Prediction of Treatment Outcomes from Multimodal Pain Rehabilitation

Rode Grönkvist

University of Gothenburg

Department of Mathematical Sciences

September 2022

Abstract

Chronic pain is a cause of suffering in a large share of the population and a leading public health problem. Multimodal pain rehabilitation (MMR) is a multidisciplinary rehabilitation method commonly used to treat chronic pain. In Sweden, MMR providers contribute patient data to a national registry (the SQRP). A Bayesian multivariate linear regression model was fitted on SQRP patient data from 2009-2016 (n=8168) targetting patient health change from before to after treatment as responses. The psychometric instruments MPI, SF-36, HADS and NRS were used to measure health, with summarized versions of MPI and SF-36 being utilized. Several significant effects were found, principally in that unemployment, low patient belief in recovery and having constant, as opposed to periodic, pain had significant negative effects on several responses. Additionally, the results of a previous study on dimension reduction of the MPI instrument were replicated.

Acknowledgements

I would first and foremost like to thank my supervisors for their invaluable assistance: Petter Mostad, for his guidance and patience regarding the mathematical and statistical, and Anna Grimby Ekman, who described the project idea and gave guidance regarding chronic pain and MMR. I would also like to thank Helene Svensdotter, whose data pre-processing was very useful, and who gave insight into MMR from the clinician's perspective; as did Isak Barvesten, who I also thank. I thank Eva Petersson and Barbro Strååt who provided insight into chronic pain and MMR from the patient's perspective. Finally, I thank my friends and family for their help and support.

Contents

1 Introduction

Chronic pain is a very common illness and a public health challenge. Estimates show that approximately 20% of adults live with at least moderate chronic pain [Dah+18], [Bre+06] with negative impacts on a wide range of health aspects, including sleep, brain function, mood and mental health, cardiovascular health, sexual function and overall quality of life [Fin11]. Lower back pain, which is a common form of chronic pain, is the leading cause of years lived with disability worldwide [Wu+20].

Multimodal pain rehabilitation (MMR) is a treatment method that combines physiotherapy, psychiatry and medicine in a rehabilitation program. There are no universal criteria for what constitutes MMR, but common components are education on pain, coping skills, physical activity and training [Spi+18]. MMR is consistent with the widely accepted biopsychosocial approach to pain, which is the view that chronic pain is influenced by both biological, psychological and social factors [Gat+07], and aims to target several of these factors for improvement via multidisciplinary interventions.

While scientific evaluation of MMR points to it giving better long-term improvement in many pain patient categories than other treatment methods [Sca+08], the results vary between patients, and little research has been conducted on what causes differing outcomes. The studies that exist often have small sample sizes or focus only on patients with specific pain disorders [Roo+13], [Ang+10]. As MMR is a costly form of intervention with limited availability, the ability to better predict which patients will experience the greatest benefits from MMR would be useful in prioritizing patients for treatment. In accordance with the biopsychosocial approach to pain, it is important to account for outcomes in physical, social and psychological health dimensions.

In Sweden, the Swedish Quality Registry for Pain Rehabilitation (SQRP) gathers patient data from MMR treatment centers concerning background and socioeconomic factors, values for several psychometric instruments at baseline and values for these instruments after treatment. The present study is a registry study based on data from the SQRP that aims to identify significant predictors of several treatment outcome dimensions, and to give estimates of effect sizes for these predictors.

The present study fits a Bayesian multivariate linear regression model on selected regressors and responses with the goal of describing multifaceted health outcomes of MMR treatment in chronic pain patients. The Bayesian framework allows explicit accounting for model uncertainty which is important to quantify certainty in predictions of treatment outcomes. The multivariate linear regression model leverages the fact that outcomes in different health domains correlate to produce stronger predictions than univariate models would. The posterior distribution of effect parameters are described, including credible intervals, measures of effect size and significance, and posterior predictive simulations for some example patients are given, demon-

strating the usefulness of the model in predicting treatment outcomes. The Bayesian multivariate approach and the choices of regressors and responses distinguish this model from a similar study conducted on SQRP data with the aim of finding predictors for treatment outcomes in chronic pain patients [Ger+16].

Health outcomes in chronic pain patients are not trivial to measure and describe, and thus psychometric instruments are commonly utilized to measure health states. One of the instruments that is used in the SQRP, the Multidimensional Pain Inventory (MPI), has been criticized for having some undesirable properties. One proposed solution is to reduce the dimension of the instrument by transforming the original 12 scales into 3 summary scales [MN11]. The present study replicates the factor analysis results of [MN11] and utilizes the summarized MPI scales, among other instruments, as outcome measures.

The present study is limited by only utilizing registry data and thus not having a control group. Still, the findings can contribute to a better understanding of predictors for treatment outcomes from multimodal pain rehabilitation.

2 Theory

2.1 Chronic Pain and Multimodal Pain Rehabilitation

Pain is defined by the International Association for the Study of Pain (IASP) as "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" [Raj+20]. This definition is to be adopted in the International Statistical Classification of Diseases and Related Health Problems 11th revision (ICD-11), which is not yet in use in Sweden [WHO21]. Chronic pain is defined in the ICD-11 as pain that persists or recurs for longer than three months [Nic+19]. The ICD-11 is used to classify medical conditions both for clinical purposes and to give a unified standard of categorization for statistical purposes.

Many medical conditions cause pain. Rheumatic disorders and migraines or other headache disorders are common medical conditions the symptoms of which include chronic pain. However, chronic pain can persist despite successful treatment of the condition that initially caused it, or itself become a larger detriment to health than any other condition. Chronic pain can also manifest without any other known condition causing it. Therefore, chronic pain can be a patient's main medical issue, and patients can need treatment specifically for chronic pain. For these reasons, classification of chronic pain has moved away from diagnosis oriented towards some underlying disease towards viewing chronic pain as a condition of its own [Nic+19].

Chronic pain is understood through a biopsychosocial framework. Biological factors such as injury or disease are clearly common causes of pain, both acute and chronic. Psychological factors, such as catastrophizing or ruminating thought patterns, or avoidant behavior, can exacerbate chronic pain, as can social factors, such as lack of social support or economic stress. Treatment methods that target only one or some of these factors are less effective than multimodal treatment [SBU21].

Multimodal pain rehabilitation (MMR) is an interdisciplinary treatment method, combining physiotherapy, psychiatry and medicine. The main aim of MMR is not to remove pain completely, as pain is a part of normal human functioning, but rather to reduce the negative impact of pain on patients' lives. Reduction of irrelevant pain or decreased pain sensitivity would be valuable to patients, but as this is difficult to achieve in isolation, MMR takes a wider approach. In Sweden, MMR is only provided to a patient after referral and evaluation. Due to limited availability of treatment and high requirements on patient engagement and activity to reach a positive outcome, MMR providers perform a selection of patients to receive treatment after an initial evaluation.

2.2 The Swedish Quality Registry for Pain Rehabilitation

The Swedish Quality Registry for Pain Rehabilitation (SQRP) is a national database that gathers patient data from providers of multimodal pain rehabilitation in Sweden. The SQRP was started in 1998 by the Swedish Association for Rehabilitative Medicine, a specialist organization within the Swedish Medical Association, with the aim of describing patients in need of pain rehabilitation and treatment outcomes of such rehabilitation. Data for the registry is supplied by pain rehabilitation teams and units connected to the registry - as of 2021, every specialist provider of pain rehabilitation in Sweden was connected to the registry, as well as 17 primary care pain units. The registry holders estimate that during 2021, 85% of patients receiving MMR in Sweden were entered into the registry [NRS22]. Yearly reports from the period 2009-2016 do not provide such coverage estimates, but similar amounts of patients were registered yearly during that period [NRS13].

Inclusion of patient data in the SQRP is based on receiving a referral to specialist pain treatment. Not every patient that receives such a referral will be provided treatment, but every patient that undergoes an evaluation for treatment with a healthcare provider connected to the SQRP will fill in a form that is submitted to the SQRP.

Each entry in the SQRP consists of the data gathered for one person during one process of referral, evaluation and treatment. This means that an entry might only contain the data gathered during evaluation, if the patient was not offered treatment after evaluation. There might also be several entries for the same person, if a patient receives multiple referrals to specialist pain treatment at different points in time. This is not uncommon for sufferers of chronic pain.

The SQRP data is gathered at three points in time for each person - an initial form is gathered as part of the patient's evaluation for treatment, a concluding form is gathered at the time of the conclusion of treatment, and a follow-up form is gathered 12 months after the conclusion of treatment. All three forms contain the same psychometric instruments, whereas the first form contains most background questions and the conclusion and follow-up forms contain some evaluative questions.

The SQRP does not contain follow-up data for patients who did not undergo treatment. This means that there is no natural way to compare health outcomes of patients receiving treatment with chronic pain sufferers who did not receive treatment.

The evaluation form contains many background variables, such as age, gender, region of origin, education, employment status, self-reported pain duration, self-reported pain severity and pain locations, and self-reported expectation of improvement. The form also contains several psychometric instruments that are described in detail in the next section. This form can vary somewhat between healthcare providers and has changed over time, but a large set of background variables and psychometric instruments are the same. An extensive description of the variables included in the SQRP dataset is available in Appendix A.

The SQRP underwent a revision in 2016, which has caused data to be divided into two epochs with some variety in the gathered data. The first epoch consists of data from the period January 2009 to July 2016, and the second of data from

August of 2016 onward. The present study was performed using data from the first epoch.

The SQRP data contains a multitude of psychometric instruments measuring different outcomes. Some are mandatory inclusions in evaluation forms for all healthcare providers connected to the SQRP, while some are voluntary and not used by a majority of treatment locations. In addition to the major instruments, the SQRP contains an NRS (numeric rating scale) for pain during the past week, where patients report their pain on a scale of 0-10; and a pain location indicator instrument, where patients can report in what parts of their body they experience pain. The following list describes the major instruments briefly, focusing on the mandatory instruments.

- HADS - Hospital Anxiety and Depression Scale. Used to indicate symptoms of anxiety and depression. Contains two subscales, one for anxiety and one for depression. Scores range from 0 to 21, with 21 indicating greatest risk of anxiety or depression. [ZS83]
- SF36 - 36-item Short-Form Health Survey. Measures quality of life on eight dimensions: Physical functioning, Role functioning/physical, Role functioning/emotional, Energy/fatigue, Emotional well-being, Social functioning, Pain, and General health. Scores range from 0 to 100 with 100 indicating the best health. [WS92]
- EQ-5D - EuroQoL 5 dimensions. Measures quality of life on five dimensions: Mobility, Self-care, Usual activities, Pain/discomfort, and Anxiety/depression. [RC01]
- MPI - Multidimensional pain inventory. Measures pain and how pain impacts the patient's life on many dimensions, including physical, mental, social and activity-related. Designed with 12 dimensions in three domains: the physical/mental domain with dimensions Interference, Social support, Pain severity, Perceived life control and Affective distress; the social domain, with dimensions Solicitous responses from significant other, Distracting responses and Negative responses; and the activity domain, with dimensions Household Chores, Outdoor work, Activities away from home and Social activities. A system of patient classification into three categories (adaptive coper, interpersonally distressed, dysfunctional) has been developed by the developers of the instrument and the category assignment probabilities are given in SQRP data. Scores range from 0 to 6. [KTR85]
- Voluntary instruments are CPAQ - Chronic Pain Acceptance Questionnaire, measuring pain acceptance; Tampa Scale of Kinesiophobia, measuring kinesiophobia, i.e. fear of movement; LiSat-11 - Life Satisfaction Questionnaire with 11 items, measuring life satisfaction.

2.3 Mathematical and statistical methods

Some mathematical and statistical methods are used to evaluate data, investigate the structure of correlation between the instruments, and propose and evaluate predictive models.

2.3.1 Psychometrics and factor analysis

The aim of psychometrics is to measure psychological properties of humans, such as emotional well-being or level of extraversion. Unlike physical properties like height or weight, there are no simple means of measuring such psychological properties, and there might even be disagreement on their definition. Properties that cannot be measured directly are called *latent* properties or *constructs*. Despite the difficulty in constructing measures that approximate latent properties and that allow comparison or mathematical manipulation, such measures may be of interest to researchers, for instance when trying to determine whether treatments for psychological problems are effective. Therefore, an *instrument* such as a test or questionnaire is created based on relevant theory and expertise in such a way that one can assume that a person's result on the instrument is related to the latent property in that person. For instance, if one wishes to evaluate patients on their level of anxiety, which is clearly a latent property, one might construct a questionnaire on the extent to which they experience symptoms of anxiety, with the assumption that patients with a greater level of anxiety will experience more symptoms thereof.

Creating a useful psychometric instrument is a technical process where many properties of the instrument must be ensured - an introduction to measurement theory is available in e.g. [DB18]. The present study does not focus on measurement theory, and will generally not go into detail regarding the creation and verification of the instruments used in the SQRP. The exception is the question of *construct validity*, that is, whether the instrument actually measures what it purports to measure. Several instruments in the SQRP contain many dimensions that are supposed to reflect different latent properties, in order to capture the complexity of pain conditions. However, it is not certain that different dimensions of some instrument actually measure independent and distinct latent properties. One way of investigating construct validity in a multidimensional instrument is to use *factor analysis*.

Factor analysis aims to find structures in data that might not correspond to the dimensions that are assumed in the design of the instrument. Given a set of responses to some instrument, one can consider whether the responses seem to correspond with the theory behind the instrument, in the sense that the dimensions are independent and measure different latent properties. One can also consider the question of whether the set of dimensions that are postulated in the instrument best explain the variability that is observed in the data, or if some other set of dimensions would better describe this variability. These questions can be attacked mathematically using *principal component analysis* (PCA). The following exposition on PCA follows [Jol02]. Even though the concepts and goals of factor analysis and PCA are closely related, there are differences in the theoretical assumptions that are described well in chapter 7 of [Jol02]. As a consequence of the close relation of concepts, the terms *principal component* and *factor* will sometimes be used interchangeably, specifically in the context of factor rotation.

PCA is a mathematical method that finds a basis for some data such that the axes of the basis are orthogonal and the first axis lies along the axis of maximal variance in the data, the second axis lies along the axis of maximal remaining variance orthogonal to the first axis, and so on. Formally, let $\mathbf{x}$ be a vector of p random variables and $\mathbf{\Sigma}$ be the covariance matrix of $\mathbf{x}$ (for this derivation $\mathbf{\Sigma}$ is assumed to be known, but for data with unknown covariance structure the derivation is similar to the one below with $\mathbf{\Sigma}$ replaced by the sample covariance matrix). Letting $\boldsymbol{\alpha}^t$

denote the transpose of $\boldsymbol{\alpha}$, the goal is to find some set of functions $\boldsymbol{\alpha}_i^t\mathbf{x}$, $i \in \{1, ..., p\}$ such that

$$\boldsymbol{\alpha}_i^t\mathbf{x} = \sum_{k=1}^{p} \alpha_{ik}x_k.$$

The $\boldsymbol{\alpha}_i$ should be such that $\boldsymbol{\alpha}_1^t\mathbf{x}$ has maximal variance, $\boldsymbol{\alpha}_2^t\mathbf{x}$ is uncorrelated with $\boldsymbol{\alpha}_1^t\mathbf{x}$ and has maximal variance, and so forth, such that $\boldsymbol{\alpha}_k^t\mathbf{x}$ has maximal variance and is uncorrelated with $\boldsymbol{\alpha}_1^t\mathbf{x}$,..., $\boldsymbol{\alpha}_{k-1}^t\mathbf{x}$. Each $\boldsymbol{\alpha}_i^t\mathbf{x}$ is a new random variable, and the kth such random variable, $\boldsymbol{\alpha}_k^t\mathbf{x}$, is called the kth principal component. To obtain the vectors $\boldsymbol{\alpha}_i^t$, the following procedure is used: consider first $\boldsymbol{\alpha}_1^t$, which maximizes $var(\boldsymbol{\alpha}_1^t\mathbf{x}) = \boldsymbol{\alpha}_1^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_1$. Since $\boldsymbol{\Sigma}$ is positive semi-definite this maximum will not be attained for finite $\boldsymbol{\alpha}_1$, so impose the constraint $\boldsymbol{\alpha}_1^t\boldsymbol{\alpha}_1 = 1$. To maximize $\boldsymbol{\alpha}_1^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_1$ subject to this constraint, the method of Lagrange multipliers is used. Maximize instead

$$\boldsymbol{\alpha}_1^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda(\boldsymbol{\alpha}_1^t\boldsymbol{\alpha}_1 - 1)$$

where λ is a Lagrange multiplier. Differentiating this expression with respect to $\boldsymbol{\alpha}_1$ and setting it equal to zero gives

$$\begin{aligned} \boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda\boldsymbol{\alpha}_1 = 0 \Leftrightarrow \\ (\boldsymbol{\Sigma} - \lambda\mathbf{I}_p)\boldsymbol{\alpha}_1 = 0 \end{aligned}$$

meaning λ is an eigenvalue of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}_1$ is the corresponding eigenvector. To determine which eigenvector of $\boldsymbol{\Sigma}$ gives the maximum variance, note that the quantity to be maximized is

$$\boldsymbol{\alpha}_1^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1^t\lambda\boldsymbol{\alpha}_1 = \lambda$$

so λ must be as large as possible, meaning $\boldsymbol{\alpha}_1$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Sigma}$. Note also from this equality that the variance of $\boldsymbol{\alpha}_1^t$ is equal to the largest eigenvalue of $\boldsymbol{\Sigma}$. In general, the kth principal component is $\boldsymbol{\alpha}_k^t\mathbf{x}$ and its variance is λ_k, the kth largest eigenvalue of $\boldsymbol{\Sigma}$. This is proved here for $k = 2$, the proof for $k \geq 3$ proceeds similarly. To find $\boldsymbol{\alpha}_2$, a maximization of $\boldsymbol{\alpha}_2^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_2$ is performed as before, but with the added constraint that $cov(\boldsymbol{\alpha}_1^t\mathbf{x}, \boldsymbol{\alpha}_2^t\mathbf{x}) = 0$. This condition can be rephrased in several ways, as

$$cov(\boldsymbol{\alpha}_1^t\mathbf{x}, \boldsymbol{\alpha}_2^t\mathbf{x}) = \boldsymbol{\alpha}_1^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_2^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2^t\lambda_1\boldsymbol{\alpha}_1.$$

Thus, any of the following equivalent conditions could be added to the maximization problem:

$$\begin{aligned} \boldsymbol{\alpha}_1^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 = 0 \qquad & \qquad \boldsymbol{\alpha}_2^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = 0 \\ \boldsymbol{\alpha}_1^t\boldsymbol{\alpha}_2 = 0 \qquad & \qquad \boldsymbol{\alpha}_2^t\boldsymbol{\alpha}_1 = 0 \end{aligned}$$

Choosing the last one of these, the new maximization problem becomes

$$\boldsymbol{\alpha}_2^t\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda(\boldsymbol{\alpha}_2^t\boldsymbol{\alpha}_2 - 1) - \phi(\boldsymbol{\alpha}_2^t\boldsymbol{\alpha}_1)$$

where λ and ϕ are Lagrange multipliers. Differentiating this with respect to $\boldsymbol{\alpha}_2$ and setting to zero gives

$$\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda\boldsymbol{\alpha}_2 - \phi\boldsymbol{\alpha}_1 = 0$$

which, when multiplied by $\boldsymbol{\alpha}_1^t$ on the right gives

$$\boldsymbol{\alpha}_1^t \boldsymbol{\Sigma} \boldsymbol{\alpha}_2 - \lambda \boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2 - \phi \boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_1 = 0.$$

Here, the first two terms are 0 by the assumed constraints and $\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_1 = 1$, so $\phi = 0$. Therefore, the maximization problem reduces to

$$\begin{aligned} \boldsymbol{\Sigma} \boldsymbol{\alpha}_2 - \lambda \boldsymbol{\alpha}_2 = 0 \Leftrightarrow \\ (\boldsymbol{\Sigma} - \lambda \mathbf{I}_p) \boldsymbol{\alpha}_2 = 0 \end{aligned}$$

which means that λ is an eigenvalue of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}_2$ is the corresponding eigenvector. Since $\boldsymbol{\alpha}_2^t \boldsymbol{\Sigma} \boldsymbol{\alpha}_2$ is to be maximized, λ is the greatest eigenvalue of $\boldsymbol{\Sigma}$ subject to the constraints. Note that unless $\boldsymbol{\Sigma}$ has repeated eigenvalues, $\lambda \neq \lambda_1$ since that would imply $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$, violating the constraint $\boldsymbol{\alpha}_1^t \boldsymbol{\alpha}_2 = 0$. Thus, λ is the second largest eigenvalue of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}_2$ is the corresponding eigenvector. As stated above, one can proceed similarly to derive further principal components $\boldsymbol{\alpha}_3^t \mathbf{x}, ..., \boldsymbol{\alpha}_p^t \mathbf{x}$, where the kth principal component will be the eigenvector of $\boldsymbol{\Sigma}$ corresponding to its kth eigenvalue λ_k, with variance $var(\boldsymbol{\alpha}_k^t \mathbf{x}) = \lambda_k$.

Given the above, principal component analysis reduces to finding the eigenvectors of the (sample) covariance matrix of the random variables of interest. Therefore, one can use the relative sizes of the eigenvalues to determine the relative amounts of the total variance that is described by the corresponding eigenvectors. The principal components ordered by size of eigenvalues are then an orthogonal basis for data where the first basis component describes the largest possible amount of variance in the data, followed by the second, and so on.

PCA is a powerful tool for dimension reduction and assessing construct validity. If some multidimensional psychometric instrument is well-designed in the sense that every dimension measures a unique construct that is uncorrelated with the other dimensions, then no significant dimension reduction should be possible and all principal components should describe similar amounts of variance. If, however, several dimensions are correlated, a few principal components will describe a much larger share of the variance, and dimension reduction will be possible.

In practical applications of PCA, it is common to normalize all data dimensions before finding the eigenvectors of the sample covariance matrix. Otherwise, dimensions with greater range will dominate over dimensions with smaller range.

There is no universal metric to assess what constitutes a "large share" of total variance, but PCA is often evaluated by a scree plot, a plot of the eigenvalues of all principal components in descending order, where a clear elbow can indicate that principal components above the elbow are significant and principal components below it can be discarded. Such a plot can be seen in Figure 4.1. Another simple evaluation tool is to consider principal components with eigenvalues greater than 1 to contribute a large share of variance.

For the purposes of interpretability, it can be of interest to rotate principal components. Once a subset of principal components (or factors) that explains a sufficiently large proportion of the variance has been selected, those factors can be rotated to provide more interpretable loadings of original variables onto dimensions, without affecting the amount of variance they explain as a basis. Usually, the goal of such rotation is to make the loadings of the original variables onto the factors either as close to or as far from zero as possible. Clearly, a factor that is just an

average of two well-understood dimensions is more interpretable than a factor that consists of a small part of every dimensions in a multivariate model.

Factor rotations are either *orthogonal* or *oblique*. An orthogonal rotation preserves the orthogonality of the factors, whereas an oblique rotation allows for non-orthogonal factors. Depending on what assumptions on orthogonality of the latent variables are made, either kind of rotation can be useful. There are many different rotation procedures of either kind, based on different simplicity criteria. For the purposes of this study, an oblique rotation known as *Promin* is used, see [Lor99]. The Promin rotation first obtains a *simple* target matrix using a weighted Varimax rotation, meaning a rotated structure where each variable loads strongly onto very few factors and have zero loadings onto the other factors, and where each factor is loaded onto strongly by a few variables and have zero loadings from the other variables. Next, the distance between this target and an oblique rotation of the factors is minimized using a Procrustes rotation. The details of the rotation and references to the rotations used in Promin are given in [Lor99].

A construct validation procedure for a high-dimensional psychometric instrument might thus be as follows. First, consider a set of measurements using some multivariate instrument on some reasonable set of people. Then, perform PCA on the measurements and reduce dimensionality if appropriate. Next, rotate the remaining principal components to increase interpretability. Finally, investigate how the result correlates with the instrument's design and theoretical knowledge in the academic field the measurement is concerned with. It must be noted that PCA does not assume or give rise to a model, but is merely a tool that describes properties of the data - thus, a set of rotated principal components is not a replacement for a theoretically sound model of some latent properties. However, if the data do not correspond with the model, the model might be insufficient.

2.3.2 Bayesian inference

Bayesian statistical methods utilize prior knowledge about some phenomenon and combine it with some evidence to update that knowledge. Bayesian inference is useful when estimating parameters in some complicated model where the exact distributions of the parameters are not known. Bayesian methods also allow more precise quantification of the level of certainty in the obtained knowledge about the studied object. In the present study, Bayesian methods are used to infer estimates of regression coefficients and error distributions.

Bayesian inference rests on Bayes' theorem. Given some model with a set of parameters θ and some data $\mathbf{X}$ and letting π indicate probability, Bayes' theorem states that

$$\pi(\theta|\mathbf{X}) = \frac{\pi(\mathbf{X}|\theta)\pi(\theta)}{\pi(\mathbf{X})}.$$

In words, the probability of the model conditional on the data (also known as the *posterior probability* of the model) is equal to the probability of the data conditional on the model (also known as the *likelihood* or *evidence*) multiplied by the unconditional probability of the model (also known as the *prior probability* of the model) divided by the unconditional probability of the data. This theorem allows model and parameter inference. By assuming that the likelihood follows some distribution (e.g. multivariate normal with parameters provided by the model) inference can be

made on the distributions of model parameters. For a linear regression problem, these model parameters are regression coefficients and the distribution of the error.

This indicates the main philosophical difference between classical (or frequentist) inference and Bayesian inference. In classical inference, the parameters one wishes to infer are assumed to be fixed but unknown. In Bayesian inference, model parameters are assumed to be uncertain, which allows the notion of a probability distribution for the model parameters. Another large difference lies in the use of a prior distribution in the Bayesian setting. In classical inference, any parameter estimates are derived solely from the data, given some chosen model. In the Bayesian setting, since the posterior distribution of a parameter is the product of the prior distribution and the likelihood, any prior knowledge of the parameters can be explicitly incorporated into the model and will affect the posterior distribution. If there is no prior knowledge of the parameter distributions, or no wish to incorporate such knowledge into the model, an *uninformative prior* can be used. This aims to minimize the role of the prior distribution and "let the data speak for themselves" [Gel13]. For example, an uninformative prior for a location parameter might be the uniform distribution over some wide interval. One might also assign equal probability to any real value for a location parameter. This results in an *improper prior*, meaning it is not actually a probability distribution as it integrates to ∞ rather than 1. Certain improper priors can yield proper posterior distributions, making them useful as a way to express minimal prior knowledge.

An important concept relating to the choice of a prior distribution is the notion of *conjugate distributions*. This means that the prior and posterior distributions are in the same probability distribution family, giving relatively simple closed-form expressions for the posterior distribution. This is desirable to improve interpretability of results and to avoid posterior that can only be described by numerical integration. The existence of conjugate prior and posterior distributions depends on the distribution of the likelihood, as the posterior is the product of the prior and the likelihood.

In classical linear regression, the p-value of regression coefficients under a t-test is commonly used to conclude which regressor variables are of importance. In the Bayesian framework, there is no exact translation of the concept of the p-value. Instead, to concisely evaluate certainty of effects, the *probability of direction* (PD) and effect size estimates will be used in the present study. PD is defined as the share of the posterior distribution of some parameter that shares a sign with the median, and indicates the certainty associated with the most probable direction (positive or negative) of the effect. If the entire posterior distribution of some parameter lies above 0, PD will be 1, representing total certainty that the effect is positive. If, on the contrary, the posterior distribution of some parameter is symmetric around 0, PD will be 0.5, representing total uncertainty in the directionality of the effect. PD will vary between 0.5 and 1, and is closely correlated to the notion of p-value [Mak+19]. An approximate translation between the two can be made by taking $p_{two-sided} = 2 * (1 - PD)$, meaning that a two-sided p-value of 0.05 corresponds approximately to a PD of 0.975.

Bayesian inference allows the notion of a *credible interval*, which is similar to the classical notion of a confidence interval, albeit with a different mathematical justification and interpretation. In the classical framework, since parameters are viewed as fixed but unknown and estimators are stochastic, the interpretation of a

95% confidence interval for some parameter θ is that it is an interval whose bounds L, U are random variables such that the probability

$$\pi(L < \theta < U) = 0.95.$$

This means that if one repeatedly generates confidence intervals for θ, 95% of these confidence intervals are expected to contain the true value of θ. In the Bayesian framework, since parameters are viewed as uncertain, the notion of a confidence interval does not work, and credible intervals are used instead. In Bayesian inference, a posterior distribution for a parameter is generated using the procedure described above. Given this posterior distribution, a 95% credible interval is an interval that covers 95% of the posterior distribution, meaning there is a 95% probability that the parameter falls within the credible interval. For any posterior distribution, a 95% credible interval can be selected in different ways, some of the most common being the highest density interval, meaning the interval such that any point within the interval has higher probability density than every point outside the interval (this is also the narrowest possible interval), and the equal-tailed interval, meaning the interval such that the probability of the parameter being greater than the upper bound of the interval is equal to the probability of the parameter being smaller than the lower bound. For unimodal and symmetric posterior distributions, these intervals coincide.

The Bayesian framework allows the computation of a *posterior predictive distribution* given that a posterior distribution for some model parameters has been computed. The posterior predictive distribution is the estimated distribution for an unobserved data point, taking into account the level of uncertainty in the posterior distribution. Formally, letting θ denote the set of parameters in the model (belonging to some parameter space Θ), $\mathbf{X}$ denote the data and $\tilde{x}$ an unobserved data point, the posterior predictive distribution of $\tilde{x}$ is

$$\pi(\tilde{x}|\mathbf{X}) = \int_{\Theta} \pi(\tilde{x}|\theta, \mathbf{X})\pi(\theta|\mathbf{X})d\theta.$$

The posterior predictive distribution is very useful if the goal is to predict future values stemming from the model, as it explicitly incorporates the level of uncertainty in the estimated model. Out-of-sample prediction that simply plugs in the best estimates of the model parameters will fail to take this uncertainty into account and can thus give distributions that are too narrow.

2.3.3 Bayesian multivariate linear regression

Regression analysis is a class of methods that aims to describe the relationship between some dependent variables (also called responses) and some independent variables (also called regressors). Linear regression is the most common setting, where the response is assumed to be some linear function of the regressors plus some error term. Ordinary (multiple) linear regression only targets one dependent variable at a time. If multiple responses are assumed to have correlated errors, multivariate regression techniques can be used to take advantage of this correlation. Regression analysis can be approached both from a Bayesian and a classical point of view, and the present work will use the Bayesian approach, utilizing Bayesian multivariate linear regression. An overview of the subject is available in e.g. [BT92].

The following exposition of the theory follows the Wikipedia article on the subject [Wik22].

Consider the regression problem

$$\begin{aligned} y_{i,1} =& \mathbf{x}_i^t \boldsymbol{\beta}_1 + \epsilon_{i,1} \\ &\vdots \\ y_{i,m} =& \mathbf{x}_i^t \boldsymbol{\beta}_m + \epsilon_{i,m} \end{aligned}$$

where $\mathbf{x}_i^t$ is a vector of k predictors for individual i out of n individuals, $\boldsymbol{\beta}_i$ are the regression coefficients corresponding to the response variable y_i and the set of errors $\{\epsilon_{i,1}, ..., \epsilon_{i,m}\}$ are correlated. This can be expressed compactly as

$$\mathbf{y}_i^t = \mathbf{x}_i^t \mathbf{B} + \boldsymbol{\epsilon_i}^t$$

where $\mathbf{B}$ denotes the (k, m) coefficient matrix

$$\mathbf{B} = \left[\left(\boldsymbol{\beta}_1 \right) \dots \left(\boldsymbol{\beta}_m \right) \right].$$

For an individual observation the noise vector $\boldsymbol{\epsilon}_i$ has a multivariate normal distribution: $\boldsymbol{\epsilon} \sim N_m(0, \Sigma_\epsilon)$ where Σ_ϵ is some covariance matrix of the noise. The entire regression problem can be written using matrices as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$$

where $\mathbf{Y}$ is an (n, m) matrix of all outcomes for all individuals, $\mathbf{X}$ is an (n, k) design matrix and $\mathbf{E}$ is an (n, m) matrix of all errors for all individuals. The standard linear least squares solution to this regression problem is to estimate $\mathbf{B}$ as $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$. In the Bayesian framework the goal is instead to find some posterior distribution for $\mathbf{B}$ and Σ_ϵ.

In the sequel, *vec* denotes the vectorization operator

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} \Rightarrow vec(A) = \begin{bmatrix} a_{1,1} \\ \dots \\ a_{m,1} \\ a_{1,2} \\ \dots \\ a_{m,n} \end{bmatrix}$$

and $\otimes$ denotes the Kronecker product

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \dots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \dots & a_{m,n}B \end{bmatrix}$$

$MN_{n,p}$ denotes an (n, p)-dimensional Matrix Normal distribution. This distribution is related to the multivariate normal distribution in the following way:

$$\mathbf{X} \sim MN_{n,p}(\mathbf{M}, \mathbf{U}, \mathbf{V}) \Leftrightarrow vec(\mathbf{X}) \sim N_{np}(vec(\mathbf{M}), \mathbf{U} \otimes \mathbf{V}).$$

The density of the (n, p) matrix-variate normal distribution is

$$\pi(\mathbf{X}|\mathbf{M}, \mathbf{U}, \mathbf{V}) = \frac{\exp(-\frac{1}{2}\text{tr}(\mathbf{V}^{-1}(\mathbf{X}-\mathbf{M})^t\mathbf{U}^{-1}(\mathbf{X}-\mathbf{M})))}{2\pi^{\frac{np}{2}}|\mathbf{V}|^{\frac{n}{2}}|\mathbf{U}|^{\frac{p}{2}}}$$

where $\mathbf{X}$ and $\mathbf{M}$ are (n, p), $\mathbf{U}$ is (n, n) and $\mathbf{V}$ is (p, p).

A conjugate prior distribution for $\boldsymbol{\beta} = vec(\mathbf{B})$ is of the form

$$\pi(\boldsymbol{\beta}, \Sigma_\epsilon) \propto \pi(\Sigma_\epsilon)\pi(\boldsymbol{\beta}|\Sigma_\epsilon),$$

where

$$\pi(\Sigma_\epsilon) \sim W^{-1}(\mathbf{V}_0, \nu_0)$$

and

$$\pi(\boldsymbol{\beta}|\Sigma_\epsilon) \sim N(\boldsymbol{\beta}_0, \Sigma_\epsilon \otimes \Lambda_0^{-1}).$$

where $\mathbf{V}_0$, ν_0, $\boldsymbol{\beta}_0$, Λ_0 are some prior hyperparameters and W^{-1} is the inverse Wishart distribution. Using this, a conjugate posterior can be found in the form:

$$\begin{aligned}\pi(\boldsymbol{\beta}, \Sigma_\epsilon|\mathbf{Y}, \mathbf{X}) \propto& \pi(\Sigma_\epsilon|\mathbf{Y}, \mathbf{X})\pi(\boldsymbol{\beta}|\Sigma_\epsilon, \mathbf{Y}, \mathbf{X}) \\ \propto& W^{-1}(\mathbf{V}_n, \nu_n)MN_{n,m}(\mathbf{B}_n, \Lambda_n^{-1}, \Sigma_\epsilon)\end{aligned}$$

Furthermore, the posterior hyperparameters are given by

$$\begin{aligned}\mathbf{V}_n =& \mathbf{V}_0 + (\mathbf{Y}-\mathbf{X}\mathbf{B}_n)^t(\mathbf{Y}-\mathbf{X}\mathbf{B}_n) + (\mathbf{B}_n-\mathbf{B}_0)^t\Lambda_0(\mathbf{B}_n-\mathbf{B}_0) \\ \nu_n =& \nu_0 + n \\ \mathbf{B}_n =& (\mathbf{X}^t\mathbf{X}+\Lambda_0)^{-1}(\mathbf{X}^t\mathbf{Y}+\Lambda_0\mathbf{B}_0) \\ \Lambda_n =& \mathbf{X}^t\mathbf{X}+\Lambda_0\end{aligned}$$

To create credible intervals for the regression coefficients, the marginal posterior distribution of $\mathbf{B}$ is needed. This marginal posterior is a matrix-variate t distribution, denoted by $T_{n,p}$

$$\pi(\mathbf{B}|\mathbf{Y}) \sim T_{n,m}(\frac{(\nu_n+m-1)}{2}, \mathbf{B}_n, \Lambda_n^{-1}, \mathbf{V}_n).$$

The matrix-variate $T_{n,p}$ distribution has the density

$$\pi(\mathbf{X}|\nu, \mathbf{M}, \mathbf{U}, \mathbf{V}) = \frac{\Gamma_p(\frac{\nu+n+p-1}{2})|\mathbf{I}_n + (\mathbf{V}^{-1}(\mathbf{X}-\mathbf{M})^t\mathbf{U}^{-1}(\mathbf{X}-\mathbf{M}))|^{-\frac{\nu+n+p-1}{2}}}{\pi^{\frac{np}{2}}\Gamma_p(\frac{\nu+p-1}{2})|\mathbf{V}|^{\frac{n}{2}}|\mathbf{U}|^{\frac{p}{2}}}$$

where ν is a scalar, $\mathbf{X}$ and $\mathbf{M}$ are (n, p), $\mathbf{U}$ is (n, n) and $\mathbf{V}$ is (p, p).

Selection of the prior can be done in different ways. By specifying the prior hyperparameters for the conjugate prior distribution any available information can be used. It is also possible to specify an improper prior, attempting to reduce its influence. One choice for improper prior, following [Mur07], can be obtained by setting the prior hyperparameters

$$\begin{aligned}|\mathbf{V}_0| =& 0 \\ \nu_0 =& -1 \\ \boldsymbol{\beta}_0 =& 0 \\ \Lambda_0 =& 0\end{aligned}$$

giving the improper prior

$$\pi(\boldsymbol{\beta}, \Sigma_\epsilon) \propto |\Sigma_\epsilon|^{-(m+1)/2}$$

which leads to a proper posterior and posterior hyperparameters according to the above description.

2.3.4 Assessing multivariate normality of the error

Any regression model makes some assumptions about the relationship it seeks to describe. It is important to investigate whether these assumptions hold. A standard linear regression model assumes that responses consist of some linear function of the predictors plus some error stemming from a normal distribution. This will lead to model residuals being t-distributed (which converges in distribution to a normal distribution as the number of degrees of freedom increases). To investigate whether model residuals are distributed in this way, visualizing the residuals by plotting them in histograms or scatter plots or making quantile-quantile (Q-Q) plots of the residuals against the quantiles of their theoretical distribution are common tools.

In the multivariate case, these visualizations are not quite sufficient. In the model proposed in 2.3.3 the errors are many-dimensional and assumed to follow a multivariate normal distribution, meaning that model residuals will belong to a multivariate t-distribution (converging in distribution to a multivariate normal distribution as the number of degrees of freedom increases). Marginal plots of the residuals are insufficient, as the marginal distributions of some multivariate distribution being normal does not prove that the joint distribution is multivariate normal (though the converse holds). Still, investigating marginal distributions is a common first step as clear deviations from univariate normality in the marginal distributions indicate that the joint distribution is not multivariate normal. Additionally, a multivariate tool that can be used is a Q-Q plot where the squared Mahalanobis distance of residuals is utilized [Tho02]. The Mahalanobis distance is a measure of distance of some point in a multivariate space from the origin relative to the distribution of interest, and can be seen as a multivariate generalization of the concept of measuring distance from the mean of some distribution in standard deviations of that distribution. The Mahalanobis distance of some point $\mathbf{x_i}$ relative to a multivariate probability distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is defined as

$$R_i = \sqrt{(\mathbf{x_i} - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x_i} - \boldsymbol{\mu})}$$

and the squared Mahalanobis distance of points belonging to a multivariate normal distribution is chi-square distributed with the number of degrees of freedom equal to the dimensionality of the multivariate normal distribution.

3 Method

3.1 Outline

In this section, the general medical and scientific goals of the study are translated into specific mathematical problem formulations and goals, providing an outline of the model in mathematical terms.

The goal of the study is to model health effects of the treatment based on some background information of the patients. Such a model might be used to increase the total beneficial effects of the treatment by allowing a better selection of patients to treat based on the initial measurements, to identify and investigate unwanted outcome discrepancies between patient categories, and to increase knowledge of how multimodal pain rehabilitation affects the patients.

To attain the goal of the study, a selection of what outcomes to model and what background information to use must be made. Given a selection of m different outcomes to model and k background variables to be used in the model, let each outcome measurement be represented by a point $y \in \mathbb{R}^m$ and each set of background variables be represented by a point $x \in \mathbb{R}^k$, such that the components of y correspond to the dimensions of various psychometric instruments and the components of x correspond to the various background variables. For each person there is one point x giving that person's background variables and (at most) a triplet (y_0, y_1, y_2), giving the values on the selected instruments before the treatment, immediately after the treatment, and 12 months after the treatment.

The health effects of the treatment must be defined in terms of y, as some function $B(y_0, y_1, y_2)$, since (y_0, y_1, y_2) are the only measurements that allow comparison of health before and after the treatment. In principle, the function B could be any function. Assuming it is some linear function will make optimization of the function simple. As the outcome dimensions selected in this project are all commonly used psychometric instruments in the context of pain research, designed and tested to measure specific aspects of health accurately, it is reasonable to choose a function B that does not transform the outcome dimensions non-linearly.

To model the health effects of the treatment before the treatment has occurred, a model that can predict (y_1, y_2) with uncertainty based on x and y_0 is required. Let y_1, y_2 be predicted by $f(x, y_0) + e(x, y_0)$ where $f(x, y_0)$ is the expected result after the treatment and $e(x, y_0)$ is an error term, a random vector with expectation 0. Then, using the linearity of B, the expected benefit of the treatment given (x, y_0) is

$$E(B(y_0, f(x, y_0) + e(x, y_0))) = B(x_0, E(f(x, y_0) + e(x, y_0))) = B(y_0, f(x, y_0)).$$

Thus, the expected treatment benefit depends on f but not on e. The details of e do matter for assessing uncertainty in the prediction.

In principle, the function f could be non-linear and complex in different ways. However, initial data analysis did not indicate any clear non-linear patterns or effects that could be modelled. Thus, f was modelled as an affine function. Non-linearities in the relationship between (x_0, y_0) and (y_1, y_2) could be accounted for by transformations of data.

The values of selected instruments after treatment y_1, y_2 might depend on x and y_0 in different ways from one another, which would need to be taken into account in modelling. However, the initial data analysis did not indicate any major differences between y_1, y_2, which is discussed further in Section 3.4.1. This led to the decision to model the measured values in each time step as

$$\begin{aligned} y_0 =& \gamma + \epsilon_0 \\ y_1 =& \gamma + \delta + \epsilon_1 \\ y_2 =& \gamma + \delta + \epsilon_2 \end{aligned}$$

where γ denotes the level of the instrument before treatment, δ denotes the change in the instrument due to treatment, and ϵ_i denotes the error in time step i. In this setup, the differences between the second and third time step are random and have no trend. Given this model, a natural choice for the function B is

$$\begin{aligned} B(y_0, y_1, y_2) =& \frac{b(y_1) + b(y_2)}{2} - b(y_0) \\ =& \frac{b(\gamma + \delta + \epsilon_1) + b(\gamma + \delta + \epsilon_2)}{2} - b(\gamma + \epsilon_0) \\ =& b(\delta) + b(\frac{\epsilon_1 + \epsilon_2}{2} - \epsilon_0) \end{aligned}$$

for some linear function b (e.g. selecting some outcome dimension). This choice of B isolates the treatment effect δ from γ and aims to minimize the impact of the error by averaging ϵ_1, ϵ_2. As this B only depends on the average of y_1, y_2 rather than their distinct values, this also led to the choice to use the observed average of these values, y', as response data, and let $f(x, y_0) + e(x, y_0)$ predict y'.

Now, the chosen B is a linear function depending only on the difference $d = y' - y_0$, which gives the identity $B(y_0, f(x, y_0)) = a \cdot d$, the dot product of d with some vector a. If all instruments are assumed to contribute equally to the overall health outcome, a can be set equal to the unit vector, but for the purposes of the present study it will be of more interest to consider the vector d itself, as a multivariate object. Depending on the perceived relative importance of different instruments, a can be adjusted to give a number corresponding to a desired model of total health outcome.

Next, an explicit $f(x, y_0)$ must be chosen. Since d is the object of interest, f is defined in terms of $f(x, y_0) - y_0$, as this is the prediction using f of d. A general formulation for affine f is

$$f(x) - y_0 = Ax + By_0 + c$$

for some matrices A, B and some vector c. However, the choice was made to exclude the By_0 term, as initial data analysis indicated that the errors were large compared to the effects and that the By_0 term would mostly capture the statistical behaviour

of the error. This decision is discussed further in 3.4.1. Therefore, $f(x) - y_0$ was modelled as

$$f(x) - y_0 = Ax + c$$

for some matrix A and vector c. Now, maximizing health gains is a matter of maximizing $b(Ax+c) = bAx+bc$ given the vector b, i.e. maximize vx where $v = bA$. To select the patients who will gain maximal health benefits from the treatment is simply a matter of multiplying their background variables x with a vector v and prioritizing those with the highest score.

It remains to produce A and c from the data. Bayesian inference is chosen as a framework to make explicit any uncertainty in estimates of these parameters. As the data is quite high-dimensional and noisy, explicit knowledge about parameter uncertainty is important.

To specify an inference procedure, a noise model must be specified. The above model is essentially a multivariate linear regression model. From the initial data analysis, it seems like a reasonable initial assumption that noise is multivariate normal and independent of y_0, and in this case there is a theory to describe such a model, a Bayesian multivariate normal linear regression model.

3.2 Variable selection

3.2.1 Responses

Since chronic pain is a multifaceted diagnosis and treatment does not aim to completely restore patients, but rather to manage symptoms, outcomes are not dichotomous. The biopsychosocial approach to pain shows the importance of measuring treatment outcomes not only in terms of physical pain but also in additional dimensions, but it is not immediately obvious which such outcomes are interesting to measure. To establish a scientific standard of treatment outcome measures for pain treatment, the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT), which is composed of leading researchers in the field, has recommended six outcome domains with several specific measures of each suggested to use in clinical trials [Dwo+05]:

- Pain
  - 11-point (0-10) numerical rating scale of pain intensity
  - Usage of rescue analgesics
  - Categorical rating of pain intensity (none, mild, moderate, severe) in circumstances in which numerical ratings may be problematic
- Physical functioning (either one of two measures)
  - Multidimensional Pain Inventory Interference scale
  - Brief Pain Inventory interference items
- Emotional functioning (at least one of two measures)
  - Beck Depression Inventory

  - Profile of Mood States

- Participant ratings of global improvement and satisfaction with treatment
  - Patient Global Impression of Change
- Symptoms and adverse events
  - Passive capture of spontaneously reported adverse events and symptoms and use of open-ended prompts
- Participant disposition
  - Detailed information regarding participant recruitment and progress throughout the trial, including all information specified in the CONSORT guidelines

Of the specific measures recommended by IMMPACT, only the 11-point numerical rating scale (NRS) of pain intensity and the Multidimensional Pain Inventory (MPI) Interference scale are available in the SQRP data. Thus, several more outcome measures were used. Furthermore, even though they are not included in the IMMPACT recommended domains, measures of quality of social life and activity were analyzed, as they are readily available and of importance to quality of life of chronic pain patients.

Note that the IMMPACT recommendations are older than the SQRP, and thus the specific measures suggested by IMMPACT may have been replaced by other instruments in the SQRP.

As described in Section 2.3.1, the instruments in the SQRP aim to measure latent properties and it is not clear that the dimensions that are formulated for the instruments correspond to actual latent properties. Although all of the mandatory instruments in the SQRP are widely used, academic investigation of their psychometric properties have given varying results. The SQRP instruments that were described in 2.2 were considered for inclusion as outcome measures in the light of previous research.

Though the MPI is prevalent in pain rehabilitation evaluation and one of its dimensions is recommended in the IMMPACT guidelines, concerns have been raised regarding its psychometric properties. One study found that a third of the study participants were assigned different MPI profiles on retest though no treatment had been administered [BJT04], and was unable to predict stability of MPI profile assignment from other factors. This indicates that MPI profile change is a highly noisy measure of patient improvement, even though other studies have found that it correlates positively with positive treatment outcomes [NNS14]. Research on the individual MPI scales indicate that several of these have insufficient range to capture the full variation in the population [MN11]. A summarized version of the MPI instrument that transforms the original 12 dimensions into 3 dimensions with improved psychometric properties has been proposed. These summary dimensions are derived through factor analysis, in a process which is described in Section 2.3.1 and which follows the paper by McKillop and Nielson.

The three proposed dimensions are described as "Impairment", being composed mainly of the dimensions "Pain severity", "Interference", "Life control" and "Affective distress" in the original MPI; "Social support", being composed mainly of the

dimensions "Support", "Negative responses", "Solicitous responses" and "Distracting responses" in the original MPI; and "Activity", being composed mainly of the dimensions "Household chores", "Outdoor work", "Activities away from home" and "Social activities" in the original MPI.

The NRS 11-point self-reported pain scale has been found to be reliable, valid and clinically useful as a measure of pain [Far+01], [Eua+22].

The psychometric properties of the SF36 and HADS instruments on SQRP data have been investigated in [LoM+19]. The focus of this investigation is to what extent the several dimensions of those instruments actually measure unique latent properties of the test-taker in a chronic pain setting. To exemplify the issue in question: the SF36 has one dimension termed "Physical functioning" and one termed "Role limitations due to physical health". One might ask whether those dimensions are actually independent, or at most weakly correlated, in a test-taker. In general, one might ask how many unique latent properties of a test-taker some instrument actually measures, and what those properties might be. The paper comes to the conclusion that the SF36 instrument, which has 8 dimensions, actually supports two independent constructs on SQRP data: physical health and mental health. This view of the SF36 has been put forth since shortly after the instrument was first developed, and its original authors have developed summary scales corresponding to these two separate constructs [War94].

Furthermore, [LoM+19] finds that the HADS instrument, which has 2 dimensions termed "Anxiety" and "Depression", only supports one general construct of overall emotional distress on SQRP data. Other studies, however, have found that HADS supports two constructs, in accordance with its design [GH09], [Her+03]. It also concludes that the EQ-5D has low reliability on SQRP data and recommends against using it.

The present study uses the three summarized MPI dimensions of McKillop and Nielson, the two summarized SF-36 dimensions, the NRS 11-step self-reported pain scale and the two original HADS dimensions. The selected outcome measures, their encoded name in the study, and the outcome domains they correspond to are given in Figure 3.1. Note that the IMMPACT-recommended domains of Participant ratings of global improvement and satisfaction with treatment, Symptoms and adverse events and Participant disposition are not included - this is due to a lack of variables in the dataset corresponding to these domains. Two domains which are not included in the IMMPACT recommendations are also listed, those being Social functioning and Activity - this is due to the MPI - Social support and MPI - Activity dimensions being readily available and seeming interesting, but not corresponding to an IMMPACT domain.

The EQ-5D was wholly omitted, as a previous study recommended against using it. When the author of the present study conducted interviews with clinicians working with pain rehabilitation, they were also critical of the EQ-5D and disputed its clinical usefulness. Additionally, the EQ-5D did not provide measures for any outcome domain that lacked alternatives, and was thus deemed superfluous.

The HADS scales were included as separate scales, despite previous results arguing that they might only reflect one construct of mental well-being in chronic pain patients. The potential advantage of reducing the two HADS dimensions to one, i.e. better construct validity and dimension reduction, were deemed not to be worth the loss of interpretability of the original scales.

The SF-36 summarized physical and mental summary scales were included, as these scales are interpretable according to the original design of the instrument and seem to be better supported as constructs than the full 8 dimensions on the SQRP dataset.

The three summarized MPI scales were included as these scales did not lose much interpretability as compared to the original 12 dimensions and had been found to have better psychometric properties than the full 12 dimensions [MN11]. Their construct validity for the data used in the study was assessed, and the present study replicates the results of the previous study on the SQRP dataset. The method of this replication is described in Section 3.3.2 and its result in Section 4.1.

3.2.2 Response rescaling

The selected outcome measures vary in what direction on the scale denotes a positive health outcome. For MPI - Impairment, NRS (self-reported level of pain during the past week), HADS - Anxiety and HADS - Depression, higher scores indicate worse health, whereas for MPI - Social support, MPI - Activity, SF36 - Physical summary and SF36 - Mental summary, a higher score indicates better health. For ease of interpretation, the scales were inverted such that a higher score indicates better health in the present study, with the exception of visualizations of scores at baseline in Figure 4.6.

The ranges of the selected responses also vary. MPI - Impairment, MPI - Social support and MPI - Activity take values on a 0-6 scale, the SF-36 summary dimensions take values on a 0-100 scale, the HADS dimensions take values on a 0-21-point scale and the NRS takes values on a 0-10 scale. In order to make effect sizes comparable, standardized versions of all outcome scores were created by subtracting the mean of the initial time step and dividing by the standard deviation of the initial time step for every score in the respective outcome dimensions. Thus, for the standardized scores the initial mean was 0 and the initial standard deviation was 1 in every dimension. The standardized scores were used to fit the regression model and compute the posterior predictive distributions for individuals using the model, as described in sections 4.3 and 4.4, to allow for easier comparison of effect sizes across outcome dimensions.

3.2.3 Regressors

A set of background variables was selected to be included in the study as regressors. The names of these variables and their coding in the database are displayed in Figure 3.2. Regressors were selected to reflect common demographic data and the character of the patient's chronic pain condition. A potentially interesting set of variables encoding the degree to which different pain mechanisms contributed to the patient's pain were omitted due to missingness over 50%.

Categorical regressors with more than two levels (Country of birth, Level of education, Employment status) were re-encoded using dummy variables to indicate the non-baseline levels of the variable, with the typical value being the baseline in each case (Sweden for Country of birth, Upper secondary school for Level of education, Employed for Employment status).

	Encoded name	Outcome domain
MPI - Impairment	Impairment	Pain, Physical functioning, Emotional functioning
MPI - Social support	Social_support	Social functioning
MPI - Activity	Activity	Activity
SF36 - Physical summary	sf36_pcs	Physical functioning
SF36 - Mental summary	sf36_mcs	Emotional functioning
NRS self-reported level of pain	SM_GMSN	Pain
HADS - Anxiety	had1	Emotional functioning
HADS - Depression	had2	Emotional functioning

Figure 3.1: Coding and outcome domain of outcome measures included in the study

Variable	Variable coding
Gender	GENDER
Age	d_age
Country of birth	
Sweden	F_LAND_SWE
Nordic countries	F_LAND_NOR
Europe	F_LAND_EUR
Other	F_LAND_OTH
Level of education	
Elementary school	UTB_NIVA_ELE
Upper secondary school	UTB_NIVA_GYM
University	UTB_NIVA_UNI
Other	UTB_NIVA_OTH

Variable	Variable coding
Employment status	
Employed	ARB_FORM_ANSTALLD
Unemployed	ARB_FORM_ARBETSSOKANDE
Student	ARB_FORM_STUDERANDE
Not in the labor force	ARB_FORM_EJARBETANDE
Days since pain debut	d_dag_smint
Level of belief in being restored	RINSTBLI
Is the pain periodic or constant	SMPERIOD_NY1
How many visits with doctor due to pain in the last year	ini_N71

Figure 3.2: Coding of background variables included in the study

3.3 Data selection and preprocessing

Data was obtained from the SQRP. The obtained dataset is from the 2009-2016 epoch of SQRP data, containing 50 808 entries. Each entry consists of the forms gathered from a patient during one process of referral, evaluation, and treatment, as described in Section 2.2.

The dataset contains the filled-out forms of every person who has been evaluated for potential treatment using multimodal pain rehabilitation with participating healthcare providers. This means that many of the responses come from individuals who were evaluated for, but weren't offered, treatment, potentially several times, and individuals who were offered treatment but didn't complete it. These responses were excluded from the analysis, leaving a dataset containing 19 607 sets of forms from unique individuals who completed multimodal pain rehabilitation. This reduced data set was obtained from Helene Svensdotter, who performed the data selection as part of her Ph.D. work. Helene Svensdotter also performed some further data pre-processing, such as re-encoding "No answer" from a numeric code to an NA variable and removing duplicate entries. From this dataset, a further reduction was performed by omitting any response where the recorded dates of the three time steps were chronologically inconsistent, such as the follow-up step happening before the initial step. After this reduction, a set of 15 101 forms forms were left.

The reduced dataset was subdivided into a set of respondents who filled out the forms in all three time steps and a set of respondents who did not respond in the final time step, the 12-month follow-up. From these subsets, a further exclusion was performed of any responses that were incomplete in any of the regressor or response variables used for regression analysis. Thus, respondents who filled out all relevant variables in the second form but did not fill out the third form at all were included. These responses were included since the group who did not fill out the third form had significantly worse values for all response variables except Social support in the initial time step than the group who did fill out the third form, as well as significantly lower mean increases in Impairment, SF36 - Physical health, reported level of pain and HADS - Anxiety (the group who did not fill out the third form did, however, have significantly higher mean Activity than those who did not). These differences were evaluated using unequal variances t-tests for equal means, the results of which are displayed in Figure 3.3, where responses with $p < 0.05$ are highlighted in orange. Excluding this group entirely from analysis might lead to overvaluation of the effect of treatment or other systematic errors. Recombining the resulting subsets of complete responses gave a final analysis dataset with 8168 responses.

3.3.1 Treatment of missing values

The approach to handling missing values chosen in this study has potential disadvantages. Removing any incomplete response reduces the power of any statistical test performed on the data as it decreases the size of the utilized sample. It also risks introducing bias into any analysis performed on the data if the set of responses that is removed is distributed differently in any variables that are considered in analysis than the set of responses that remain. To investigate the risk of introducing bias, unequal-variances t-tests were performed to find potential differences in means of

	Difference in mean (fol filled - not filled)	p of t-test for equal mean
Impairment	0.1450440	0.0000000
Social_support	0.0043828	0.8757299
Activity	0.1417367	0.0000000
sf36_pcs	0.6508765	0.0024941
sf36_mcs	1.5242687	0.0000357
SM_GMSN	0.2109126	0.0000062
had1	0.5364524	0.0000639
had2	0.6946069	0.0000000

(a) Initial values

	Difference in mean (fol filled - not filled)	p of t-test for equal mean
Impairment	0.0958965	0.0000958
Social_support	-0.0378872	0.0516313
Activity	-0.0469986	0.0192395
sf36_pcs	0.8787065	0.0000544
sf36_mcs	0.4773869	0.1847051
SM_GMSN	0.2454154	0.0000076
had1	0.3562031	0.0012186
had2	-0.0194788	0.8574746

(b) Change from before treatment to after

Figure 3.3: Unequal variances t-tests for equal means of respondents who did or did not fill out follow-up form

	Difference in mean (included - excluded)	p of t-test for equal mean
Impairment	0.0331143	0.0280675
Social_support	0.0405129	0.0217484
Activity	0.0490719	0.0011623
sf36_pcs	-0.4336620	0.0023966
sf36_mcs	0.8438391	0.0001694
SM_GMSN	0.1020308	0.0006028
had1	0.2802622	0.0003744
had2	0.1988102	0.0068529

(a) Initial values

	Difference in mean (included - excluded)	p of t-test for equal mean
Impairment	0.0562624	0.0016245
Social_support	0.0022154	0.8745316
Activity	0.0142717	0.3304622
sf36_pcs	-0.0171074	0.9047535
sf36_mcs	0.6008356	0.0077476
SM_GMSN	0.0856544	0.0158084
had1	0.2558080	0.0000956
had2	0.2920009	0.0000078

(b) Change from before treatment to after

Figure 3.4: Unequal variances t-tests for equal means of included and excluded responses

response variables between the responses that were kept and removed. These tests were performed both for initial values, reported in table 3.4a, and changes from before treatment to after treatment, reported in table 3.4b. Responses with $p < 0.05$ are highlighted in orange. For these tests, raw response scores were used instead of standardized values, but for ease of comparison responses were flipped such that a greater value indicated better health.

Several significant differences between the included and excluded groups were found. For the initial values, there were significant differences between included and excluded responses for every response variable, and for every variable except SF36 - Physical health, the included group had higher scores indicating better health. For changes from before to after treatment, there were significant differences between groups for Impairment, SF36 - Mental health, self-reported level of pain, HADS - Anxiety and HADS - Depression, and for all of these responses the included group improved more. Though many of the differences were significant, none of them were very large, with every mean difference being smaller than $\frac{1}{10}$ of the estimated standard deviation of the included responses for that variable.

An alternative to excluding all non-complete responses would be to perform item imputation, replacing missing values in relevant variables with a value determined

by some method. Either approach to handling missing values has advantages and disadvantages. Imputation will keep the utilized sample as large as possible, which will give any performed test higher power. However, common methods of imputation will cause other issues. Two common methods of imputation are mean substitution and regression imputation. Mean substitution replaces any missing variable with the sample mean for that variable, and regression substitution uses a regression model estimated from the sample to predict a missing value using the non-missing values of the response. These methods cause opposite problems: mean substitution tends to reduce any sample correlations involving variables that are imputed, as the imputed value is independent of any other variables in the same response. This can lead to e.g. overestimation of uncertainty and underestimation of effect size in regression analysis. Regression imputation, on the other hand, tends to artificially strengthen confidence in any relationship that is observed in the sample.

3.3.2 MPI instrument summarization

In accordance with the procedure of [MN11], principal component analysis was performed on the 12-dimensional MPI data in the initial time step. A scree plot was used to decide on the number of components to keep. These principal components were the rotated using a Promin rotation. The results are detailed in 4.1.

3.4 Modelling

3.4.1 Model assumptions

The three data points gathered from each individual in the dataset can be fitted into several modelling approaches. These approaches all build on assumptions about the way the treatment impacts the relevant outcome measures and how time after completed treatment affects this impact.

In the present study, the outcome measures used are assumed to be noisy observations of some latent variables, such as abstract well-being or health, rather than the only source of error being e.g. ticking the wrong box on the questionnaire. This is relevant for modelling because it means there is some potentially large amount of randomness to the measurements taken from each individual, and thus to the change measured. An individual that is, by chance, having particularly little pain on the day of initial measurement might report higher pain on the day of conclusion of treatment simply due to chance, even if treatment has reduced average daily pain. A consequence of this, especially if the errors are large compared to the effect of treatment, is that a pattern of regression to the mean is to be expected in the data, yielding a negative correlation between initial value and change in value. Those patients who have a particularly large negative error in the first measurement will tend to score higher in the second measurement, and vice versa, simply because it is unlikely that a similarly extreme error will occur in consecutive measurements. If the initial measurement is modelled as a regressor of the change in measurement, this can lead to the spurious conclusion that this regressor is highly significant, even if the effect is only a statistical artefact. On the other hand, the initial level of a variable might have an actual effect on the change in measurement that would be overlooked by excluding it from modelling. In the present study, baseline values

of variables were not used as regressors, as the data was deemed so noisy and the effects of the treatment so small on average that the statistical effects of the error would likely dominate. The lack of a control group that did not receive treatment precluded the use of a comparison to such a group to distinguish treatment effect from error. Initial data analysis using simple linear regression on the marginal distributions of the responses indicated that baseline values indeed displayed negative correlation with change in those values. A previous study using SQRP data utilized baseline values as regressors, and reported significant effects of all baseline values on the corresponding outcomes [Ger+16]. That study discusses that the effect might be due to regression to the mean but does not investigate the possibility further.

A fundamental question that any model must take into account is whether to use data from the conclusion of treatment or the 12-month follow-up. If the assumption is made that there should be no systematic difference in some measure between the conclusion of treatment and the 12-month follow-up, then the follow-up data might be used as a second measurement of treatment effect, giving better estimates of regression coefficients. On the other hand, if some change in measure is assumed to occur during the 12-month period following treatment, that change should be modelled explicitly. The difference in mean between the "con" time step, immediately after treatment, and the "fol" time step, 12 months after treatment, was tested using t tests with unequal variance. The results are reported in Figure 3.5. From these tests, there is no significant difference in the mean between time steps for Impairment, Activity or HADS - Anxiety. The mean is significantly larger in the con timestep for Social support, SF36 - mental summary and reported level of pain, while it is significantly larger in the fol timestep for SF36 - Physical summary and HADS - Depression. This indicates that while there are significant differences between the time steps, they are not uniform in their direction, and for some responses there seems to be no difference at all. It is not clear what gives rise to this pattern of changes, and it could be investigated further, especially as to how it impacts modelling of treatment outcomes. For the purposes of the present study, these differences were deemed minor enough that no change to the described modelling approach was necessary. This means that the response values in the regression model were the values in the initial time step subtracted from the averages of the values in the second and third time-step, for each dimension.

It should also be noted that the t-tests only test for equality of the first moment of the distributions in the different time steps, and there might be other differences that would impact modelling, e.g. greater variance in the fol time step, and that such differences might also impact modelling. Initial data visualization did not provide any reason to believe such differences existed, but more in-depth testing might be useful.

3.4.2 Bayesian multivariate linear regression

Bayesian multivariate linear regression was implemented using the responses and regressor described in sections 3.2.1 and 3.2.3 respectively, yielding a model with 16 regressors (including the intercept) and 8 responses. Uninformative priors were assumed according to what is described in 2.3.3.

The posterior parameters were computed in R using a simulation approach. The MixMatrix R package was used for its implementation of the distribution functions

	Estimated mean difference (con - fol)	p of t-test for equal mean
Impairment	0.0168862	0.4033053
Social_support	0.0906931	0.0000007
Activity	0.0046840	0.7558178
sf36_pcs	-0.9162362	0.0000004
sf36_mcs	0.5656005	0.0226273
SM_GMSN	0.1344048	0.0010869
had1	0.0431355	0.6063156
had2	-0.3760685	0.0000040

Figure 3.5: Unequal variances t-tests for equal means of responses in con and fol time steps

and random samplers for the matrix normal and matrix t distributions [Geo21]. R code was written to set up the design matrix $\mathbf{X}$ and response matrix $\mathbf{Y}$ and compute the posterior hyperparameters. Then, 1000000 samples were simulated from the marginal posterior distribution of $\mathbf{B}$. The bayestestR package was used to describe the posterior distribution, construct credible intervals and compute probability of direction for all regression coefficients [MBL19].

To demonstrate the posterior predictive results of the model and to provide illustrative examples, the posterior predictive distribution of treatment outcomes for two example patients were simulated. The simulation was carried out by randomly selecting one of the simulated marginal posterior distributions of $\mathbf{B}$, multiplying the vector of regressors for both individuals by it, simulating an error distribution from the posterior distribution of Σ_ϵ and then adding the simulated error to the previous product. This process was carried out 100000 times.

4 Results

4.1 MPI instrument summarization

The PCA gave similar results to that in the earlier work, with three dimensions having eigenvalues clearly greater than 1 and the rest having eigenvalues clearly smaller than 1, with a scree test indicating a three-factor solution, as displayed in Figure 4.1. This three-factor solution accounted for 58% of total variance, very close to the 60% of total variance explained by a three-factor solution in previous work. After Promin rotation, this provided a set of loadings similar to that found in [MN11], detailed in table 4.2, with similar correlations between dimensions, detailed in table 4.3. Notably, the eigenvalues of PCA vectors as well as the loadings found in the present work tended to be smaller in absolute terms than those found in previous research, but the proportions of both the eigenvalues and the loadings were very similar to those in previous research. As a system of principal components is scale invariant, this has no bearing on the result, and comparisons of relative weights of different factors can be performed by scaling the entire set of factor loadings by some constant. The present work found greater negative correlation between the first and third factor than McKillop and Nielson, and some very weakly loaded factors have different signs in the present study, but the major patterns are consistent and support the three-factor interpretation proposed in the previous study.

MPI data was converted to this three-dimensional setting as in [MN11]. An Impairment score was calculated by averaging the MPI dimensions Pain Severity, Interference, Life Control and Affective distress (after inverting the Life Control dimension). A Social Support score was computed by averaging the MPI dimensions Support, Negative Responses, Solicitous Responses and Distracting Responses (after inverting the Negative Responses dimension). An Activity score was created by averaging the MPI dimensions Household Chores, Outdoor Work, Activities Away From Home and Social Activities. This provided three new dimensions with scores ranging from 0 to 6, with a higher Impairment score indicating greater negative impact of pain, and higher Social Support and Activity scores indicating a smaller negative impact of pain. Descriptive statistics for these new dimensions are given in Section 4.2, together with descriptives for other responses.

4.2 Exploratory analysis

A description of frequencies and, where applicable, distributions of regressors in the dataset is given in figures 4.4 and 4.5. In the histograms over age and days since pain debut, some extremely high values were omitted to make the figures more readable. For age, 6 individuals over the age of 70 were omitted out of a total of

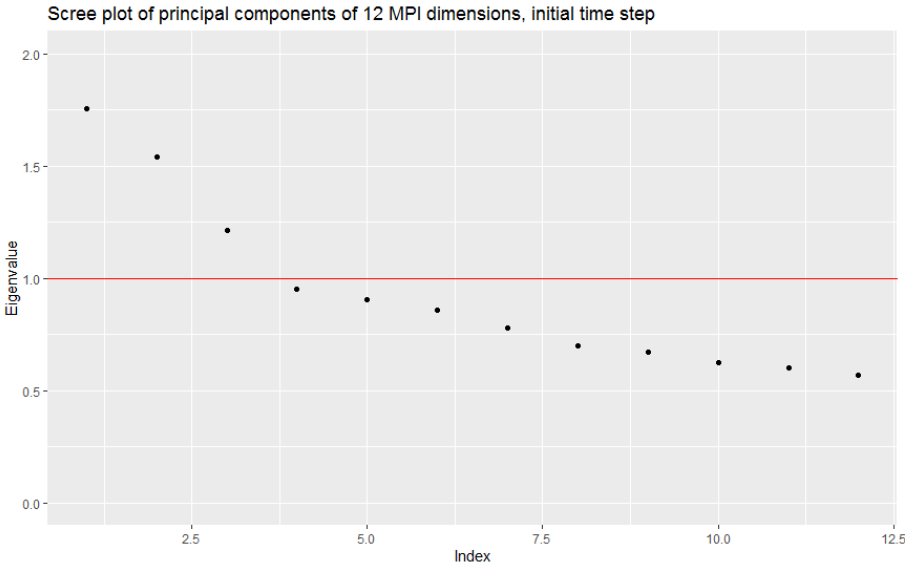


Figure 4.1: Scree plot of PCA on MPI, initial time step.

	Factor		
	1	2	3
Pain severity	**0.4696770**	0.1144426	-0.0660353
Interference	**0.4643654**	0.0637422	0.0779295
Life control	**-0.4605090**	0.0625738	-0.0690103
Affective distress	**0.4984420**	-0.0794429	-0.0185242
Support	0.0587303	**0.5039564**	0.0371933
Negative responses	0.2849415	**-0.3432337**	-0.0858807
Solicitous responses	0.0690655	**0.5330206**	-0.0073091
Distracting responses	0.0952522	**0.4857870**	-0.0810459
Household chores	-0.0005224	-0.1663605	**-0.4913777**
Outdoor work	0.0291200	-0.1301796	**-0.4402904**
Activities away from home	-0.0163521	0.0983927	**-0.5374437**
Social activities	-0.0675044	0.1631864	**-0.4956500**

Figure 4.2: Loadings of original MPI dimensions onto new factors. The greatest absolute loading for each original dimension is bolded.

Correlation		
1	2	3
1.000000	0.0502870	-0.3364710
0.050287	1.0000000	-0.0277247
-0.336471	-0.0277247	1.0000000

Figure 4.3: Correlation matrix of the new MPI factors.

Variable	Percentage
Country of birth	
Sweden	85.0%
Nordic countries	2.8%
Europe	4.2%
Other	8.0%
Level of education	
Elementary school	11.0%
Upper secondary school	55.4%
University	27.9%
Other	5.7%

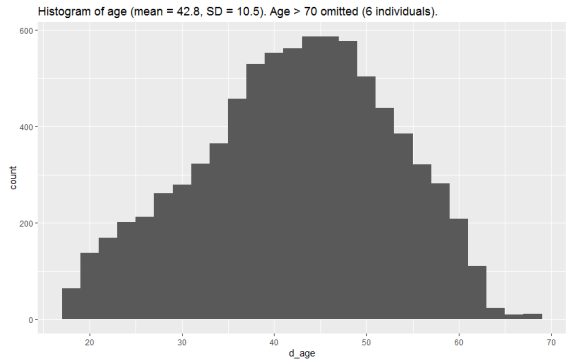
Variable	Percentage
Gender	78.2% (female)
Is the pain periodic or constant	84.8% (constant)
Employment status	
Employed	73.0%
Unemployed	15.5%
Student	2.9%
Not in the labor force	6.2%

Figure 4.4: Frequency of regressors in the dataset

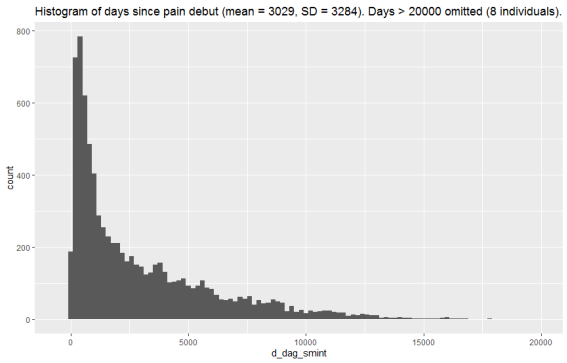
8168 individuals in the dataset, the most extreme of which was aged 87. For days since pain debut, 8 individuals with more than 20000 days since pain debut were omitted, the most extreme of which had 35010 days since their pain debut. Note that the variable *ini-N71* which describes the amount of visits with a doctor due to pain the past year is not continuous but categorical with three levels: level 0 indicating 0-1 visits, level 1 indicating 2-3 visits and level 2 indicating 4 or more visits.

The patients in the dataset were most commonly female (78.2%), born in Sweden (85%) and employed (73%). It was much more common for them to have constant pain, as opposed to pain being recurring (84.8% had constant pain).

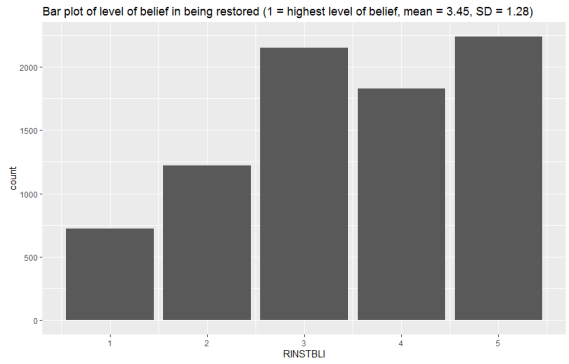
Baseline levels of response variables are displayed in Figure 4.6. Note that in this figure, the raw response values are used, meaning that better health is indicated by high levels of Social support, Activity, SF36 - Physical summary, SF36 - Mental summary; and low levels of Impairment, self-reported level of pain, HADS - Anxiety and HADS - Depression.



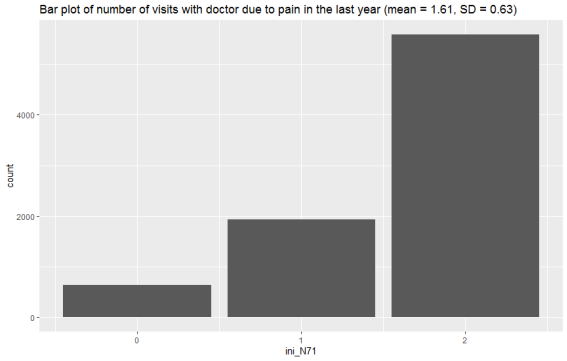
(a) Age



(b) Days since pain debut



(c) Level of belief in recovery



(d) Number of visits with doctor due to pain in the last year

Figure 4.5: Histograms and bar plots of non-categorical or multi-categorical regressors in the dataset

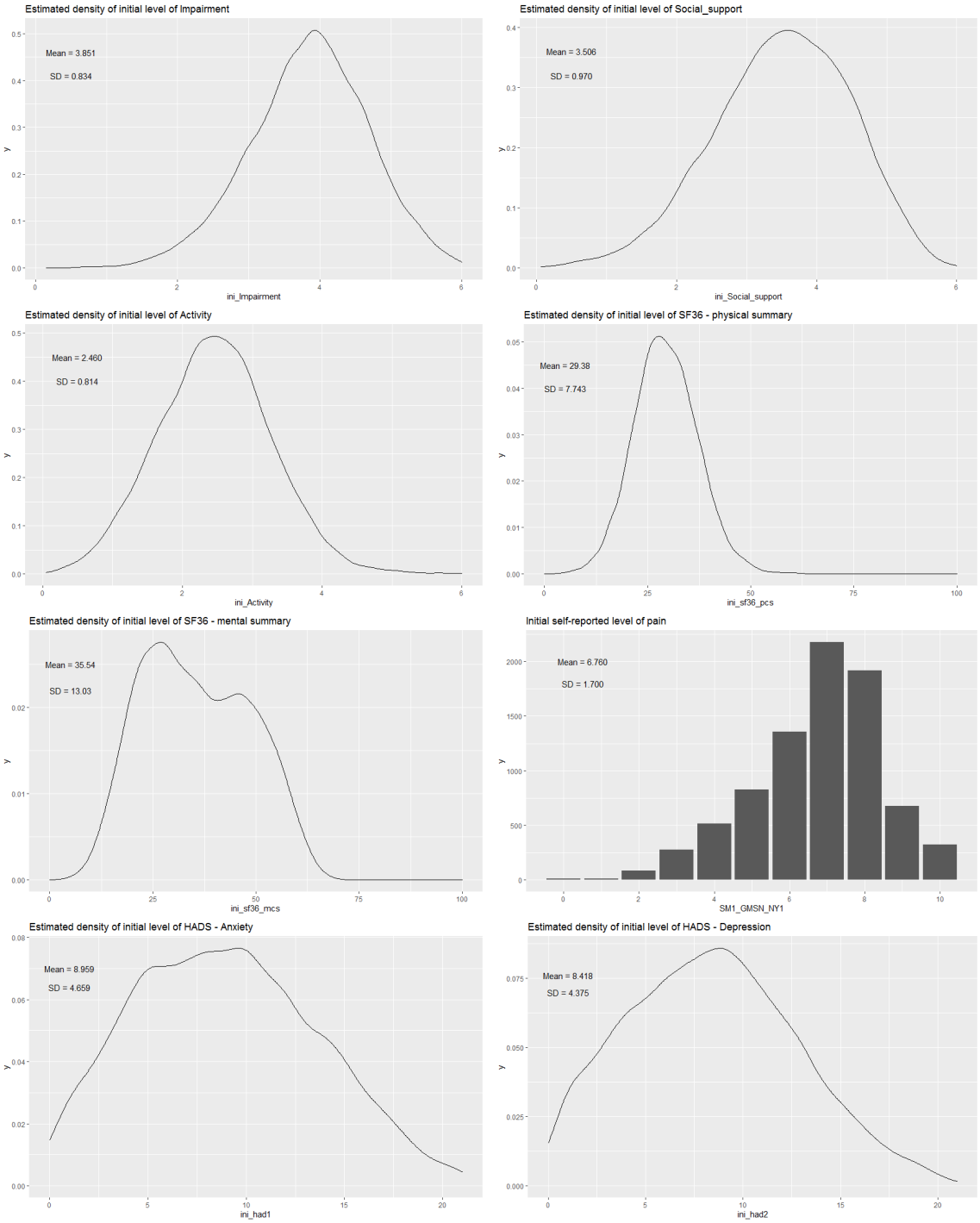


Figure 4.6: Baseline levels of response variables (scales not inverted)

Observed changes in response variables, from the first time-step to the average of the two last time-steps, are displayed in Figure 4.7. The changes in responses are all unimodal and somewhat bell-shaped, though most of them display some skewness. Note that in this figure and onward, some response values are reversed such that better health is indicated by a higher value for all responses.

4.3 Regression model

Bayesian multivariate linear regression was performed according to Section 3.4.2. The full results of this regression are displayed in Appendix B. In this section, regressors that are significant in the sense of having probability of direction larger than 0.975 will be discussed. Statistics describing the posterior distribution of significant regression coefficients are given in table 4.8. Note that intercepts are not included in this table - intercepts were significant for all responses except for Social support. As the regression model was fitted using standardized response values, as described in Section 3.2.2, all effect sizes and credible intervals are to be interpreted such that a score of 1 indicates a change equal to the standard deviation in the initial time step of that dimension. Statistics for the posterior distributions of all effects and all responses, using non-standardized response values, can be found in Appendix B.

For every response variable except Social support, the intercept had PD greater than 0.9995, indicating great certainty in effect direction. Furthermore, for every response except Social support, the mean estimated intercept was positive, indicating improved health. For Social support, the mean estimated intercept was negative, indicating worsening social support after treatment, but the PD for this estimate was notably lower, at 0.9648, below the 5% significance threshold.

Note that due to the way categorical regressors are set up in the model, the intercept corresponds to the mean estimated treatment effect on a man who is 0 years old, is born in Sweden, has upper secondary school education, is employed, has gone 0 days since his pain debut, is highly confident that he will get better as a result of treatment, has periodically recurring pain rather than constant pain, and has sought medical aid for his pain at most once in the past year. Examples of the model's estimated treatment effect for several realistic patient types are given in Section 4.4.

For Impairment, significant regressors were university education, unemployment, days since pain debut, belief in recovery and constancy of pain. The strongest effects were associated with a very low belief in recovery (the estimated effect of stating the lowest belief in recovery was -0.212 points of change) and being unemployed (estimated at -0.154 points of change).

No significant regressors were found for Social support.

For Activity, being unemployed and days since pain debut were significant regressors.

For SF36 - Physical summary, being born in Europe outside of the Nordic countries or being born outside Europe, having at most elementary school education, being unemployed and having low belief in recovery were significant regressors. The strongest effects were associated with very low belief in recovery (-0.385 points of change), being born in Europe outside the Nordic countries (-0.187 points of change) and being unemployed (-0.179 points of change).

For SF36 - Mental summary, constancy of pain was the only significant regressor.

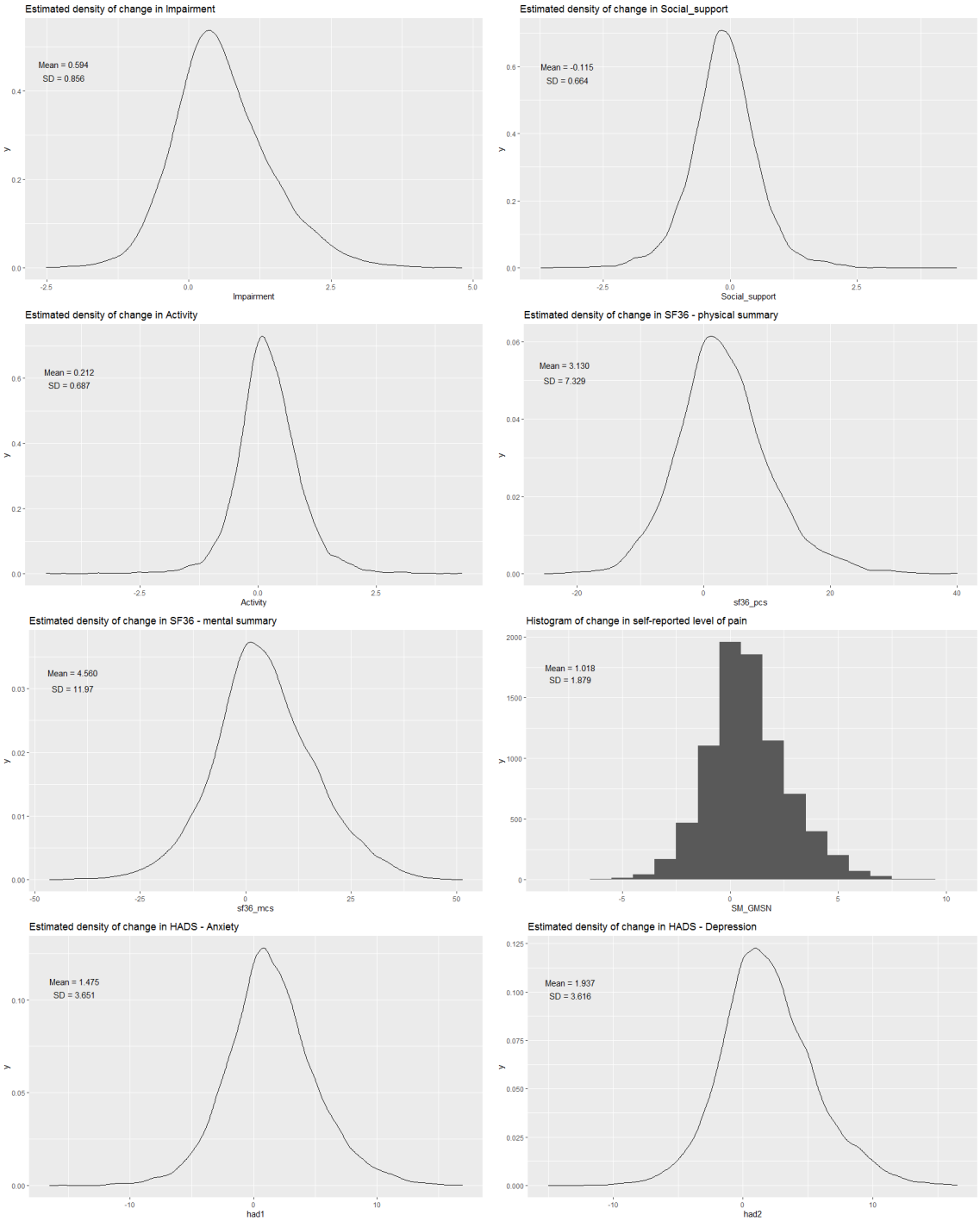


Figure 4.7: Change in levels of response variables

	Median	CI_low	CI_high	pd
Impairment				
Impairment.UTB_NIVA_UNI	0.0763377	0.0015032	0.1499884	0.977871
Impairment.ARB_FORM_ARBETSSOKANDE	-0.1539336	-0.2426588	-0.0654332	0.999661
Impairment.d_dag_smint	-0.0000105	-0.0000205	-0.0000003	0.979799
Impairment.RINSTBLI	-0.0424141	-0.0683326	-0.0165019	0.999317
Impairment.SMPERIOD_NY1	-0.1089320	-0.1989338	-0.0193037	0.991240
Activity				
Activity.ARB_FORM_ARBETSSOKANDE	-0.0736624	-0.1455949	0.0004513	0.975969
Activity.d_dag_smint	-0.0000161	-0.0000245	-0.0000079	0.999925
SF36 - Physical summary				
sf36_pcs.F_LAND_EUR	-0.1872167	-0.3301501	-0.0440320	0.994782
sf36_pcs.F_LAND_OTH	-0.1182876	-0.2263321	-0.0120376	0.984684
sf36_pcs.UTB_NIVA_ELE	-0.0997946	-0.1956618	-0.0049100	0.980029
sf36_pcs.ARB_FORM_ARBETSSOKANDE	-0.1792961	-0.2611171	-0.0986464	0.999988
sf36_pcs.RINSTBLI	-0.0770934	-0.1009494	-0.0534528	1.000000

	Median	CI_low	CI_high	pd
SF36 - Mental summary				
sf36_mcs.SMPERIOD_NY1	-0.1285919	-0.2092190	-0.0480586	0.999140
Pain during past week				
SM_GMSN.ARB_FORM_ARBETSSOKANDE	-0.1657931	-0.2604270	-0.0696961	0.999687
SM_GMSN.RINSTBLI	-0.0588616	-0.0869299	-0.0310982	0.999984
SM_GMSN.SMPERIOD_NY1	-0.1179212	-0.2149134	-0.0219658	0.991651
HADS - Anxiety				
had1.d_age	-0.0025488	-0.0049857	-0.0001010	0.979610
had1.SMPERIOD_NY1	-0.1110630	-0.1797213	-0.0422932	0.999249
HADS - Depression				
had2.GENDER	0.0978951	0.0359053	0.1597076	0.998954
had2.UTB_NIVA_UNI	0.0646227	0.0040008	0.1237509	0.982710

Figure 4.8: Posterior distribution statistics for significant regression coefficients (standardized)

For self-reported pain during the past week, unemployment, low belief in recovery and chronicity of pain were significant regressors. The strongest effect was associated with very low belief in recovery (-0.295 points of change on a 10-point scale).

For HADS - Anxiety, age and chronicity of pain were significant regressors.

For HADS - Depression, gender and level of education were significant regressors.

For each response, there were one or more regressors with probability of direction in the range of 0.9 to 0.975. This indicates effects where there is some evidence of an effect, but where that evidence doesn't reach the threshold of significance used in the present study. These effects can have small or moderate influence on the predicted treatment effect. Non-standardized statistics for these effect estimates can be found in Appendix B.

The obtained model contains significant regressors for most responses. Impairment, SF36 - Physical summary and Self-reported level of pain have the largest amount of significant regressors and the greatest effect from them, whereas Activity, SF36 - Mental summary, HADS - Anxiety and HADS - Depression have fewer significant regressors and a smaller effect, and Social support has no significant regressors. This indicates that the effect of treatment on dimensions of emotional functioning or mental health is more uniform across patients than its effect on physical functioning and pain. This conclusion is reinforced by the example cases in Section 4.4, where there is a very large difference in PD between best and worst cases for Impairment, SF36 - Physical summary and Self-reported level of pain, and smaller differences for other dimensions.

Certain regressors appeared more important than others by virtue of having significant effect on more responses and those effects being stronger. Periodicity of pain, unemployment and belief in improvement were regressors that had strong effects across several responses. These regressors differ in interesting ways. Periodicity of pain is a medical fact, employment status is a social condition that cannot be directly changed by treatment but can conceivably be changed in some other way, and belief in improvement is a matter of attitude that could be directly addressed

in treatment. The exact phrasing of the item concerning belief in improvement is, "How convinced are you that you will be restored" (in Swedish: "Hur övertygad är du om att bli återställd?"), which is contrary to the scope of multimodal pain rehabilitation being improved quality of life and ability to manage pain, not complete restoration. Although this question is a useful predictor of treatment outcome, it is poorly worded in the context of multimodal pain rehabilitation.

Several regressors had no significant effects. The number of visits with a doctor, having "Other" education, being a student or being outside the labor force, and being born in a Nordic country other than Sweden, showed no significant effects for any response. Gender and age only had significant effects on one response each, as did having elemental school education. Having university education had a significant effect on two responses. It is likely that a more parsimonious regression model could give similarly powerful predictions as the present model by discarding some of the non-significant regressors.

4.3.1 Social support

The Social support response produces results that are different in several ways from other responses. It is the only response with a negative mean, and the only response with no significant regressors. To understand why, it is instructive to investigate the MPI items that contribute to Social support, most of which ask the respondent to rate how often their "significant other" (defined in the questionnaire as the other person to whom the respondent feels closest) responds to them in different ways when they are in pain, where helpful, supportive or distracting actions will contribute to a higher Social support score and punishing actions will contribute to a lower score. Any other items contributing to Social support asks the respondent to rate how the pain interferes with their relationship to their significant other, their family, and about the support they receive from them. As these things are at least partially outside the control of the respondent, it seems reasonable that the treatment would not have a large effect on them and that no regressor would have a significant effect.

It is interesting that the mean change in Social support is negative. In the example cases in Section 4.4, both the best-case and worst-case patients were predicted to have negative responses for Social support, although the prediction has fairly low certainty regarding the effect direction. One reason might be that several test items contributing to Social support are difficult to interpret in terms of unidimensional health gains or losses, and might score lower as a result of improved health. For instance, stating that a significant other often gets the patient pain medication will contribute to a higher Social support score, but if a patient experiences less pain as a result of treatment, the patient might have lesser need for pain medication and thus not get it as often from their significant other, contributing to a lower Social support score.

In other cases, it is not completely clear that behaviours from significant others that contribute to a high Social support score are strictly beneficial from a health standpoint. For instance, a significant other that tries to get the patient to rest often will contribute to a high Social support score, as this is coded as a supportive behaviour, but an important part of multimodal pain rehabilitation is to reduce fear of movement and increase the level of activity.

The lack of correlation between Social support and other responses means that the model is unable to leverage other responses to give a better estimate for change in Social support. This lack of correlation is interesting and somewhat surprising, as it seems reasonable, for example, that getting less support and more punishment from one's significant other could have adverse effects on mental health. On the other hand, the countervailing effects described above, where improved health lessens the need for social support, might counteract any such effect.

Overall, Social support might not be suitable as an aggregated unidimensional measure of health, but its components might be useful in a treatment setting where a clinician can take the patient's unique situation and complexity into account.

4.3.2 Assessing assumptions on data

The BMLR model requires an assumption on the distribution of the response data. In the present work, the errors are assumed to stem from a multivariate normal distribution. The marginal plots of response values in Figure 4.7 give some indication that the response distributions are skewed. To more accurately assess whether the multivariate normality assumption holds, a Q-Q plot of squared Mahalanobis distances for each residual compared to their theoretical χ^2 quantiles was created and is displayed in Figure 4.9. From this figure, it is clear that the multivariate residual distribution is heavy-tailed in the sense that points with great Mahalanobis distance have a greater such distance than is expected from a multivariate normal distribution. As residuals are assumed to follow a multivariate t-distribution, not a multivariate normal distribution some heavy-tailedness is to be expected, but seeing as the amount of data points is very high compared to the amount of dimensions, the residuals should behave similarly to a multivariate normal distribution.

Marginal Q-Q plots of the responses are given in Figure 4.10. It is clear that the marginal response distributions have heavy tails to different extents, the most extreme being Activity and Social support, and with SF36 - Mental summary being among the least extreme. Impairment and SF36 - Physical summary display clear right-tailedness, meaning that those individuals who improve most in these responses improve more than expected from a normal distribution.

Marginal normal Q-Q plots of residuals and scatterplots of fitted values against residuals are given in Appendix C, in Figure C.2. These plots are similar to the Q-Q plots for the data, again indicating heavy-tailedness in all responses and skewness in several responses. For most residuals, the heavy-tailedness is so large that it can scarcely be explained by the fact that residuals are expected to be t-distributed, given the many degrees of freedom. The scatterplots of fitted values against residuals in Figure C.1 show no clear signs of heteroskedasticity when accounting for the varying amount of points at different levels of the fitted values.

Another assumption that is made in the present model choice is that the responses have correlated errors for an individual. Strictly speaking, the BMLR model works even if all responses are uncorrelated, but there is no advantage gained from the increased model complexity in this case, as opposed to modelling different responses as univariate. To investigate whether this assumption is reasonable, the sample correlation for the responses is displayed in Figure 4.11. From this figure, it is clear that most responses have moderate to strong correlation with several

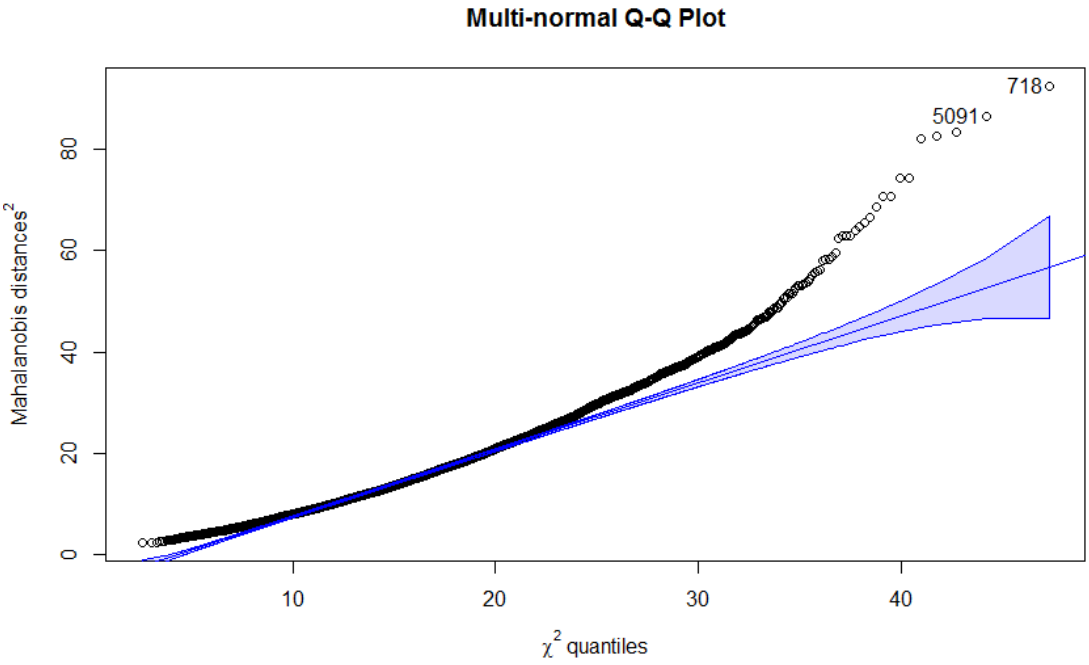


Figure 4.9: Q-Q plot of squared Mahalanobis distance of residuals against theoretical quantiles

other responses, the main exception being Social support, which has no correlation absolutely greater than 0.1 with any other response.

4.4 Examples and illustrations

The regression model shows that there are several significant predictors that affect the expected outcome of the treatment in several dimensions. However, it is not immediately interpretable how impactful these predictors are relative to the uncertainty of the treatment outcome. In a classical linear regression model, a measure such as the coefficient of determination R^2 could be used to measure this impact, but that measure does not work in the Bayesian setting.

To illustrate the size of the model effects relative to random error, the model's predictions in extreme examples were compared. Two patients from the data set were chosen to represent a "worst-case" and "best-case" patient under the model. To do this, the significant effects in the model were inspected and a profile of categorical regressors that would produce maximal or minimal estimated change in all dimensions was obtained. The worst-case patient would be male, born in Europe outside of the Nordic countries, have elementary school education, be unemployed, have very low belief in improvement and have constant pain as opposed to periodic pain. The best-case patient would be female, born in Sweden or another Nordic country, have university education, not be unemployed, have very high belief in improvement and have periodic pain as opposed to constant pain.

Two sets of patients fitting these criteria were obtained. There were 60 patients in the dataset fitting the best-case criteria but only one fitting the worst-case criteria. To select a final best-case example, the best-case candidate with the lowest number of

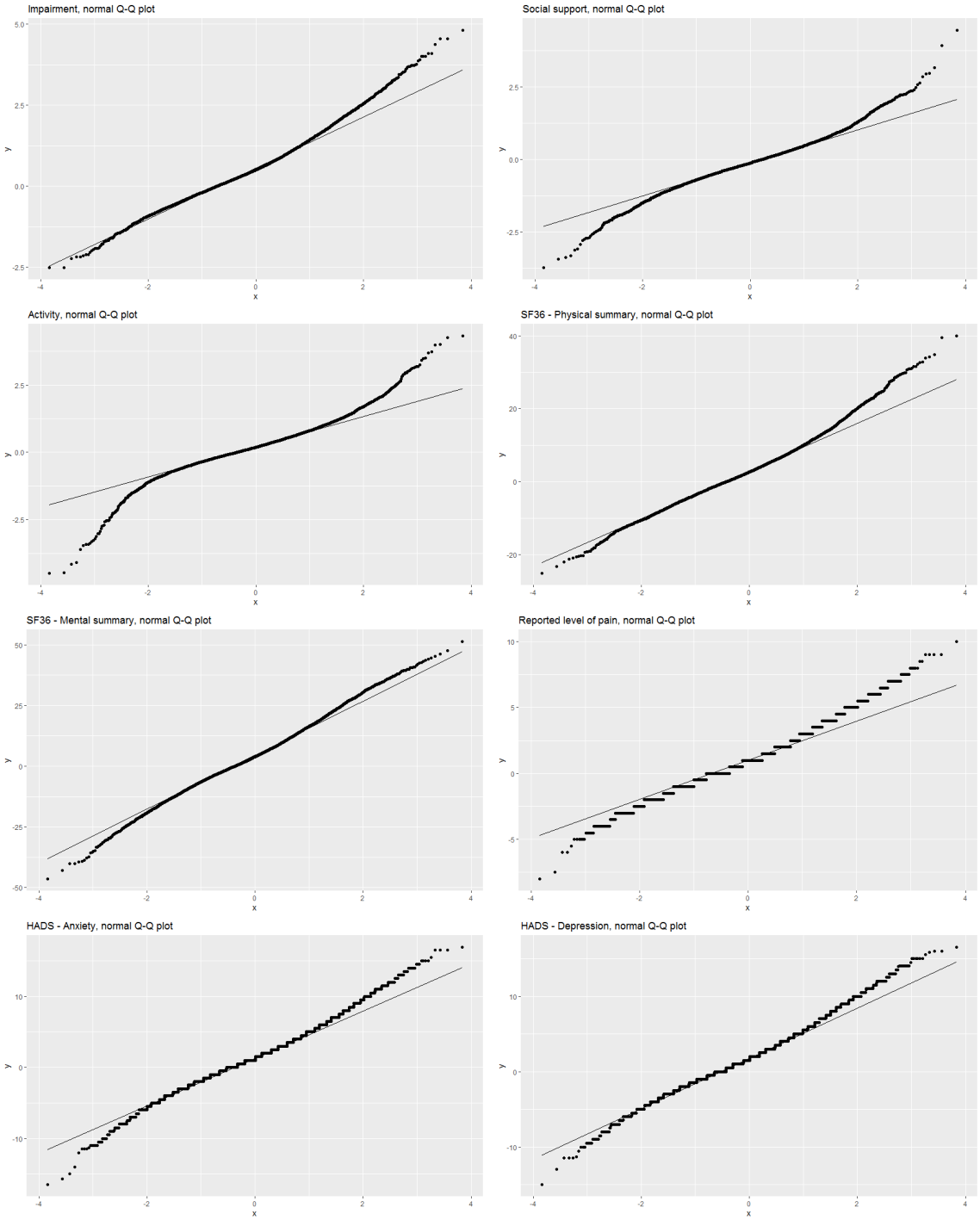


Figure 4.10: Marginal normal Q-Q plots for responses

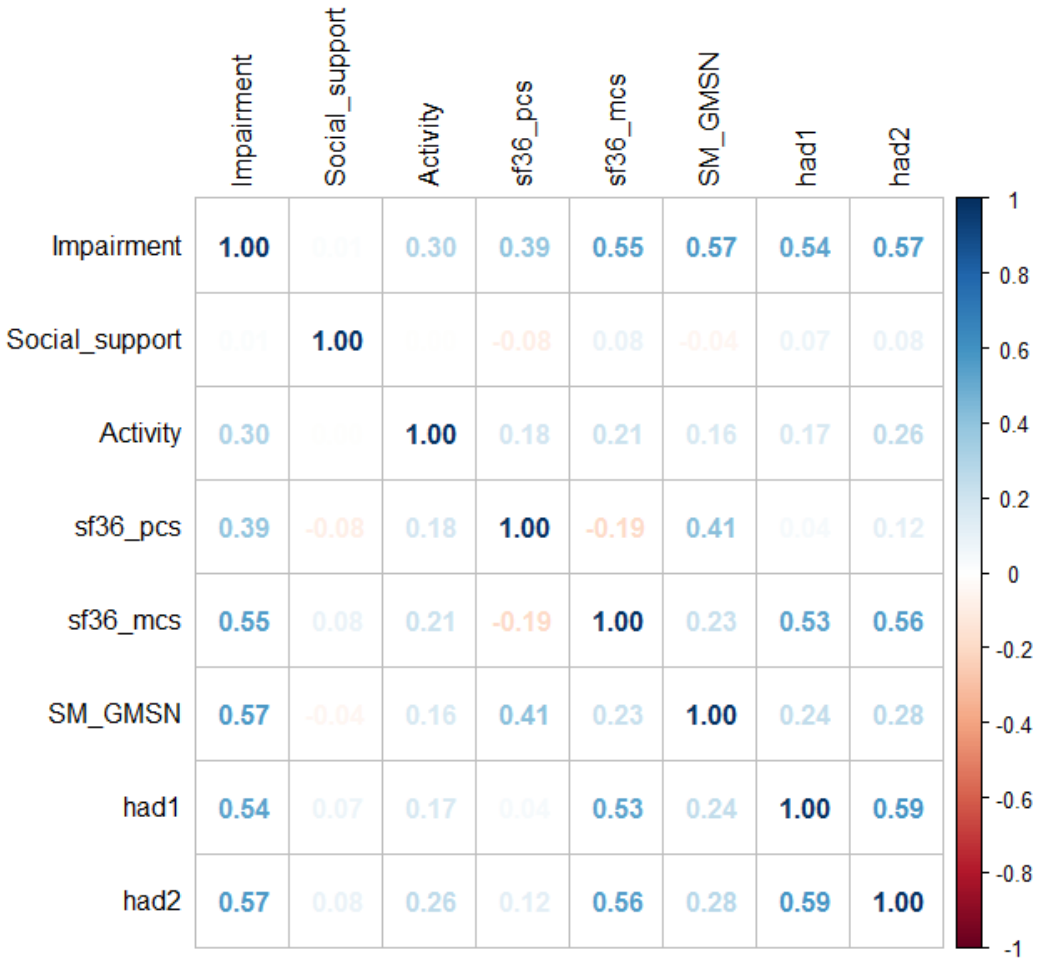


Figure 4.11: Correlation between responses

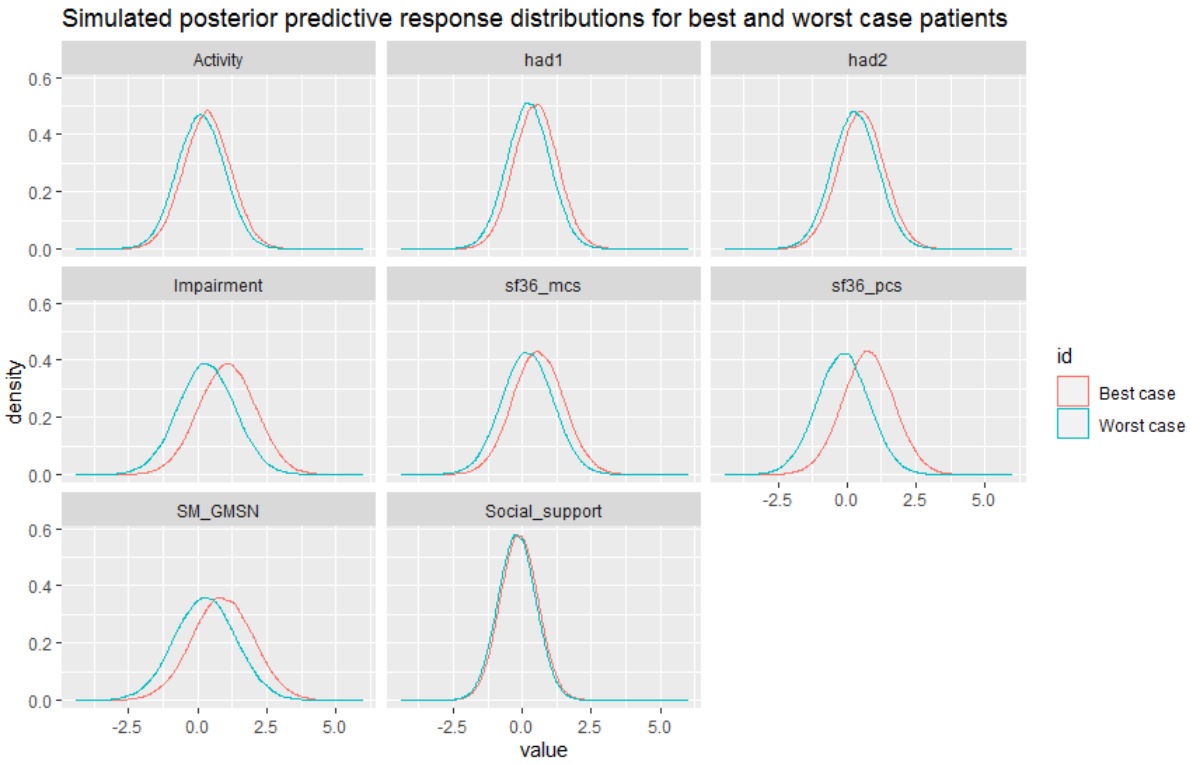


Figure 4.12: Posterior predictive distributions for all responses (standardized responses)

days since their pain debut was selected, since this variable had significant negative effect on two results, whereas the other significant continuous variable, age, had a significant negative effect on just one result.

Using these patients, the posterior predictive distributions for their treatment outcomes were simulated by simulating coefficient values from the posterior distribution of the coefficients and errors from the posterior distribution of the error. 100000 simulated values were obtained. The posterior predictive densities were estimated from this simulation and are displayed in Figure 4.12. Medians, 95% credible intervals and probability of direction for all responses are displayed in Figure 4.13. This figure also contains the actual observed changes in responses for these patients, in the column "Result". For the purposes of illustration the standardized response measures were used rather than raw scores.

In these extreme cases, certain responses display a clear difference depending on the case and others do not. The posterior predictive distribution for Social support is very similar in both cases, with the medians of both distributions being slightly negative, the 95% credible interval boundaries being very close and a slightly higher probability of direction for the worst case (0.611 for the worst case as compared to 0.565 for the best case). The posterior predictive distributions for HAD-Anxiety, HAD-Depression and Activity are also similar in both cases, with positive medians in both cases but greater PD by about 0.1 for the best case. SF36-mental summary displays a slightly larger distinction between cases, with positive medians in both cases but a larger distinction between them, and PD of 0.577 in the worst case as compared to 0.733 in the best case.

Figure 4.13: Response statistics for example patients (standardized responses)

	Median	CI_low	CI_high	pd	Result
Impairment	0.3121663	-1.703771	2.311892	0.62091	0.1436335
Social_support	-0.1942116	-1.537480	1.127124	0.61060	1.4111602
Activity	0.1143338	-1.521393	1.774544	0.55407	-0.3908235
sf36_pcs	-0.1485450	-2.005450	1.676968	0.56265	-0.1560687
sf36_mcs	0.1804136	-1.637081	1.965072	0.57724	0.2913156
SM_GMSN	0.2744412	-1.855065	2.461418	0.59831	-0.1415086
had1	0.2259992	-1.329152	1.737332	0.61454	0.9570737
had2	0.2833957	-1.318692	1.921933	0.63417	1.2382487

(a) Statistics for worst-case patient

	Median	CI_low	CI_high	pd	Result
Impairment	1.0725351	-0.8975392	3.102464	0.85128	3.3264382
Social_support	-0.1130304	-1.4679259	1.210470	0.56482	-0.5855279
Activity	0.3389991	-1.3278378	1.978780	0.65806	1.1701020
sf36_pcs	0.7600932	-1.0567647	2.612619	0.79285	2.0442650
sf36_mcs	0.5676513	-1.2123650	2.388112	0.73262	1.2424647
SM_GMSN	0.9013621	-1.2259557	3.087582	0.79563	2.7995655
had1	0.4929036	-1.0162169	2.047002	0.73368	1.0644015
had2	0.5133942	-1.1078368	2.143649	0.73352	1.6953646

(b) Statistics for best-case patient

The differences are larger in the posterior predictive distributions for Impairment and reported level of pain. Both medians are positive for both cases, but the PD is larger by 0.197 for reported level of pain and 0.230 for Impairment, with large differences in medians and 95% CI bounds. The difference is largest for SF36-physical summary, which is the only response where the median in the worst and best cases have different signs. The worst-case patient is expected to worsen by a small amount after treatment with PD of 0.563, and the best-case patient is expected to score better after treatment with PD of 0.793.

The fact that the treatment is not very likely to help worst-case patients is interesting. A worst-case patient has a lower than 64% chance to see improvement in any single response, and is expected to worsen in Social support and SF36 - Physical summary. It should be noted that bad cases were not rare in the dataset, as for some important regressors with significant negative effects across responses (belief in recovery, periodicity of pain) the worst-case value was typical among respondents.

5 Discussion

5.1 Limitations

An issue with the present study is the lack of a control group. As the SQRP only contains initial data for patients who did not receive treatment, there is no way to compare health outcomes for sufferers of chronic pain who received treatment with those who did not. This leads to two main issues:

- Predicting which patients will receive the greatest benefits from treatment is not possible without knowing what their health outcomes would be given no treatment. Hypothetically, those patients who see the greatest improvement from treatment could see similar improvement without treatment while those who receive the least benefit from treatment could see greatly worsened health without treatment. Without a control group, this is difficult to investigate.

- The lack of a control group makes it difficult to assess the effect of baseline values for outcome measures on the change in those measures. Previous studies indicate that baseline values of outcome measures have the largest effect of any regressor in a regression model, but the uncertain interpretation of this effect led to their exclusion in the present study. With a control group, it would be possible to disentangle the effect of baseline values on the treatment result from the effect of the error by comparing the treatment and control groups.

The issue of health outcomes in chronic pain patients who do not receive treatment could be attacked by consulting the literature. However, a systematic review by the Swedish agency for health technology assessment and assessment of social services (SBU) found that while the literature indicates that patients undergoing multimodal treatment improve in time, it cannot be ascertained whether this is a direct result of the interventions or contingent on other factors [SBU21]. Thus, further study on health outcomes of chronic pain sufferers who do not get treatment are necessary before a model such as this can be used to predict what patients would receive the largest benefits from the treatment.

5.2 Data preprocessing and modelling

In the course of data processing, several choices were made that could affect modelling outcomes. No item imputation was performed, and incomplete responses were dropped. This choice is likely to have biased the model results, since the excluded responses saw significantly worse results from treatment for most responses, as shown in Section 3.3.1. The model thus likely overestimates the benefits of treatment

slightly for some of the responses, and any conclusions drawn from the present study should take this into consideration. Using some common item imputation methods, as discussed in Section 3.3.1, would likely introduce other issues that might be harder to account for, such as over- or underestimating the significance of the effects due to performing regression on synthetic data. Still, attempting these methods, or more sophisticated imputation methods, would be an interesting development of the modelling approach of the present study. One such possibility would be to expand the Bayesian framework of the model by utilizing a Gibbs sampling technique. During simulation of the posterior, a Gibbs sampler could alternate between sampling missing values as they are predicted using the current model parameters, and simulating from the posterior of the model using the full data.

The choice to use the average value of the two time steps after treatment as the response value could be further investigated. To the author's knowledge, no research on whether treatment effects are permanent or change with time has been performed, and it would be of great interest to gain insight into this. If there is in fact some change with time, even if only for certain responses, this would make a different method of choosing response values preferable. Such findings would also be interesting in and of themselves, as long-term improvement is a highly desirable result of the treatment.

From analysis of model residuals there is strong evidence that the true errors in the data are not, in fact, normally distributed. Fitting a model assuming more heavy-tailed error distributions, such as a multivariate t-distributed error, could be of interest. For some outcomes, residual distributions appear to be particularly right-tailed. In these cases, some transformation of the data could also be of interest. Note that the fact that error distributions are more heavy-tailed than assumed is unlikely to impact the direction of the estimated effects. However, it does impact the level of certainty in the model.

In the model, most regressors are categorical but some are continuous. The modelling approach carries an assumption that any one step of these categorical regressors has equal effect to any other step. Particularly for the regressor "Amount of days since pain debut", this might be an unfounded assumption, as it seems reasonable that the difference between having had chronic pain for one or two years would have a larger impact than the difference of having had chronic pain for 19 or 20 years. A logarithmic transformation of this regressor might have given a better outcome. In general, investigating the possible existence of non-linear effects would be a sensible next step after the current model.

5.3 Future research

One of the main features of the Bayesian statistical approach is the ability to encode pre-existing knowledge into prior distributions. The present study used uninformative priors, which is an attempt to minimize the influence of the prior on the posterior distribution. This choice is not uncommon as a first step in investigating data, but it is not necessarily ideal. Investigating the results of using different priors and attempting to find reasonable proper priors could improve the confidence of the model.

Many chronic pain studies ask questions regarding health economics, and thus measure outcomes such as reduced sick leave expenses. As these outcomes are

not, strictly speaking, health outcomes, the present study has disregarded them - however, the modelling approach utilized in the present study is suitable in principle for investigating questions of health economy. The simplest outcome measure to add to the model to provide a health economic perspective would be sick leave status.

The present study has also not utilized sick leave status as a regressor. Sick leave status could be an indicator of severity of illness, and could be an interesting factor in explaining treatment outcomes. Of particular interest would be the interaction between employment status and sick leave status, as being unemployed had a high impact on treatment outcomes.

Finding a way to assess the health outcomes of untreated sufferers of chronic pain would be of great interest. This would give the ability to make predictions of what patients would see the greatest benefit from treatment, and thus make the model usable as a decision aid in the clinical setting. It would also allow investigation of whether baseline levels of outcome variables are, in fact, the most important regressors, or whether such results merely indicate regression to the mean. The gold standard way of obtaining the desired data would be a clinical trial, which would likely have a smaller sample size and might thus be unsuitable for a highly complex model. Investigating the registry data on those patients who were evaluated for treatment several times but did not receive it might be a way to approximate health outcomes of untreated sufferers of chronic pain, but this approach comes with issues of representativity.

5.4 Conclusion

This study shows several significant effects on background variables on treatment outcomes from MMR in various health dimensions for patients suffering from chronic pain. Previous findings on dimension reduction of the MPI instrument are replicated on a different and larger set of patients. The potential predictive power of a model for treatment outcomes is demonstrated. For the purposes of prediction, more knowledge is needed on health outcomes of chronic pain sufferers who do not receive treatment.

A Regression tables

This appendix contains lists of non-instrument variables in the SQRP. All psychometric instruments that are included in the dataset are listed in Section 2.2 Some questions may be omitted due to being purely technical or administrative in nature. This appendix is based on a legacy variable list corresponding to the epoch of SQRP data the study is based on. Due to the SQRP questionnaire changing over time, even within an epoch, there is some mismatch between the variable list and the dataset, meaning that some variables in the dataset may be omitted here (notably age, which is present in the dataset).

Variable name	Explanation	Values
DATAVSLUT	Treatment end date	Date
DIAGNOS1	Primary diagnosis	Diagnosis codes
DIAGNOS2-4	Secondary diagnoses	Diagnosis codes
DIAGNOSX	ICD-10 diagnosis for eventual outer cause of injury	Diagnosis codes
FILLOUT	Patient has filled out questionnaire	Yes/No
FVDFOR	Expected future sustained source of income - partially unpredictable	0%, 25%, 50%, 75%, 100%
FVEJFOR	Expected future sustained source of income - wholly unpredictable	0%, 25%, 50%, 75%, 100%
FVFORP	Expected future sustained source of income - Sickness compensation/activity compensation	0%, 25%, 50%, 75%, 100%
FVLON	Expected future sustained source of income - Salary	0%, 25%, 50%, 75%, 100%
IDIOPAT	Contribution of pain of unclear origin to overall pain	1-4 (1 greatest contribution)
NEUR_CEN	Contribution of central neurogenic pain to overall pain	1-4 (1 greatest contribution)
NEUR_PER	Contribution of peripheral neurogenic to overall pain	1-4 (1 greatest contribution)
NOCICEPT	Contribution of nociceptive pain to overall pain	1-4 (1 greatest contribution)
PSYKOGEN	Contribution of psychogenous pain to overall pain	1-4 (1 greatest contribution)

Variable name	Explanation	Values
MALKATEG	Patient category	1: Only for evaluation 2: Evaluation and follow-up 3: Evaluation and rehabilitation 4: Other, follow-up 5: Other, with follow-up
MCEID	Patient identifier	ID number
N101	Have you been treated respectfully and considerately?	0: Yes 1: Partially 2: No
N102	Were you satisfied with caregiver facilities?	1: Yes 0: No
N103	How was the information you received on practical matters during treatment?	0: Very good 1: Good 2: Satisfactory 3: Not very good 4: Bad
N104	Do you feel included in planning your rehabilitation to the extent that you wish?	0: Yes, fully 1: Partially 2: No
N105	How do you feel the rehab team cooperated in your case?	0: Very well 1: Well 2: Satisfactorily 3: Not very well 4: Poorly
N106	Did the people you are close to participate in your rehabilitation?	0: Yes 1: No, I had no one who could participate 2: No, I didn't want them to participate 3: No, they weren't offered to participate
N107	Has your rehab period changed your pain experience?	0: Greatly decreased pain experience 1: Somewhat decreased pain experience 2: No effect on pain experience 3: Somewhat increased pain experience 4: Greatly increased pain experience
N108	Has your rehab period changed your ability to manage your life situation in general?	0: Greatly improved 1: Somewhat improved 2: No change 3: Somewhat worsened 4: Greatly worsened
N211	Rehab planning	0: No 1: Yes

Variable name	Explanation	Values
ORSAVS	Reason for treatment ending/stoppage	1: Planned rehab completed 2: Stoppage due to need for other medical contact 3: Patient decides to stop 4: Other reasons
REKNYARB	Recommendation for return to previous or new work	Same employer, same tasks 1: Same employer but changed tasks 2: New employer, same tasks 3: New employer, changed tasks 4: Cannot be predicted
SV_SPRAK	Ability to speak Swedish	1: No difficulty 3: Misinterpretations occur 5: Dependent on interpreter
ARB_FORM	Employment status	Employed/self employed, Unemployed, Outside the labor force, Student
ARB_HELTID	I work/study 100%	1: Yes 0: No
IARBETE	Are you currently employed or studying to any extent (Followup time step)	0: No 1: Yes
ARB_OMFATTNING	If YES above - to what extent do you work/study?	Percentage, rounded to whole point
F3AKASSA	Source of income - Unemployment insurance or activity support	0%, 25%, 50%, 75%, 100%
F3FALDRP	Source of income - Parental benefit	0%, 25%, 50%, 75%, 100%
F3FORP	Source of income - Sickness compensation/activity compensation	0%, 25%, 50%, 75%, 100%
F3LON	Source of income - salary	0%, 25%, 50%, 75%, 100%
F3OVRIGT	Source of income - Other (followup time step)	0%, 25%, 50%, 75%, 100%

Variable name	Explanation	Values
F3SJBIDR	Source of income - Time limited sickness compensation/activity compensation	0%, 25%, 50%, 75%, 100%
F3SJPENG	Source of income - Sickness benefit/rehab benefit	0%, 25%, 50%, 75%, 100%
F3SOCBID	Source of income - subsistence allowance	0%, 25%, 50%, 75%, 100%
F3UTBILD	Source of income - Student aid via CSN	0%, 25%, 50%, 75%, 100%
N71	How many times have you visited a doctor for your pain in the past year	0: 0-1 times 1: 2-3 times 2:4 or more times
summons_sent_date	The date the questionnaire was sent to the patient	Date
WORKACTIVITY	I'm work training or trying other activities to regain work ability (at Followup)	1: Yes 0: No 2: Work training
ANKOMAR	What year did you come to Sweden	Year
ARB_VARD	What importance does work have to you beyond being a source of income	1: Very large 2: Large 3: Some 4: Barely any 5: None
CONCENT	Patient consents to being registered	0: No 1: Yes
DAT1ERBJ	Date of first visit offered	Date
DAT1REEL	Date of first actual visit	Date
DatSMIHA	If pain is continuous, since when?	Date
DATSMINT	When did you first feel the pain that is now troubling you?	Date

Variable name	Explanation	Values
DAT_IARB	When were you last employed or studied?	Date
F_LAND	Country of birth	0: Sweden 1: Nordic countries 2: Europe 3: Other
INREMITT	Original point of referral	1: Inpatient care or hospital 2: Primary care 3: Company care 4: Private specialt 5: Social Insurance Agency 6: Employment Agency 7: Employer 8: Direct applicant 9: Other
RINSTBLI	How convinced are you that you will be restored?	1-5 (1 completely certain)
RINSTHUR	What do you think it will be like to return to work, to start studying or to expand your working hours?	1-5 (1 very easy)
RINSTNAR	When do you expect to be able to return to work, begin studying or expand your working hours?	1-5 (1 immediately, 5 never)
SIHVDLOK	In what part of your body is your worst pain located?	1: Head or face 2: Throat or neck 3: Shoulder or arm 4: Chest 5: Abdomen 6: Genitals and crotch 7: Upper back 8: Lower back 9: Hip 10 Leg
SMPERIOD	Is your pain periodic or constant?	1: Periodic 2: Constant
UTB_NIVA	Highest level of completed education	1: Elementary school 2: Upper secondary school 3: University 4: Other

B Residual analysis - figures

This appendix contains tables with information on estimated regression coefficients. Each table contains information on all regressors as they pertain to one particular response. For each regressor-response pair, the median estimated coefficient is reported, as are the bounds of a 95% credible interval for the coefficient and its probability of direction. Regressors with PD greater than 0.975 (corresponding to significance at the 95% level) are highlighted in green, and regressors with PD greater than 0.9 but lower than 0.975 are highlighted in yellow.

	Median	CI_low	CI_high	pd
Impairment.GENDER	0.0301209	-0.0345846	0.0930842	0.823090
Impairment.d_age	-0.0023125	-0.0049551	0.0003487	0.956298
Impairment.F_LAND_NOR	0.0676635	-0.0906699	0.2263434	0.798521
Impairment.F_LAND_EUR	-0.0397455	-0.1706365	0.0895349	0.725302
Impairment.F_LAND_OTH	-0.0345622	-0.1320808	0.0624047	0.757997
Impairment.UTB_NIVA_ELE	-0.0819700	-0.1682676	0.0053427	0.967763
Impairment.UTB_NIVA_UNI	0.0635080	0.0019325	0.1253666	0.978197
Impairment.UTB_NIVA_OTH	-0.0183492	-0.1339759	0.0970978	0.621936
Impairment.ARB_FORM_ARBETSSOKANDE	-0.1280802	-0.2022934	-0.0548573	0.999676
Impairment.ARB_FORM_STUDERANDE	-0.0004122	-0.1603626	0.1598390	0.502015
Impairment.ARB_FORM_EJARBETANDE	-0.0653740	-0.1764753	0.0442620	0.877705
Impairment.d_dag_smint	-0.0000088	-0.0000172	-0.0000004	0.979604
Impairment.RINSTBLI	-0.0353267	-0.0571352	-0.0138809	0.999342
Impairment.SMPERIOD_NY1	-0.0906048	-0.1649766	-0.0153489	0.991201
Impairment.ini_N71	0.0048594	-0.0373463	0.0474214	0.589054
Impairment.intercept	0.9962682	0.7977054	1.1907432	1.000000

Regression for Impairment

	Median	CI_low	CI_high	pd
Social_support.GENDER	0.0487361	-0.0007935	0.0987579	0.972653
Social_support.d_age	0.0000906	-0.0019762	0.0021680	0.533997
Social_support.F_LAND_NOR	0.0194508	-0.1038062	0.1438007	0.620900
Social_support.F_LAND_EUR	-0.0304903	-0.1316050	0.0715682	0.721897
Social_support.F_LAND_OTH	-0.0051686	-0.0814470	0.0702779	0.553412
Social_support.UTB_NIVA_ELE	0.0017694	-0.0657864	0.0699658	0.520359
Social_support.UTB_NIVA_UNI	-0.0246102	-0.0728097	0.0237388	0.841849
Social_support.UTB_NIVA_OTH	-0.0566962	-0.1457225	0.0348384	0.890881
Social_support.ARB_FORM_ARBETSSOKANDE	-0.0335017	-0.0917570	0.0235010	0.873005
Social_support.ARB_FORM_STUDERANDE	-0.0295462	-0.1539599	0.0963330	0.678274
Social_support.ARB_FORM_EJARBETANDE	0.0153186	-0.0703739	0.1018424	0.635882
Social_support.d_dag_smint	0.0000046	-0.0000020	0.0000111	0.915560
Social_support.RINSTBLI	0.0127553	-0.0040656	0.0296316	0.930863
Social_support.SMPERIOD_NY1	-0.0036736	-0.0622670	0.0544332	0.549476
Social_support.ini_N71	-0.0311625	-0.0645705	0.0017449	0.967273
Social_support.intercept	-0.1415916	-0.2954797	0.0109090	0.965048

Regression for Social support

	Median	CI_low	CI_high	pd
Activity.GENDER	0.0176045	-0.0334337	0.0692637	0.748612
Activity.d_age	-0.0009707	-0.0030888	0.0011859	0.813119
Activity.F_LAND_NOR	-0.0184874	-0.1465248	0.1091329	0.611435
Activity.F_LAND_EUR	-0.0078802	-0.1127315	0.0974162	0.558981
Activity.F_LAND_OTH	0.0143436	-0.0644078	0.0923973	0.639662
Activity.UTB_NIVA_ELE	-0.0304004	-0.1006278	0.0391704	0.803015
Activity.UTB_NIVA_UNI	0.0220629	-0.0278227	0.0717162	0.807304
Activity.UTB_NIVA_OTH	-0.0027829	-0.0951472	0.0911587	0.523514
Activity.ARB_FORM_ARBETSSOKANDE	-0.0598339	-0.1186850	0.0002359	0.975751
Activity.ARB_FORM_STUDERANDE	0.0736082	-0.0556800	0.2020473	0.868135
Activity.ARB_FORM_EJARBETANDE	-0.0173493	-0.1057233	0.0717475	0.648774
Activity.d_dag_smint	-0.0000131	-0.0000199	-0.0000064	0.999929
Activity.RINSTBLI	-0.0085171	-0.0260413	0.0088257	0.831611
Activity.SMPERIOD_NY1	-0.0031037	-0.0629907	0.0573414	0.540280
Activity.ini_N71	0.0255354	-0.0087610	0.0596998	0.928335
Activity.intercept	0.2792323	0.1199967	0.4363959	0.999690

Regression for Activity

	Median	CI_low	CI_high	pd
sf36_pcs.GENDER	0.0988102	-0.4435638	0.6454817	0.638867
sf36_pcs.d_age	0.0041708	-0.0184319	0.0267543	0.641445
sf36_pcs.F_LAND_NOR	-0.0882137	-1.4538870	1.2499354	0.551151
sf36_pcs.F_LAND_EUR	-1.4482525	-2.5522100	-0.3329767	0.994751
sf36_pcs.F_LAND_OTH	-0.9158889	-1.7409378	-0.0848501	0.984966
sf36_pcs.UTB_NIVA_ELE	-0.7731512	-1.5048769	-0.0291267	0.979957
sf36_pcs.UTB_NIVA_UNI	0.1270705	-0.4021107	0.6508014	0.681607
sf36_pcs.UTB_NIVA_OTH	0.2272325	-0.7637988	1.2073850	0.674490
sf36_pcs.ARB_FORM_ARBETSSOKANDE	-1.3888811	-2.0134739	-0.7567457	0.999992
sf36_pcs.ARB_FORM_STUDERANDE	0.4877951	-0.8848507	1.8427944	0.757951
sf36_pcs.ARB_FORM_EJARBETANDE	-0.5320896	-1.4768134	0.4003819	0.866357
sf36_pcs.d_dag_smint	-0.0000504	-0.0001217	0.0000212	0.916464
sf36_pcs.RINSTBLI	-0.5968288	-0.7790953	-0.4104826	1.000000
sf36_pcs.SMPERIOD_NY1	-0.4645323	-1.1016367	0.1702495	0.923853
sf36_pcs.ini_N71	-0.1364837	-0.4974189	0.2250940	0.770470
sf36_pcs.intercept	6.5699690	4.8991599	8.2464810	1.000000

Regression for SF36 - Physical summary

	Median	CI_low	CI_high	pd
sf36_mcs.GENDER	0.6224006	-0.2762367	1.5181622	0.912877
sf36_mcs.d_age	-0.0315820	-0.0690262	0.0056334	0.951688
sf36_mcs.F_LAND_NOR	0.8977980	-1.3394429	3.1175525	0.785659
sf36_mcs.F_LAND_EUR	0.0817314	-1.7377601	1.9302637	0.534825
sf36_mcs.F_LAND_OTH	0.0101655	-1.3494980	1.3873431	0.505767
sf36_mcs.UTB_NIVA_ELE	-0.7787629	-2.0058704	0.4303368	0.894493
sf36_mcs.UTB_NIVA_UNI	0.7427103	-0.1276424	1.6127084	0.952662
sf36_mcs.UTB_NIVA_OTH	0.0151806	-1.6216004	1.6252353	0.507370
sf36_mcs.ARB_FORM_ARBETSSOKANDE	-0.1794499	-1.2143859	0.8599768	0.632563
sf36_mcs.ARB_FORM_STUDERANDE	-0.1209745	-2.3748761	2.1311828	0.542255
sf36_mcs.ARB_FORM_EJARBETANDE	-0.1070416	-1.6605225	1.4370830	0.553717
sf36_mcs.d_dag_smint	-0.0000142	-0.0001325	0.0001031	0.593778
sf36_mcs.RINSTBLI	-0.2290551	-0.5315382	0.0763606	0.929930
sf36_mcs.SMPERIOD_NY1	-1.6738294	-2.7206676	-0.6156199	0.999147
sf36_mcs.ini_N71	0.1367838	-0.4606619	0.7328077	0.673191
sf36_mcs.intercept	9.0159564	6.2898652	11.8164814	1.000000

Regression for SF36 - Mental summary

	Median	CI_low	CI_high	pd
SM_GMSN.GENDER	0.0272663	-0.1130076	0.1680004	0.648705
SM_GMSN.d_age	0.0021246	-0.0037341	0.0079584	0.761790
SM_GMSN.F_LAND_NOR	0.0060976	-0.3459112	0.3511111	0.513629
SM_GMSN.F_LAND_EUR	-0.0595648	-0.3465249	0.2263020	0.658149
SM_GMSN.F_LAND_OTH	0.1658412	-0.0476466	0.3793686	0.935839
SM_GMSN.UTB_NIVA_ELE	-0.0409705	-0.2313414	0.1501110	0.663394
SM_GMSN.UTB_NIVA_UNI	0.0550457	-0.0806870	0.1911412	0.786300
SM_GMSN.UTB_NIVA_OTH	0.0165131	-0.2374903	0.2708852	0.550580
SM_GMSN.ARB_FORM_ARBETSSOKANDE	-0.2818068	-0.4431571	-0.1180917	0.999622
SM_GMSN.ARB_FORM_STUDERANDE	-0.0842721	-0.4331948	0.2699845	0.680454
SM_GMSN.ARB_FORM_EJARBETANDE	-0.0927054	-0.3357608	0.1489375	0.773587
SM_GMSN.d_dag_smint	-0.0000097	-0.0000281	0.0000088	0.848258
SM_GMSN.RINSTBLI	-0.1000098	-0.1477389	-0.0527323	0.999977
SM_GMSN.SMPERIOD_NY1	-0.2005499	-0.3645612	-0.0357932	0.991725
SM_GMSN.ini_N71	-0.0101002	-0.1040825	0.0829625	0.583981
SM_GMSN.intercept	1.6954056	1.2662433	2.1299094	1.000000

Regression for self-reported level of pain

	Median	CI_low	CI_high	pd
had1.GENDER	0.0453612	-0.2312598	0.3162120	0.627165
had1.d_age	-0.0118741	-0.0233740	-0.0005971	0.979625
had1.F_LAND_NOR	0.0046221	-0.6795894	0.6795081	0.505164
had1.F_LAND_EUR	0.2977619	-0.2616736	0.8556198	0.852465
had1.F_LAND_OTH	0.2220276	-0.1899305	0.6447316	0.851823
had1.UTB_NIVA_ELE	-0.3529179	-0.7242423	0.0214218	0.968184
had1.UTB_NIVA_UNI	0.2337836	-0.0306025	0.4985224	0.958670
had1.UTB_NIVA_OTH	-0.0570643	-0.5501689	0.4400066	0.589724
had1.ARB_FORM_ARBETSSOKANDE	-0.1972489	-0.5151144	0.1169664	0.889015
had1.ARB_FORM_STUDERANDE	-0.2824842	-0.9779643	0.3957158	0.790339
had1.ARB_FORM_EJARBETANDE	0.1740442	-0.2974599	0.6488071	0.764081
had1.d_dag_smint	-0.0000048	-0.0000409	0.0000309	0.603971
had1.RINSTBLI	0.0232364	-0.0703262	0.1150860	0.687870
had1.SMPERIOD_NY1	-0.5172683	-0.8369058	-0.1955902	0.999197
had1.ini_N71	-0.0424507	-0.2233409	0.1410764	0.675923
had1.intercept	2.8810326	2.0300566	3.7165813	1.000000

Regression for HADS - Anxiety

	Median	CI_low	CI_high	pd
had2.GENDER	0.4281488	0.1555974	0.6978862	0.999003
had2.d_age	-0.0063521	-0.0176746	0.0048767	0.864941
had2.F_LAND_NOR	0.1266784	-0.5409391	0.8066822	0.644029
had2.F_LAND_EUR	-0.1166437	-0.6674362	0.4373058	0.660409
had2.F_LAND_OTH	-0.2862075	-0.7038202	0.1233113	0.914219
had2.UTB_NIVA_ELE	-0.1762524	-0.5498673	0.1863577	0.826212
had2.UTB_NIVA_UNI	0.2826804	0.0199610	0.5446668	0.982620
had2.UTB_NIVA_OTH	-0.0233425	-0.5112251	0.4683888	0.537266
had2.ARB_FORM_ARBETSSOKANDE	-0.1365035	-0.4464577	0.1794021	0.803923
had2.ARB_FORM_STUDERANDE	-0.3190479	-0.9902755	0.3674656	0.822089
had2.ARB_FORM_EJARBETANDE	0.1073187	-0.3582090	0.5784622	0.673711
had2.d_dag_smint	0.0000090	-0.0000265	0.0000448	0.689971
had2.RINSTBLI	0.0870565	-0.0050436	0.1785332	0.968430
had2.SMPERIOD_NY1	-0.2518017	-0.5699564	0.0648112	0.939797
had2.ini_N71	0.0568821	-0.1238988	0.2372280	0.731590
had2.intercept	1.9105395	1.0694078	2.7371736	0.999994

Regression for HADS - Depression

C SQRP non-instrument variables

This appendix contains plots of residuals and normal Q-Q plots of residuals for all responses.

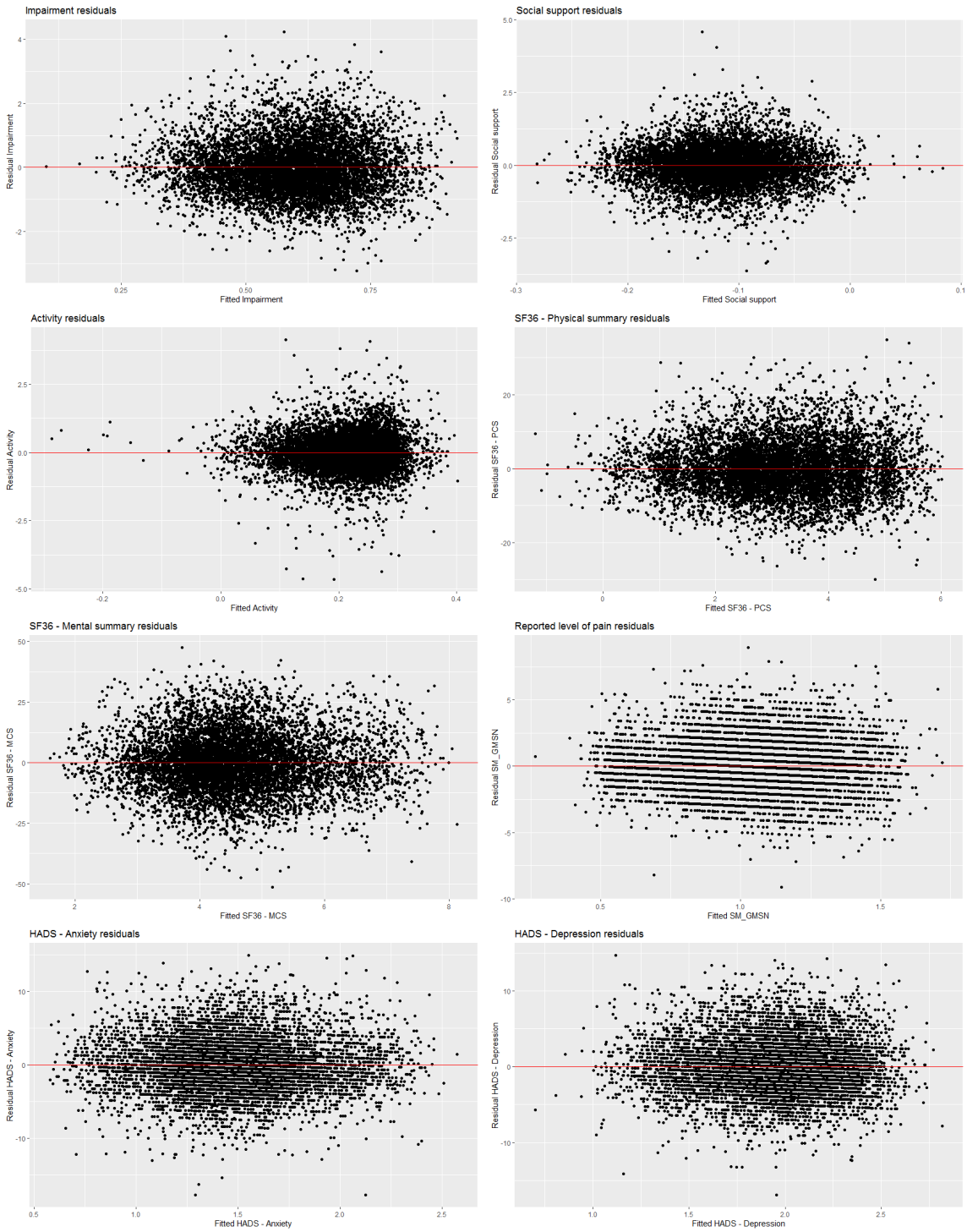


Figure C.1: Residuals of modelled responses

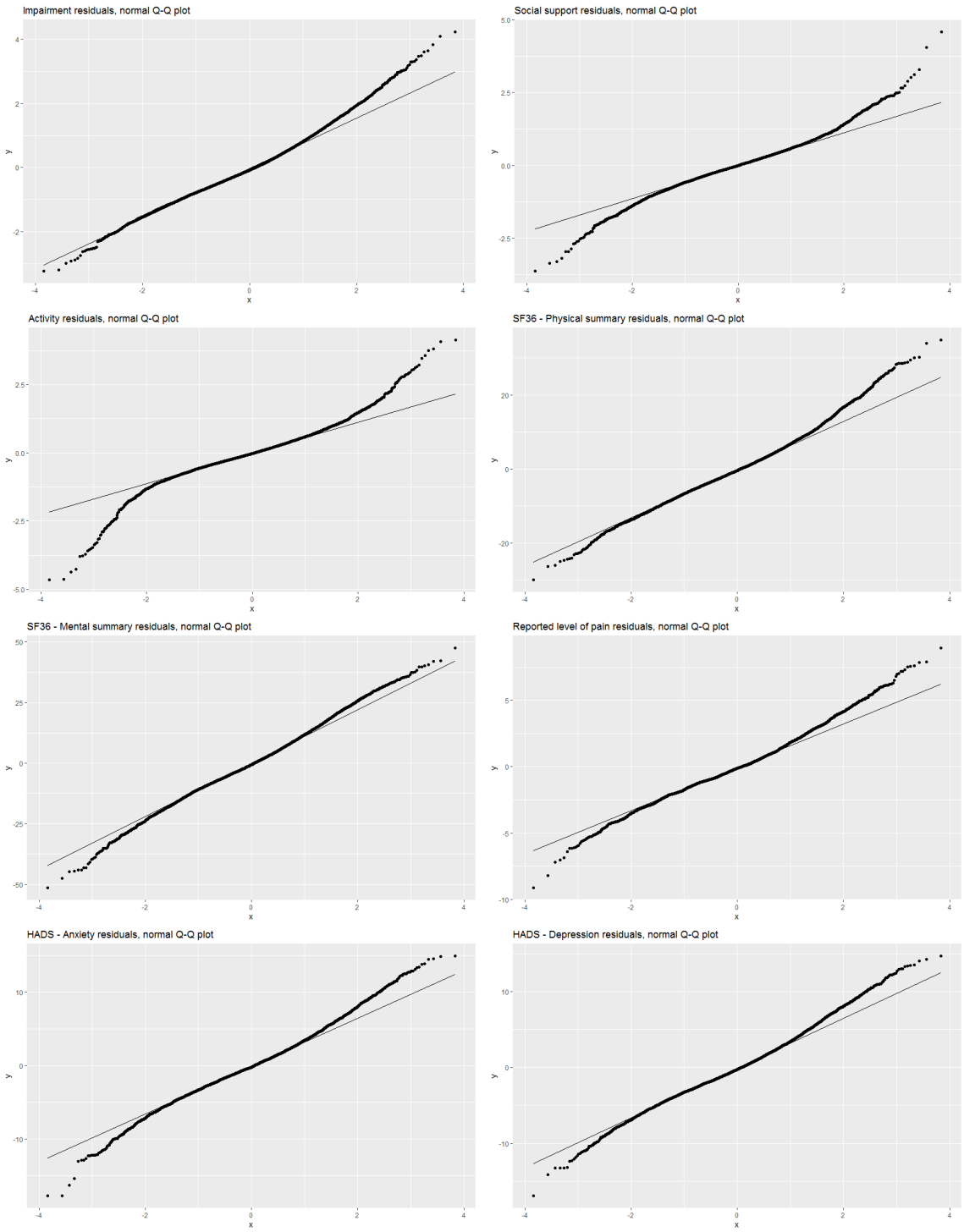


Figure C.2: Marginal normal Q-Q plots for residuals of modelled responses

Bibliography

[Ang+10] Dennis C. Ang, Matthew J. Bair, Teresa M. Damush, Jingwei Wu, Wanzhu Tu, and Kurt Kroenke. "Predictors of Pain Outcomes in Patients with Chronic Musculoskeletal Pain Co-morbid with Depression: Results from a Randomized Controlled Trial". In: *Pain Medicine* 11.4 (Apr. 2010), pp. 482–491. DOI: 10.1111/j.1526-4637.2009.00759.x.

[BJT04] Joan E. Broderick, Doerte U Junghaenel, and Dennis C Turk. "Stability of patient adaptation classifications on the multidimensional pain inventory". In: *Pain* 109.1 (May 2004), pp. 94–102. DOI: 10.1016/j.pain.2004.01.014.

[Bre+06] Harald Breivik, Beverly Collett, Vittorio Ventafridda, Rob Cohen, and Derek Gallacher. "Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment". In: *European Journal of Pain* 10.4 (May 2006), pp. 287–287. DOI: 10.1016/j.ejpain.2005.06.009.

[BT92] George E.P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, Inc., Apr. 1992. DOI: 10.1002/9781118033197.

[Dah+18] James Dahlhamer, Jacqueline Lucas, Carla Zelaya, Richard Nahin, Sean Mackey, Lynn DeBar, Robert Kerns, Michael Von Korff, Linda Porter, and Charles Helmick. "Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults — United States, 2016". In: *MMWR. Morbidity and Mortality Weekly Report* 67.36 (Sept. 2018), pp. 1001–1006. DOI: 10.15585/mmwr.mm6736a2.

[DB18] Christopher D. Desjardins and Okan Bulut. *Handbook of Educational Measurement and Psychometrics Using R*. Chapman and Hall/CRC, Sept. 2018. DOI: 10.1201/b20498.

[Dwo+05] Robert H. Dworkin, Dennis C. Turk, John T. Farrar, Jennifer A. Haythornthwaite, Mark P. Jensen, Nathaniel P. Katz, Robert D. Kerns, Gerold Stucki, Robert R. Allen, Nicholas Bellamy, Daniel B. Carr, Julie Chandler, Penney Cowan, Raymond Dionne, Bradley S. Galer, Sharon Hertz, Alejandro R. Jadad, Lynn D. Kramer, Donald C. Manning, Susan Martin, Cynthia G. McCormick, Michael P. McDermott, Patrick McGrath, Steve Quessy, Bob A. Rappaport, Wendye Robbins, James P. Robinson, Margaret Rothman, Mike A. Royal, Lee Simon, Joseph W. Stauffer, Wendy Stein, Jane Tollett, Joachim Wernicke, and James Witter. "Core outcome measures for chronic pain clinical trials: IMMPACT recommendations". In: *Pain* 113.1 (Jan. 2005), pp. 9–19. DOI: 10.1016/j.pain.2004.09.012.

[Eua+22] Pramote Euasobhon, Raviwon Atisook, Kulsiri Bumrungchatudom, Nantthasorn Zinboonyahgoon, Nattha Saisavoey, and Mark P. Jensen. "Reliability and responsivity of pain intensity scales in individuals with chronic pain". In: *Pain* Publish Ahead of Print (May 2022). DOI: 10.1097/j.pain.0000000000002692.

[Far+01] John T. Farrar, James P. Young, Linda LaMoreaux, John L. Werth, and Michael R. Poole. "Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale". In: *Pain* 94.2 (Nov. 2001), pp. 149–158. DOI: 10.1016/s0304-3959(01)00349-9.

[Fin11] Perry G. Fine. "Long-Term Consequences of Chronic Pain: Mounting Evidence for Pain as a Neurological Disease and Parallels with Other Chronic Disease States". In: *Pain Medicine* 12.7 (July 2011), pp. 996–1004. DOI: 10.1111/j.1526-4637.2011.01187.x.

[Gat+07] Robert J. Gatchel, Yuan Bo Peng, Madelon L. Peters, Perry N. Fuchs, and Dennis C. Turk. "The biopsychosocial approach to chronic pain: Scientific advances and future directions." In: *Psychological Bulletin* 133.4 (2007), pp. 581–624. DOI: 10.1037/0033-2909.133.4.581.

[Gel13] Andrew Gelman. *Bayesian Data Analysis Third Edition 3rd Edition Chapman HallCRC Texts in Statistical Science*. CRC Press, 2013. ISBN: 9781439840955.

[Geo21] Geoffrey Thompson. *MixMatrix: Classification with Matrix Variate Normal and t Distributions*. R package version 0.2.6. 2021. URL: https://CRAN.R-project.org/package=MixMatrix.

[Ger+16] Björn Gerdle, Peter Molander, Gunilla Stenberg, Britt-Marie Stålnacke, and Paul Enthoven. "Weak outcome predictors of multimodal rehabilitation at one-year follow-up in patients with chronic pain—a practice based evidence study from two SQRP centres". In: *BMC Musculoskeletal Disorders* 17.1 (Nov. 2016). DOI: 10.1186/s12891-016-1346-7.

[GH09] Karla Gough and Peter Hudson. "Psychometric Properties of the Hospital Anxiety and Depression Scale in Family Caregivers of Palliative Care Patients". In: *Journal of Pain and Symptom Management* 37.5 (May 2009), pp. 797–806. DOI: 10.1016/j.jpainsymman.2008.04.012.

[Her+03] M.J. Herrero, J. Blanch, J.M. Peri, J. De Pablo, L. Pintor, and A. Bulbena. "A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population". In: *General Hospital Psychiatry* 25.4 (July 2003), pp. 277–283. DOI: 10.1016/s0163-8343(03)00043-4.

[Jol02] I.T. Joliffe. *Principal Component Analysis*. Springer-Verlag, 2002. DOI: 10.1007/b98835.

[KTR85] Robert D. Kerns, Dennis C. Turk, and Thomas E. Rudy. "The West Haven-Yale Multidimensional Pain Inventory (WHYMPI)". In: *Pain* 23.4 (Dec. 1985), pp. 345–356. DOI: 10.1016/0304-3959(85)90004-1.

[LoM+19] Riccardo LoMartire, Björn Olov Äng, Björn Gerdle, and Linda Vixner. "Psychometric properties of Short Form-36 Health Survey, EuroQol 5-dimensions, and Hospital Anxiety and Depression Scale in patients with chronic pain". In: *Pain* 161.1 (Sept. 2019), pp. 83–95. DOI: 10.1097/j.pain.0000000000001700.

[Lor99] Urbano Lorenzo-Seva. "Promin: A Method for Oblique Factor Rotation". In: *Multivariate Behavioral Research* 34.3 (July 1999), pp. 347–365. DOI: 10.1207/s15327906mbr3403_3.

[Mak+19] Dominique Makowski, Mattan S. Ben-Shachar, S. H. Annabel Chen, and Daniel Lüdecke. "Indices of Effect Existence and Significance in the Bayesian Framework". In: *Frontiers in Psychology* 10 (Dec. 2019). DOI: 10.3389/fpsyg.2019.02767.

[MBL19] Dominique Makowski, Mattan S. Ben-Shachar, and Daniel Lüdecke. "bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework." In: *Journal of Open Source Software* 4.40 (2019), p. 1541. DOI: 10.21105/joss.01541. URL: https://joss.theoj.org/papers/10.21105/joss.01541.

[MN11] Jeffrey M McKillop and Warren R Nielson. "Improving the Usefulness of the Multidimensional Pain Inventory". In: *Pain Research and Management* 16.4 (2011), pp. 239–244. DOI: 10.1155/2011/873424.

[Mur07] Kevin P. Murphy. *Conjugate Bayesian analysis of the Gaussian distribution*. Tech. rep. University of British Columbia, 2007.

[Nic+19] Michael Nicholas, Johan W.S. Vlaeyen, Winfried Rief, Antonia Barke, Qasim Aziz, Rafael Benoliel, Milton Cohen, Stefan Evers, Maria Adele Giamberardino, Andreas Goebel, Beatrice Korwisi, Serge Perrot, Peter Svensson, Shuu-Jiun Wang, and Rolf-Detlef Treede and. "The IASP classification of chronic pain for ICD-11: chronic primary pain". In: *Pain* 160.1 (Jan. 2019), pp. 28–37. DOI: 10.1097/j.pain.0000000000001390.

[NNS14] V. Nyberg, M. Novo, and B. Sjölund. "Changes in Multidimensional Pain Inventory profile after a pain rehabilitation programme indicate the risk of receiving sick leave benefits one year later". In: *Journal of Rehabilitation Medicine* 46.10 (2014), pp. 1006–1013. DOI: 10.2340/16501977-1872.

[NRS13] NRS. "Nationella registret över Smärtrehabilitering: Årsrapport 2012 del 1". In: (2013).

[NRS22] NRS. "Nationella registret över Smärtrehabilitering: Årsrapport 2021 del 1". In: (2022).

[Raj+20] Srinivasa N. Raja, Daniel B. Carr, Milton Cohen, Nanna B. Finnerup, Herta Flor, Stephen Gibson, Francis J. Keefe, Jeffrey S. Mogil, Matthias Ringkamp, Kathleen A. Sluka, Xue-Jun Song, Bonnie Stevens, Mark D. Sullivan, Perri R. Tutelman, Takahiro Ushida, and Kyle Vader. "The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises". In: *Pain* 161.9 (Sept. 2020), pp. 1976–1982. DOI: 10.1097/j.pain.0000000000001939.

[RC01] Rosalind Rabin and Frank de Charro. "EQ-SD: a measure of health status from the EuroQol Group". In: *Annals of Medicine* 33.5 (Jan. 2001), pp. 337–343. DOI: 10.3109/07853890109002087.

[Roo+13] Aleid de Rooij, Marike van der Leeden, Leo D Roorda, Martijn PM Steultjens, and Joost Dekker. "Predictors of outcome of multidisciplinary treatment in chronic widespread pain: an observational study". In: *BMC Musculoskeletal Disorders* 14.1 (Apr. 2013). DOI: 10.1186/1471-2474-14-133.

[SBU21] SBU. *Multimodala och interdisciplinära behandlingar vid långvarig smärta: en systematisk översikt och utvärdering av effekter på hälsa och hälsoekonomiska aspekter*. SBU Utvärderar 341. Available from: https://www.sbu.se/341. Statens beredning för medicinsk och social utvärdering (SBU): Stockholm, 2021.

[Sca+08] L. Scascighini, V. Toma, S. Dober-Spielmann, and H. Sprott. "Multidisciplinary treatment for chronic pain: a systematic review of interventions and outcomes". In: *Rheumatology* 47.5 (Jan. 2008), pp. 670–678. DOI: 10.1093/rheumatology/ken021.

[Spi+18] L. Spinord, A. Kassberg, G. Stenberg, R. Lundqvist, and B. Stålnacke. "Comparison of two multimodal pain rehabilitation programmes, in relation to sex and age". In: *Journal of Rehabilitation Medicine* 50.7 (2018), pp. 619–628. DOI: 10.2340/16501977-2352.

[Tho02] Henry C. Thode. *Testing For Normality*. CRC Press, Jan. 2002. DOI: 10.1201/9780203910894.

[War94] John E. Ware. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Health Assessment Lab, New England Medical Center, 1994.

[WHO21] World Health Organization (WHO). *International Classification of Diseases, Eleventh Revision (ICD-11)*. 2019/2021. URL: https://icd.who.int/browse11.

[Wik22] Wikipedia. *Bayesian multivariate linear regression — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Bayesian%20multivariate%20linear%20regression&oldid=1089211109. [Online; accessed 29-August-2022]. 2022.

[WS92] J. E. Ware and C. D. Sherbourne. "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection." In: *Medical care* 30 (6 June 1992), pp. 473–483. ISSN: 0025-7079. ppublish.

[Wu+20] Aimin Wu, Lyn March, Xuanqi Zheng, Jinfeng Huang, Xiangyang Wang, Jie Zhao, Fiona M. Blyth, Emma Smith, Rachelle Buchbinder, and Damian Hoy. "Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017". In: *Annals of Translational Medicine* 8.6 (Mar. 2020), pp. 299–299. DOI: 10.21037/atm.2020.02.175.

[ZS83] A. S. Zigmond and R. P. Snaith. "The Hospital Anxiety and Depression Scale". In: *Acta Psychiatrica Scandinavica* 67.6 (June 1983), pp. 361–370. DOI: 10.1111/j.1600-0447.1983.tb09716.x.