusters found in this study emphasise that mood-pain relationships may differ between (groups of) individuals. The varying relationships between mood and pain, as shown by the clusters, highlights that such variability should be taken into account when considering expected future associations - for example, in a clinical prediction model, an approach of personalising forecasts could be taken.

Going beyond association to look at mechanism and causation, we stress that we have not performed causal inference and so results should all be interpreted as indicative of (potential) association magnitudes rather than as causal statements. Nevertheless, the interpretability of the observed clusters and their diversity in terms of e.g. conditions and sites of pain represented suggests that there may be associated endotypes – i.e. clusters representing distinct mechanisms
of disease. If such causally distinct groups exist, then our hypothetical investigation of interventions that target either mood or pain individually suggests that we might expect clinically significant differences from different treatment depending on an individual’s endotype.

Our study has some limitations that should be borne in mind when interpreting results. The first of these is, as discussed above, that we consider associations rather than causation. Furthermore, we have assumed missing values – primarily arising when participants did not enter data on one day – can be ignored and so have removed them; although this is not a major component of the data an alternative would be to model non-response as a separate value. Along related lines, the simplification of the state space, while necessary for the EM algorithm to produce plausible transition matrices for each cluster, involves some information loss and this leaves open the possibility of more sophisticated methodology to perform the clustering. Also, factors common to all observational studies such as this one are important to bear in mind, particularly that individuals are selected from the general rather than a clinical population.

Extension of the work presented here could include applying the same methodology to more datasets to check if the phenotypes found are reproducible. This would further strengthen the likelihood of different causal relationships holding within clusters. To make a fuller assessment of likely causation, however, expected relationships between all observed and unobserved variables would need to be specified, and ideally intervention studies run. Additionally, this work can be extended by including socio-economic factors, extra latent variables like sleep quality, environment etc. Another direction would be to apply different techniques to this dataset, such as linear model based approaches that can identify latent classes (Proust-Lima et al., 2015; Komárek and Komárková, 2013). Different methods may allow the Markovian assumption made in our work to be relaxed, allowing for e.g. consideration of patterns in longer sequences of data, but at the cost of the ability to model out of sample behaviour as Markov chains allow.

Ultimately, our hope is that work on observational data such as that presented here can aid with hypothesis generation for future clinical studies of more personalised interventions for common problems such as low mood and chronic pain.
Financial disclosure

RD and TH are supported by the Engineering and Physical Sciences Research Council. ML, JMcB and BBY are supported by Centre for Epidemiology Versus Arthritis. TH is also supported by the Royal Society, the Medical Research Council and the Alan Turing Institute for Data Science and Artificial Intelligence. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgement

We thank Dr Elaine Mackey and Prof Will Dixon for helping with the Data Availability Statement.

Ethics statement

This is a secondary analysis. Ethical approval of the primary data was obtained from the University of Manchester Research Ethics Committee (ref: ethics/15522) and from the NHS IRAS (ref: 23/NW/0716).

Conflicts of interest

The authors have declared that no competing interests exist.

Author contributions

RD performed the analysis. All authors contributed to the formulation of research questions and the writing of the paper.
| Score | Mood Text | Mood Binary | Pain Score | Pain Text | Pain Binary |
|-------|-----------|-------------|------------|-----------|-------------|
| 1     | Depressed | Bad         | 5          | Very severe pain | High |
| 2     | Feeling low | Bad        | 4          | Severe pain | High |
| 3     | Not very happy | Bad      | 3          | Moderate pain | High |
| 4     | Quite happy | Good       | 2          | Low pain | Low |
| 5     | Very happy | Good        | 1          | No pain | Low |

Table 1: Mood and pain scores, descriptions, and binary classifications. *Score* is the value on a Likert scale available to participants, *Text* is the description presented to them when recording these data, and *Binary* is our binary classification into ‘Good’ (G) or ‘Bad’ (B) for Mood, and ‘Low’ (L) and ‘High’ (H) for Pain.
Figure 1: A is the heatmap of probabilities of transitions from one state to another. B is the scatter plot of expected values and the residuals. C shows a histogram of the residuals as well as a blue curve giving the probability density function of a normal distribution having the same mean and variance as the residuals. D is a heatmap of the matrix of residuals based on the model specified by Eqn. (1).
Figure 2: Heatmaps of the transition probability matrices for the four clusters where G, B, L and H imply good mood, bad mood, low pain and high pain respectively.
Figure 3: A stationary distributions for the four clusters. B age distributions for each cluster. C and D represent log odds ratio of condition and site of pain respectively, per cluster.

Figure 4: Change in stationary distributions by an intervention. In A, $\beta_M = 0.15$ is added to probabilities of transitioning from bad mood to good mood, while in B, $\beta_P = 0.15$ has been added to the probabilities of transitions from high pain to low pain.
References

Prince, M.; Patel, V.; Saxena, S.; Maj, M.; Maselko, J.; Phillips, M. R.; Rahman, A. No health without mental health. *The Lancet* 2007, *370*, 859–877.

Organization, W. H.; Others, *Depression and other common mental disorders: global health estimates*; 2017.

Jahoda, M. *Current concepts of positive mental health.*; Basic Books, 1958.

Galderisi, S.; Heinz, A.; Kastrup, M.; Beezhold, J.; Sartorius, N. Toward a new definition of mental health. *World Psychiatry* 2015, *14*, 231.

Sheng, J.; Liu, S.; Wang, Y.; Cui, R.; Zhang, X. The link between depression and chronic pain: neural mechanisms in the brain. *Neural Plasticity* 2017.

Breivik, H.; Collett, B.; Ventafridda, V.; Cohen, R.; Gallacher, D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European Journal of Pain* 2006, *10*, 287–333.

van den Berg-Emons, R. J.; Schasfoort, F. C.; de Vos, L. A.; Bussmann, J. B.; Stam, H. J. Impact of chronic pain on everyday physical activity. *European Journal of Pain* 2007, *11*, 587–593.

Fordyce, W. E. *Behavioral methods for chronic pain and illness*; Mosby St. Louis, 1976; Vol. 1.

Sternbach, R. A. *Pain patients: Traits and treatment*; Academic Press New York, 1974.

Von Knorring, L.; Perris, C.; Eisenmann, M.; Eriksson, U.; Perris, H. Pain as a symptom in depressive disorders: I. Relationship to diagnostic subgroup and depressive symptomatology. *Pain* 1983.

Fishbain, D. A.; Cutler, R.; Rosomoff, H. L.; Rosomoff, R. S. Chronic pain-associated depression: antecedent or consequence of chronic pain? A review. *The Clinical Journal of Pain* 1997, *13*, 116–137.

Zis, P.; Daskalaki, A.; Bountouni, I.; Sykioti, P.; Varrassi, G.; Paladini, A. Depression and chronic pain in the elderly: links and management challenges. *Clinical Interventions in Aging* 2017, *12*, 709.

Whiteford, H. A.; Degenhardt, L.; Rehm, J.; Baxter, A. J.; Ferrari, A. J.; Erskine, H. E.; Charlson, F. J.; Norman, R. E.; Flaxman, A. D.; Johns, N., et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet* 2013, *382*, 1575–1586.

Collins, P. Y.; Patel, V.; Joestl, S. S.; March, D.; Insel, T. R.; Daar, A. S.; Bordin, I. A.; Costello, E. J.; Durkin, M.; Fairburn, C., et al. Grand challenges in global mental health. *Nature* 2011, *475*, 27–30.

Romano, J. M.; Turner, J. A. Chronic pain and depression: does the evidence support a relationship? *Psychological Bulletin* 1985, *97*, 18.

Tang, N. K.; Salkovskis, P. M.; Hodges, A.; Wright, K. J.; Hanna, M.; Hester, J. Effects of mood on pain responses and pain tolerance: an experimental study in chronic back pain patients. *Pain* 2008, *138*, 392–401.

Ji, R.-R.; Chameissian, A.; Zhang, Y.-Q. Pain regulation by non-neuronal cells and inflammation. *Science* 2016, *354*, 572–577.
Wright, C.; Strike, P.; Brydon, L.; Steptoe, A. Acute inflammation and negative mood: mediation by cytokine activation. *Brain, Behavior, and Immunity* **2005**, *19*, 345–350.

Irwin, M. Psychoneuroimmunology of depression: clinical implications. *Brain, Behavior, and Immunity* **2002**, *16*, 1–16.

Rosenblat, J. D.; Cha, D. S.; Mansur, R. B.; McIntyre, R. S. Inflamed moods: a review of the interactions between inflammation and mood disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **2014**, *53*, 23–34.

Jones, B. D.; Daskalakis, Z. J.; Carvalho, A. F.; Strawbridge, R.; Young, A. H.; Mulsant, B. H.; Husain, M. I. Inflammation as a treatment target in mood disorders. *BJPsych Open* **2020**, *6*.

Lwin, M. N.; Serhal, L.; Holroyd, C.; Edwards, C. J. Rheumatoid arthritis: the impact of mental health on disease: a narrative review. *Rheumatology and Therapy* **2020**, *7*, 457–471.

Gijsen, R.; Hoeymans, N.; Schellevis, F. G.; Ruwaard, D.; Satariano, W. A.; van den Bos, G. A. Causes and consequences of comorbidity: a review. *Journal of Clinical Epidemiology* **2001**, *54*, 661–674.

Vicente, A. M.; Ballensiefen, W.; Jönsson, J.-I. How personalised medicine will transform healthcare by 2030: the ICPERMed vision. *Journal of Translational Medicine* **2020**, *18*, 180.

Tunks, E. R.; Crook, J.; Weir, R. Epidemiology of chronic pain with psychological comorbidity: prevalence, risk, course, and prognosis. *The Canadian Journal of Psychiatry* **2008**, *53*, 224–234.

Agüera, L.; Failde, I.; Cervilla, J. A.; Díaz-Fernández, P.; Mico, J. A. Medically unexplained pain complaints are associated with underlying unrecognized mood disorders in primary care. *BMC Family Practice* **2010**, *11*, 1–8.

Nemeroff, C. B.; Owens, M. J. Treatment of mood disorders. *Nature Neuroscience* **2002**, *5*, 1068–1070.

Patel, K. V.; Amtmann, D.; Jensen, M. P.; Smith, S. M.; Veasley, C.; Turk, D. C. Clinical outcome assessment in clinical trials of chronic pain treatments. *Pain Reports* **2021**, *6*.

Portenoy, R. K. Current pharmacotherapy of chronic pain. *Journal of Pain and Symptom Management* **2000**, *19*, 16–20.

Asmundson, G. J.; Katz, J. Understanding the co-occurrence of anxiety disorders and chronic pain: state-of-the-art. *Depression and Anxiety* **2009**, *26*, 888–901.

Shin, J.-Y.; Park, M.-J.; Lee, S. H.; Choi, S.-H.; Kim, M.-H.; Choi, N.-K.; Lee, J.; Park, B.-J. Risk of intracranial haemorrhage in antidepressant users with concurrent use of non-steroidal anti-inflammatory drugs: nationwide propensity score matched study. *BMJ* **2015**, *351*.

Hou, P.-C.; Lin, F.-J.; Lin, S.-Y.; Hwang, T.-J.; Wang, C.-C. Risk of intracranial hemorrhage with concomitant use of antidepressants and nonsteroidal anti-inflammatory drugs: a nested case-control study. *Annals of Pharmacotherapy* **2021**, *55*, 941–948.

Harrison, J. P.; Lee, A. The Role of E-Health in the Changing Health Care Environment. *Nursing Economics* **2006**, *24*, 283–8, 279; quiz 289, Copyright - Copyright Anthony J. Jannetti, Inc. Nov/Dec 2006; Document feature - ; Tables; Last updated - 2014-03-29.

Kwankam, S. Y. What e-Health can offer. *Bulletin of the World Health Organization* **2004**, *82*, 800–801.
Kendall, M.; Milson, L.; Abeler-Dörner, L.; Wymant, C.; Ferretti, L.; Briers, M.; Holmes, C.; Bonsall, D.; Abeler, J.; Fraser, C. Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: a preliminary analysis. *The Lancet Digital Health* 2020, 2, e658–e666.

Ginn, S.; Horder, J. “One in four” with a mental health problem: the anatomy of a statistic. *BMJ* 2012, 344, e1302.

Naslund, J. A.; Aschbrenner, K. A.; Araya, R.; Marsch, L. A.; Unützer, J.; Patel, V.; Bartels, S. J. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. *The Lancet Psychiatry* 2017, 4, 486–500.

Reade, S.; Spencer, K.; Sergeant, J. C.; Sperrin, M.; Schultz, D. M.; Ainsworth, J.; Lakshminarayana, R.; Hellman, B.; James, B.; McBeth, J., et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR mHealth and uHealth* 2017, 5, e6496.

Dixon, W. G.; Beukenhorst, A. L.; Yimer, B. B.; Cook, L.; Gasparini, A.; El-Hay, T.; Hellman, B.; James, B.; Vicedo-Cabrera, A. M.; Maclure, M., et al. How the weather affects the pain of citizen scientists using a smartphone app. *NPJ Digital Medicine* 2019, 2, 1–9.

Druce, K. L.; McBeth, J.; van der Veer, S. N.; Selby, D. A.; Vidgen, B.; Georgatzis, K.; Hellman, B.; Lakshminarayana, R.; Chowdhury, A.; Schultz, D. M., et al. Recruitment and ongoing engagement in a UK smartphone study examining the association between weather and pain: cohort study. *JMIR mHealth and uHealth* 2017, 5, e168.

Bishop, Y.; Fienberg, S. E.; Holland, P. W. *Discrete Multivariate Analysis: Theory and Practice*; Massachusetts Institute of Technology Press: Cambridge, 1975.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 1977, 39, 1–22.

Stirzaker, D. *Elementary Probability*; Cambridge University Press: Mathematical Institute and St. John’s College, University of Oxford, 2003.

Proust-Lima, C.; Philipps, V.; Liquet, B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *arXiv preprint arXiv:1503.00890 2015*.

Komárek, A.; Komárková, L. Clustering for multivariate continuous and discrete longitudinal data. *The Annals of Applied Statistics* 2013, 7, 177–200.

Dorea, C. C.; Goncalves, C. R.; Resende, P. Simulation results for Markov model selection: AIC, BIC and EDC. Proceedings of World Congress on Engineering and Computer Science. 2014; pp 899–901.

Yin, P.; He, Q.; Liu, X.; Lee, W.-C. It takes two to tango: Exploring social tie development with both online and offline interactions. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 2016, 9, 174–187.
S.1 Supplementary Text

S.1.1 Summary statistics of data and results

S.1.1.1 Main data

Here we provide an overview of the data. Both mood and pain were rated on a five point scale where 1 for mood is the worst score while 1 for pain is the best. Similarly 5 means the mood is the best and pain is at its worst. Mean values of mood and pain are 3.6 and 2.7 respectively. Number of NA’s: 344784 in mood and 349760 in pain. After removing these NA’s from the data, we are left with 9990 participants instead of 10584. More details about the data can be found at www.cloudywithachanceofpain.com.

With five possibilities of each of mood and pain, there are $5 \times 5$ i.e. 25 possible (mood, pain) pairs, which would become the states of our Markov processes, leading to transition matrices with $25 \times 25 = 625$ entries. But instead, we reduce the total number of states by regrouping the scores of mood and pain into good and bad, and low and high categories respectively and then taking pairs of these regrouped scores. For Mood, the Bad (B) scores are $\{1,2,3\}$ while the Good (G) ones are $\{4,5\}$. For Pain, Low (L) is $\{1,2\}$ while High (H) is $\{3,4,5\}$.

For Mood, number of B’s and G’s are, respectively, 153922 and 288145. For Pain the number of H’s and L’s are, respectively, 245344 and 196723. The frequencies with which the four (Mood, Pain) states are observed is shown in Fig S1. The frequencies for BH, BL, GH and GL are 113632, 40290, 131712 and 156433, respectively.

Fig S2 gives the overall age distribution of the cohort. It shows that the mean age for the women and men are approximately 47 years and 52 years respectively.

Tables S1 and S2 give the characteristics of the 9990 study participants included in our analysis.

S.1.1.2 Clustered data

We have also included the clustered heatmaps without any regrouping of the categorical variables in Fig S12. Note that the probability of moving to a state gets reduced from 0.25 (in case of grouped data) to 0.04, which introduces problems of interpretability and generalisation, but can help in understanding the contribution of a transition probability to the grouped clusters, in combination with Fig S11.

We see that within cluster composition of the conditions and sites of pain are both more or less the same across the clusters as shown in Figs S3A and S3B. The biggest difference which can be immediately noted is that in the composition of conditions in cluster 2, Fibromyalgia and Neuropathic pain are less while unspecified arthritis is comparatively more. Next, when we look at how much of a cluster constitutes a condition, we find the proportions are in sync with the in general sizes of the clusters. Noticeable difference in S3C is that cluster 2 amounts to very little proportion of Fibromyalgia, while for site of pain, we find in S3D, cluster 2 constitutes very little of site of pain as the face, while cluster 1 takes up a high proportion compared to the rest of its contributing proportions.

Table S3 gives the mean age in years and the response rate in percentage per cluster. Mean age was calculated by taking the average of age in a group, after removing all the NA values. Response rate is the percentage of participants in a cluster who gave their date of birth details.
S.1.2 Definition of the log odds ratio

Tables S4 and S5 give the log odds ratio of a condition and site of pain respectively in a cluster compared to the other clusters. In this case, we are considering the log odds ratio as defined for 2×2 contingency tables (Bishop et al., 1975), which is a similar concept to that in logistic regression but not fully isomorphic. To calculate such a log odds ratio, we consider the following contingency table first:

|                  | Cluster | Remaining clusters |
|------------------|---------|--------------------|
| Condition        | $n_{11}$| $n_{12}$           |
| Remaining conditions | $n_{21}$| $n_{22}$           |

Explicitly:

$n_{11}$ is the number of participants with a specific condition in a cluster.

$n_{12}$ is the number of participants with the specific condition not in the cluster.

$n_{21}$ is the number of participants without the specific condition in the cluster.

$n_{22}$ is the number of participants without the specific condition not in the cluster.

The log odds ratio is then given by:

$$L = \log \left( \frac{n_{11}n_{22}}{n_{12}n_{21}} \right) = \log(n_{11}) + \log(n_{22}) - \log(n_{12}) - \log(n_{21})$$

The standard error of this quantity is asymptotically equal to

$$\sigma = \sqrt{n_{11}^{-1} + n_{12}^{-1} + n_{21}^{-1} + n_{22}^{-1}}),$$

as shown in Bishop et al. (1975). Therefore, the 95% Confidence Interval is approximately $L \pm 1.96 \sigma$.

Similarly we calculate for site of pain by building the following contingency table:

|                  | Cluster | Remaining clusters |
|------------------|---------|--------------------|
| Site of pain     | $n_{11}$| $n_{12}$           |
| Remaining sites of pain | $n_{21}$| $n_{22}$           |

Explicitly:

$n_{11}$ is the number of participants with a specific site of pain in a cluster.

$n_{12}$ is the number of participants with the specific site of pain not in the cluster.

$n_{21}$ is the number of participants without the specific site of pain in the cluster.

$n_{22}$ is the number of participants without the specific site of pain not in the cluster.

And we can then calculate a log odds ratio as for conditions. In general, a positive log odds ratio indicates that the site or condition is more commonly in a cluster, and a negative that it is less commonly so.

S.1.3 Description of the EM algorithm

The matrix $\Gamma$ is initialised randomly with probabilities chosen such that every row sums to 1. The mixture of Markov chains is then specified by a weight vector $\omega$ of length $K$ in which

$$\hat{\omega}_k = \frac{\sum_{s=1}^{S} \Gamma_{sk}}{S}$$

(S1)
Using the mixture weights and count matrices, we then estimate the parameters of the per-cluster Markov chains, which are transition probability matrices $M$. For cluster $k$, the estimate is

$$
\hat{M}_{kij} = \frac{\sum_{s=1}^{S} \Gamma_{sk} C_{sij}}{\sum_{k=1}^{K} \sum_{s=1}^{S} \Gamma_{sk} C_{sij}}. \tag{S2}
$$

The rows of the a participant’s count matrix are taken to follow a Multinomial distribution and so define an $S \times K$ matrix of expected likelihoods whose entry $\Lambda_{sk}$ gives the likelihood of observing participant $s$’s trajectory given that the participant $s$ belongs to cluster $k$. It is given by

$$
\Lambda_{sk} = \prod_{i,j=1}^{n} M_{kij}^{C_{sij}},
$$

where we have suppressed a multinomial coefficient that does not depend on the parameters of the mixture model and so does not affect maximum-likelihood estimates. The log-likelihood for participant $s$ and cluster $k$ is thus given by,

$$
\log \Lambda_{sk} = \sum_{i,j=1}^{n} C_{sij} \log (M_{kij}). \tag{S3}
$$

Using Eqn. (S3), we can specify the algorithm steps as follows:

- **Expectation step:** The expected values of the matrix of class membership probabilities are computed using

$$
\hat{\Gamma}_{sk} = \frac{\omega_k \Lambda_{sk}}{\sum_{c=1}^{K} \omega_c \Lambda_{sc}}
$$

- **Maximisation step:** This involves re-estimating the parameters of the mixture using Eqns. (S1) and (S2).

One performs the steps in alternation until the matrix $\Gamma$ converges. That is, one keeps track of the two most recent estimates of $\Gamma$ — call them $\hat{\Gamma}$ and $\hat{\Gamma}'$ — and continues iterating until a convergence criterion such as

$$
|\hat{\Gamma} - \hat{\Gamma}'| < \epsilon,
$$

for some sufficiently small $\epsilon$ is met.

### S.1.4 Choosing the number of clusters

To find the total log-likelihood of the observed data, we made use of the log-likelihood per participant per cluster $\log(\Lambda_{sk})$ as found in Eqn. (S3):

$$
\sum_{s=1}^{S} \log \left( \sum_{k=1}^{K} \omega_k \Lambda_{sk} \right), \tag{S4}
$$

where $s$ denotes the participant and $k$ ranges over the clusters.

Fig S4 shows the negative log-likelihood as a function of the number of clusters. As expected, we see the largest drop between $K = 1$ and $K = 2$, with diminishing returns as $K$ increases. Ideally a quantitative trade-off between model complexity and reducing loss would be made, however most such formal measures do not perform well for large datasets such as the one we consider here, and so here we use the elbow method and assess optimal value of $K$. To this end, we continue to see relatively big decreases as we increase the number of components to $K = 3$ and $K = 4$. The curve continues falling after $K = 4$, but as the decreases in negative
log-likelihood are modest, following this, 4 clusters is indicated by the elbow method. To make this reasoning clearer, we extrapolated consecutive negative Log-Likelihood differences in Fig S5 to show that $K = 4$ indeed sits at the crossover between the small- and large-$K$ regimes.

To demonstrate the difficulties in using a quantitative trade-off for data of our scale, we consider the Bayesian Information Criterion (BIC), which is a method to compare statistical models by calculating the information loss between the true and evaluated model by penalising the sample size to address the problem of overestimating the number of parameters (Dorea et al., 2014).

First we compute the likelihood $L$ for the model in consideration. Then, we write the BIC as

$$BIC = -2\log L + k \log(n),$$

where $\log L$ same as that given in S4. $k$ is the total number of parameters and $n$ is the number of observations.

For the problem in question, $k = K \times \text{size}(T) \times (\text{size}(T) - 1) + K - 1$ where $K$ is the total number of clusters and $T$ is the transition probability matrix therefore, for the 4-state Markov mixture models, we get $k = K \times 4 \times (4 - 1) + K - 1 = 13K - 1$. Number of observations is the total number of transitions which is 432077. We select the optimal number of clusters in the similar way as that done for negative log-likelihood above. Its plot is given in Fig S7. However, we do not see much difference from the previous model selection plot in Fig S4. This is because in comparison to the large size of the dataset, the penalty in BIC is too small and does not give a criterion value much different from the negative log-likelihood. It is common to have BIC and other criteria to keep on decreasing in case of large datasets and other authors such as Yin et al. (2016) used similar reasoning to ours when performing model selection.

Additionally, we performed clustering into 5 to 8 components as shown in Figs S13, S14, S15 and S16. We see that for $K > 4$, some clusters are very small, involve repeated patterns, and are hard to interpret (in contrast to the discussion as in the Discussion section of the main paper) also leading us to prefer $K = 4$.

**S.1.5 Residual analysis: A second model**

In order to fit a better model, let’s us go back to the observed transition matrix plotted in Fig 1A. The observation that the diagonal elements are high has already been incorporated into the model described by Eqns. (1). If we look at the heatmap more carefully, we can see that there are more dark bands indicating high probabilities in certain off-diagonal regions as well.

Let $m$ and $p$ denote the original mood and pain scores respectively. The states are in the pairs of form $(m, p)$ where, $m, p \in \{1, 2, 3, 4, 5\}$. Since people tend not only to remain in the same state, but also to move a single step up or down in either mood or pain, it is interesting to extend the simple model of Eqn. (1) to capture these features. Let the probability that a person remains in the same state $(m, p)$ be $\pi_{m,p}$ and the probabilities that they move to a state with $p \pm 1$ be $\pi_{m,p \pm 1}$ and probability of moving to a state with $m \pm 1$ be $\pi_{m \pm 1,p}$.

Assuming independence holds, the model is re-defined using the following distribution. $P_{(m,p),(m',p')}$ is the probability of people moving from state $(m, p)$ on a day to $(m', p')$ the next day. For $m, p \in \{1, 2, 3, 4, 5\}$,
\[ P_{(m,p),(m',p')} = \begin{cases} 
\pi_{m,p} & \text{if } (m', p') = (m, p) \\
\pi_{m,p \pm 1} & \text{if } (m' = m) \text{ and } (p' = p \pm 1) \\
\pi_{m \pm 1, p} & \text{if } (m' = m \pm 1) \text{ and } (p' = p) \\
\text{uniform} & \text{otherwise} 
\end{cases} \] (S5)

The new model is thus that the probabilities for staying at the same state or moving to states whose mode or pain scores differ by 1 agree with those implicit, but transitions to all other states are equally likely. When we overlay a standard normal curve on the histogram of standardised residuals, as shown in Fig S9, we find once again that the residuals do not appear to be normally distributed.

Using the same standardised residual formula as in equation (3), the heatmap of the residuals shown in Fig S9 is obtained. Comparing it with Fig 1, the first noticeable difference is that the range of residuals for the new model has decreased, indicating a better fit. Also, the diagonal region connecting the top left to bottom right has smoothed out a bit.

We could in principle carry on, constructing models of increasing complexity and reducing the largest residuals until those that remain have the expected, near-normal distribution. But this modelling effort was only meant to be exploratory: our main goal was the clustering analysis as discussed before.

Fig S9A compares the expected values and residuals obtained from Fig S5, and Fig S9B shows how the normal distribution curve fits the histogram of residuals.

### S.1.6 Clusters

Transition probability matrix based on the regrouped states is given in Fig S10.

Before clustering, we take a look at the transition probability matrix again, but with the new states where we have regrouped the states into two categories, Good (G) and Bad (B) for mood, and Low (L) and High (H) for pain. We see trends similar to those in Fig 1A, where the probability to remain in any given state is high. Additionally, here we can also see that probability of moving from (Mood, Pain) state (B, L) to (G, L) is high.

Once clustering is done, in Fig S11 we note the distribution of transitions amongst the clusters. Here, the sum of probabilities for a particular transition across clusters add up to 1.

### S.1.7 Computing the shift for interventions

Here, we show how we have shifted the transition matrices to intervene with either improving mood or improving pain. This is a bespoke method to find a shift in the given scenario, designed to ensure that we follow the laws of probability. This need not be the optimal solution and in general it would be preferable to build a model to incorporate intervention effects by parameterisation of the state transitions as sets of odds ratios that could be calibrated to studies.

To our knowledge, building such a general model for Markov chains is an unsolved problem in discrete multivariate statistics and so we outline the problem-specific approach we adopt in this work below.

Step 1: We first calculate an intermediate result \( \alpha \) by taking the maximum of the maximum of the probabilities of transitioning from bad mood to good mood over the clusters:

\[ \alpha = \max \{ \Pr(\text{mood tomorrow} = G \mid \text{mood today} = B \text{ and } \text{cluster} = k) \} , 1 \leq k \leq K . \]
Step 2: Given $\alpha$, we calculate $\beta = (1 - \alpha)/2$. Then our new probabilities become

$$
\Pr'(\text{mood = good tomorrow} \mid \text{mood = bad today}) = \Pr(\text{mood = good tomorrow} \mid \text{mood = bad today}) + \beta.
$$

To ensure that the probabilities to add up to 1, we impose $\Pr'(\text{mood = bad} \& \text{pain = low} \mid \text{mood = bad}) = 0.8 \times (\Pr'(\text{mood = good} \mid \text{mood = bad})$ and $\Pr'(\text{mood = bad} \& \text{pain = high} \mid \text{mood = bad}) = 0.2 \times (\Pr'(\text{mood = good} \mid \text{mood = bad})$. We split in the ratio of 4:1 to give lesser importance to the least-ideal state BH. For our data, we get the approximate maximum value of $\beta$ as 0.15. So we shift by 0.15 and compare with the clusters. In a similar way, we calculated $\beta_P$. 
### S.2 Supplementary Tables

| Diagnosis                  | N (% approx.) |
|---------------------------|---------------|
| Chronic headache          | 1040 (10.4)   |
| Fibromyalgia               | 2668 (26.7)   |
| Gout                      | 340 (3.4)     |
| Neuropathic pain          | 1519 (15.2)   |
| Osteoarthritis            | 2283 (22.8)   |
| Rheumatoid arthritis      | 1838 (18.3)   |
| Spondyloarthropathy       | 865 (8.65)    |
| Unspecified arthritis     | 3418 (34.2)   |

Table S1: Conditions reported by the participants of the study.

| Site of pain       | N (% approx.) |
|--------------------|---------------|
| Head               | 1963 (19.6)   |
| Face               | 740 (7.4)     |
| Mouth or jaws      | 1606 (16)     |
| Neck or shoulder   | 5692 (57)     |
| Back               | 5910 (59.1)   |
| Stomach            | 1713 (17.1)   |
| Hip                | 5160 (51.6)   |
| Knee               | 6260 (62.6)   |
| Hands              | 5778 (57.8)   |
| Feet               | 4749 (47.5)   |

Table S2: Sites of chronic pain reported by the participants of the study.

| Sex     | Overall | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---------|---------|-----------|-----------|-----------|-----------|
| Female  | 47      | 46 (96)   | 51 (97)   | 47 (97)   | 46 (96)   |
| Male    | 52      | 50 (96)   | 56 (96)   | 53 (95)   | 50 (93)   |

Table S3: Rounded off values of mean age and response rate
| Cluster | Condition          | Log OR | Std. Error | CI low  | CI high |
|---------|--------------------|--------|------------|---------|---------|
| 1       | Cluster 1          | -0.04  | 0.07       | -0.17   | 0.10    |
| 2       | Cluster 2          | -0.15  | 0.07       | -0.29   | -0.00   |
| 3       | Cluster 3          | 0.16   | 0.07       | 0.03    | 0.29    |
| 4       | Cluster 4          | -0.01  | 0.05       | -0.11   | 0.10    |
| 5       | Cluster 1          | -0.08  | 0.06       | -0.20   | 0.04    |
| 6       | Cluster 2          | -0.25  | 0.07       | -0.38   | -0.12   |
| 7       | Cluster 3          | -0.30  | 0.06       | -0.41   | -0.18   |
| 8       | Cluster 4          | 0.40   | 0.05       | 0.30    | 0.49    |
| 9       | Cluster 1          | -0.34  | 0.08       | -0.51   | -0.18   |
| 10      | Cluster 2          | 0.47   | 0.12       | 0.25    | 0.70    |
| 11      | Cluster 3          | 0.03   | 0.09       | -0.14   | 0.20    |
| 12      | Cluster 4          | -0.00  | 0.07       | -0.15   | 0.14    |
| 13      | Cluster 1          | 0.02   | 0.14       | -0.26   | 0.30    |
| 14      | Cluster 2          | 0.13   | 0.16       | -0.19   | 0.44    |
| 15      | Cluster 3          | 0.01   | 0.14       | -0.26   | 0.28    |
| 16      | Cluster 4          | -0.08  | 0.11       | -0.30   | 0.14    |
| 17      | Cluster 1          | 0.22   | 0.06       | 0.11    | 0.33    |
| 18      | Cluster 2          | -0.26  | 0.06       | -0.38   | -0.14   |
| 19      | Cluster 3          | 0.07   | 0.05       | -0.04   | 0.17    |
| 20      | Cluster 4          | -0.04  | 0.04       | -0.13   | 0.05    |
| 21      | Cluster 1          | -0.97  | 0.06       | -1.08   | -0.85   |
| 22      | Cluster 2          | 1.64   | 0.10       | 1.45    | 1.84    |
| 23      | Cluster 3          | -0.41  | 0.06       | -0.52   | -0.31   |
| 24      | Cluster 4          | 0.32   | 0.05       | 0.23    | 0.41    |
| 25      | Cluster 1          | -0.46  | 0.08       | -0.61   | -0.31   |
| 26      | Cluster 2          | 0.59   | 0.11       | 0.37    | 0.81    |
| 27      | Cluster 3          | 0.27   | 0.09       | 0.10    | 0.44    |
| 28      | Cluster 4          | -0.11  | 0.07       | -0.24   | 0.02    |
| 29      | Cluster 1          | -0.72  | 0.06       | -0.84   | -0.59   |
| 30      | Cluster 2          | 1.08   | 0.11       | 0.87    | 1.29    |
| 31      | Cluster 3          | -0.23  | 0.07       | -0.36   | -0.09   |
| 32      | Cluster 4          | 0.24   | 0.06       | 0.13    | 0.35    |

Table S4: 8618 out of 9990 participants of the study, reported their chronic pain condition. Log odds ratio of a condition in a cluster with 95% Confidence Interval
| Cluster | Condition                | Log OR | Std. Error | CI low  | CI high |
|---------|--------------------------|--------|------------|---------|---------|
| 1       | Cluster 1 Head           | -0.71  | 0.06       | -0.82   | -0.59   |
| 2       | Cluster 2 Head           | 0.75   | 0.08       | 0.58    | 0.91    |
| 3       | Cluster 3 Head           | 0.06   | 0.06       | -0.07   | 0.18    |
| 4       | Cluster 4 Head           | 0.08   | 0.05       | -0.02   | 0.18    |
| 5       | Cluster 1 Face           | -0.80  | 0.09       | -0.96   | -0.63   |
| 6       | Cluster 2 Face           | 1.16   | 0.16       | 0.85    | 1.47    |
| 7       | Cluster 3 Face           | -0.23  | 0.09       | -0.41   | -0.05   |
| 8       | Cluster 4 Face           | 0.28   | 0.08       | 0.13    | 0.44    |
| 9       | Cluster 1 Mouth or jaws  | -0.59  | 0.07       | -0.72   | -0.46   |
| 10      | Cluster 2 Mouth or jaws  | 0.96   | 0.10       | 0.77    | 1.16    |
| 11      | Cluster 3 Mouth or jaws  | -0.21  | 0.07       | -0.34   | -0.08   |
| 12      | Cluster 4 Mouth or jaws  | 0.12   | 0.06       | 0.01    | 0.22    |
| 13      | Cluster 1 Neck or shoulder | -0.62  | 0.06       | -0.74   | -0.50   |
| 14      | Cluster 2 Neck or shoulder | 0.53  | 0.06       | 0.42    | 0.65    |
| 15      | Cluster 3 Neck or shoulder | -0.28  | 0.06       | -0.39   | -0.17   |
| 16      | Cluster 4 Neck or shoulder | 0.22  | 0.04       | 0.13    | 0.30    |
| 17      | Cluster 1 Back           | -0.91  | 0.07       | -1.04   | -0.78   |
| 18      | Cluster 2 Back           | 0.81   | 0.06       | 0.70    | 0.93    |
| 19      | Cluster 3 Back           | -0.39  | 0.06       | -0.50   | -0.28   |
| 20      | Cluster 4 Back           | 0.25   | 0.04       | 0.16    | 0.33    |
| 21      | Cluster 1 Stomach        | -0.69  | 0.06       | -0.82   | -0.57   |
| 22      | Cluster 2 Stomach        | 1.06   | 0.10       | 0.87    | 1.26    |
| 23      | Cluster 3 Stomach        | -0.05  | 0.07       | -0.18   | 0.08    |
| 24      | Cluster 4 Stomach        | 0.04   | 0.05       | -0.06   | 0.15    |
| 25      | Cluster 1 Hip            | -0.60  | 0.06       | -0.71   | -0.48   |
| 26      | Cluster 2 Hip            | 0.60   | 0.06       | 0.48    | 0.71    |
| 27      | Cluster 3 Hip            | -0.34  | 0.05       | -0.44   | -0.23   |
| 28      | Cluster 4 Hip            | 0.22   | 0.04       | 0.14    | 0.31    |
| 29      | Cluster 1 Knee           | -0.39  | 0.06       | -0.51   | -0.26   |
| 30      | Cluster 2 Knee           | 0.39   | 0.06       | 0.28    | 0.51    |
| 31      | Cluster 3 Knee           | -0.38  | 0.06       | -0.49   | -0.26   |
| 32      | Cluster 4 Knee           | 0.22   | 0.05       | 0.14    | 0.31    |
| 33      | Cluster 1 Hands          | -0.32  | 0.06       | -0.44   | -0.21   |
| 34      | Cluster 2 Hands          | 0.20   | 0.06       | 0.09    | 0.32    |
| 35      | Cluster 3 Hands          | -0.19  | 0.06       | -0.30   | -0.09   |
| 36      | Cluster 4 Hands          | 0.19   | 0.04       | 0.11    | 0.28    |
| 37      | Cluster 1 Feet           | -0.43  | 0.06       | -0.54   | -0.32   |
| 38      | Cluster 2 Feet           | 0.40   | 0.06       | 0.28    | 0.51    |
| 39      | Cluster 3 Feet           | -0.28  | 0.05       | -0.38   | -0.17   |
| 40      | Cluster 4 Feet           | 0.21   | 0.04       | 0.13    | 0.29    |

Table S5: 9146 out of 9990 participants of the study reported their site of pain. Log odds ratio of a site of pain in a cluster with 95% Confidence Interval
S.3 Supplementary Figures

Figure S1: Frequency of states

Figure S2: Overall age distribution, included for comparison with Fig 3B.
Figure S3: A and B indicate the proportion of participants in a cluster reporting, respectively, a given condition and site of pain. C and D show the proportions of participants with, respectively, a given condition or site of pain who fall into each cluster.
Figure S4: Negative Log Likelihood as a function of the number of components

Figure S5: Difference of negative Log-Likelihoods of models with number of clusters k+1 and k
Figure S6: Dotted lines represent negative Log Likelihood gradients extrapolated from the difference between clusters 1 and 2, and clusters 9 and 10.

Figure S7: Bayesian Inference Criterion (BIC) per model
Figure S8: Residual heatmap of the model given by Eqn. (S5)
Figure S9: **A** is the scatter plot of expected values and the residuals. **B** shows a histogram of the residuals as well as a blue curve giving the probability density function of a normal distribution having the same mean and variance as the residuals.

Figure S10: Transition probability matrix based on the regrouped scales

Figure S11: The ratio of the entries in the transition probability matrices for the clusters to the transition probabilities estimated from the whole sample without clustering.
Figure S12: Four clusters without regrouping (Mood, Pain) states

Figure S13: Heatmaps of transition probability matrices when number of clusters is 5
Figure S14: Heatmaps of transition probability matrices when number of clusters is 6

Figure S15: Heatmaps of transition probability matrices when number of clusters is 7
Figure S16: Heatmaps of transition probability matrices when number of clusters is 8