Dynamic Bayesian networks for clinical decision support on ankylosing spondylitis patients under biological therapies

Ana Rita Lopes Borges Martins
ana.rita.lopes.borges.martins@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

July 2021

Abstract

Rheumatic and musculoskeletal diseases (RMDs) are among the most common chronic diseases worldwide, having an enormous burden on individuals. Ankylosing spondylitis (AS) is a type of inflammatory arthritis that affects primarily the spine and the sacroiliac joints, with patients experiencing not only physical limitations, but also reduced quality of life.

Currently, there is no cure for the disease, but there are treatments that aim at slowing its progression and relieving its symptoms, appearing as promising recently developed biological therapies, as tumor necrosis factor (TNF) antagonists. Nevertheless, not all patients respond to these therapies and some of them even experience severe adverse events. Furthermore, TNF blocking therapy is costly, being essential to identify which patients are likely to benefit from these agents and which are not.

The goal of this work is to find clinical predictors of therapy response and to distinguish between three different anti-TNF agents, adalimumab, etanercept and infliximab, in order to evaluate which therapy best suits each patient. This was done using dynamic Bayesian network (DBN) models built from data from the Rheumatic Diseases Portuguese Register.

This study allowed to corroborate existing literature findings, as male patients responding better to all therapies studied, and also to produce new insights. Firstly, patients with greater BMI seem to response worse when treated with adalimumab. Secondly, HLA-B27 negative patients seem to respond worse to infliximab. Finally, for older patients or patients with higher disease duration, infliximab seems to be a better therapeutic option.

Keywords: Ankylosing spondylitis, dynamic Bayesian networks, anti-TNF-alpha therapies, therapy outcome prediction

1. Introduction

Rheumatic and musculoskeletal diseases (RMDs) are among the most common chronic diseases worldwide, comprising more than 200 different diseases affecting mainly the joints, tendons, ligaments, bones and muscles. Among these diseases there is ankylosing spondylitis (AS), a type of inflammatory arthritis that affects primarily the spine and the sacroiliac joints [2, 22]. AS patients experience not only physical limitations, which can lead to implications for employment and in extreme cases even result in inability of working, but also to a reduced quality of life, possibly experiencing anxiety and depression. It is a complex and unpredictable disease, being its pathogenesis poorly understood [13], nevertheless, due to its serious consequences, there is an urgent need of better understanding the disease and improving the knowledge of the causes and mechanisms of AS, to improve early diagnosis and prevention and develop innovative therapies [9]. Currently, there is no cure for the disease, but there are treatments that aim at slowing its progression and relieving its symptoms [21]. Recently developed biological therapies, as tumor necrosis factor (TNF) antagonists (anti-TNF), appear as promising treatments, however, not all patients respond to these therapies and some of them even experience severe adverse events. Furthermore, TNF blocking therapy is costly, being associated with a great economic burden. Therefore, to minimise risks and costs associated with these therapies, it is essential to identify which patients are likely to benefit from anti-TNF agents and which are not. Additionally, there is no specific drug selection criteria regards deciding which biologic should be given to each patient, being this choice in some cases based on logistic issues, rather than on scientific evidence. This leads to patients switching between different

therapies until finding one to which they respond positively, nevertheless, this lack of accuracy in the therapeutic choice should not be underestimated. This way, being able to predict in advance the patient's response for each treatment, according to the patient's characteristics, would be of great benefit.

The main goal of this work is to identify predictors of response to the different biological therapies and to be able to produce insights about which treatment suits each patient best, in the context of a personalized medicine approach, investigating differences between the existing anti-TNF therapies. This is proposed to be achieved by developing Bayesian network (BN) models, more concretely dynamic Bayesian networks (DBNs), using data of AS patients on biological therapies taken from the Rheumatic Diseases Portuguese Register (RNDR), the Reuma.pt database.

This document is organized as follows. First, in section 2 a theoretical background on both ankylosing spondylitis and Bayesian networks, respectively, is presented. Next, in section 3 the object of study of this work is described and the methodology followed is presented, from data acquisition and processing, to the implementation of the DBN models. In section 4 the obtained results are presented and discussed. Finally, in section 5 the conclusions reached with this work are summarized and suggestions for future work are presented.

2. Background

2.1. Ankylosing Spondylitis

Ankylosing Spondylitis (AS) is a type of inflammatory arthritis that affects primarily the spine and the sacroiliac joints, causing inflammation and leading to chronic pain and discomfort [2, 22]. This disease can be associated with new bone formation in the spine and fusion of the vertebrae, causing stiffness and pain and leading to possible loss of physical function and spinal mobility. Nevertheless, AS doesn't affect only the spine and peripheral arthritis and extra-articular features can also be present [18]. AS starts at a relatively young age, usually in the second or third decade of life, with men being two to three times more affected than women. Also, the pattern of the disease is different between the two, with women having milder disease and more extra spinal involvement [13]. The worldwide prevalence of AS is believed to be between 0,1% and 1,4% [8], while in Europe is around 0,25% [23].

The etiology and pathogenesis of AS are still poorly understood, but it is believed that it develops through complex interactions between genetic and environmental factors [13, 27]. Nevertheless, AS is strongly associated with the human leukocyte antigen B27 (HLA-B27) [14], a protein located on the surface of white blood cells found in 90-95% of AS patients. The tumor necrosis factor alpha (TNF-alpha) is a proinflammatory cytokine also believed to be implicated in the pathogenesis of AS. Tumor necrosis factor antagonists are one type of the recently developed biological therapies.

Clinical monitoring of the disease is of great importance, not only to assess disease status but also to better understand patients' response to treatment and guide therapeutic decisions, and it became even more relevant after the emergence of anti-TNF therapies [28, 29]. Therefore, over time, a lot of clinical measures have been developed and used in clinical practice, being the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), the Bath Ankylosing Spondylitis Functional Index (BASFI) and the Ankylosing Spondylitis Disease Activity Score (ASDAS) among the most commonly used. BASDAI is a score used to assess patient-reported disease activity, consisting of 6 questions related to fatigue, spinal pain, peripheral arthritis, enthesitis and duration and severity of morning stiffness [20]. BASFI is a score used to assess the degree of functional limitation of patients, based on 10 questions about daily activities, like dressing, bending and standing, and ability to cope with everyday life. ASDAS is a score developed to also assess disease activity, being the first to use both self-reported items and objective measures [28]. This score includes patient-reported assessments of back pain, duration of morning stiffness, global assessment of disease activity and peripheral joint pain/swelling, and an acute phase-reactant. Acute phase reactants are inflammation markers that exhibit significant changes in serum concentration during inflammation [3], and the ones used in ASDAS are either the erythrocyte sedimentation rate (ESR), in mm/h, or the C-reactive protein (CRP), in mg/L, depending on the ASDAS score's version used, being ASDAS-CRP the preferred one.

Identified predictors of response to biological therapy are younger age [6, 1, 12], male gender [1, 12], higher ASDAS [1] and BASDAI [6, 12] scores at baseline, lower BASFI at baseline [6, 12], raised levels of ESR [6, 15] and CRP [6, 15, 12] at baseline, presence of peripheral arthritis [1], smaller disease duration [6] and higher patient's global assessment of disease activity [1]. Furthermore, HLA-B27-positive patients have been reported to respond better to TNF antagonists [25, 12]. On the other hand, female gender [1, 16], absence of peripheral arthritis [1], higher BMI [17, 11], higher baseline BASFI score [15] and lower baseline levels of ESR and CRP [1] have been identified as predictors of discontinuation of TNF-alpha blocking therapy.

2.2. Bayesian Networks

A Bayesian Network (BN) is a graphical model for representing relationships among random variables,

using conditional probabilities. A BN $\mathbf{B}$ is defined as a triple $\mathbf{B} = (\mathbf{X}, G, \Theta)$ where:

- $\mathbf{X} = \{X_1, ..., X_n\}$ is a n-dimensional finite random vector, where each random variable X_i, with $i \in \{1, ..., n\}$, takes values in the set $\{x_{\mathrm{i1}}, ..., x_{\mathrm{ir_i}}\}$, being x_{ik} the k-th value of X_i;
- $G = (N, E)$ is a directed acyclic graph (DAG), with $\mathbf{N}$ being the set of nodes/vertices in $\mathbf{X}$, representing the random variables and $\mathbf{E}$ is the set of directed edges that represents the dependencies between the variables. If there is an edge from node i to node j, i is said to be a parent of j and j is said to be a child of i;
- $\Theta = \{\theta_{\mathrm{ijk}}\}$ is a set of parameters specifying the conditional probability distribution of each variable according to the structure of the graph G, being

$$\theta_{\mathrm{ijk}} = P_B(X_i = x_{\mathrm{ik}} | \mathbf{pa}(X_i) = w_{\mathrm{ij}}), \quad (1)$$

  where $\mathbf{pa}(X_i)$ denotes the set of parents of X_i in G, and w_{ij} is the j-th configuration of the set of parents $\mathbf{pa}(X_i)$ [7, 5].

A BN defines a unique joint probability distribution given by the product, over all the nodes of the graph, of a conditional distribution for each node conditioned on its parents. Therefore, for a BN with n nodes, the joint distribution is given by [17]:

$$P_B(X_1, ..., X_n) = \prod_{i=1}^{n} P_B(X_i | \mathbf{pa}(X_i)), \quad (2)$$

meaning that given its parents, each node is conditionally independent of all other non descendant nodes [5, 19]. Dynamic Bayesian networks (DBNs) are an extension of Bayesian networks to represent the evolution in time of a system. In fact, DBNs consist of a series of time slices that represent the state of all variables at a certain time t, being allowed not only intra-slice connections, that are the connections within the time slices, but also inter-slice connections, that are connections between variables from different slices. These inter-slice connections follow the direction of time, meaning that a variable cannot have descendants in previous time slices [26, 24].

The problem of learning a Bayesian Network given a dataset of observations of a set of variables, consists in finding the network that is most probable to have generated the data, this is, learning its structure and parameters, i.e., the DAG and the corresponding conditional probability tables [10].

3. Methods

3.1. Reuma.pt database

The object of study of this dissertation is the Rheumatic Diseases Portuguese Register (RNDR), i.e., the Reuma.pt database, more precisely the data from patients with AS who undergo or have undergone biologic therapy, retrieved on July 22$^{\mathrm{nd}}$ 2019. This online platform was developed by the Portuguese Society of Rheumatology (SPR) and it became active in June 2008 due to a need of improving monitoring in patients with rheumatic diseases. Reuma.pt allows better data production and evaluation, leading to a better insight about these diseases and therefore allowing the development of better therapies. The database contains several information for each patient, including identification data, demographic data, work status, life style habits, anthropomorphic data, comorbidities, previous medical history, past and current therapies, disease activity and functional assessment scores, laboratory measurements, and others. Among the data, both static and dynamic variables can be found: static variables are either time-independent covariates, like gender, or data gathered only at the patient's first visit, like smoking status; dynamic variables are measurements that are not constant over the whole study, collected over subsequent patient's visits [4]. For the present study, there were 3 relevant datasets: one with general information on the patients, a second one with patients' past and current therapies and a third one with the patients' registered appointments, and corresponding measurements.

3.2. Pre-processing

The goal of this work is to create DBNs from the data, in order to better understand therapy response in ankylosing spondylitis patients. Before moving forward, it should be noted that only the first biological agent received by each patient was considered in this study, so, more precisely, this study evaluates the patient's response for the first biological therapy received, which is known to be different than the response from the subsequent biologics. For creating the nerworks, it was used the `bnstruct` R package, which through state-of-the art algorithms learns the network that may have generated a particular set of data, even in the presence of missing values, which is the case of the present data and a common situation in the clinical context.

Before inserting data into the models, several processing steps were performed. First, to identify inconsistencies in the data, including either incorrect, incomplete or incoherent information, a detailed analysis of the dataset was performed. Several issues were spotted, and corrected when possible, with the help of medical professionals. Next, it was necessary to select the variables that were going to be used in the models. This process was done based on expert knowledge, after discussion with medical professionals who pointed out relevant variables, and also based on the amount of missing

data for each variable. The selected variables are presented in Tables 1 and 2, respectively static, dynamic, along with the corresponding discretization. In turn, the outcomes of interest are presented in Table 3, being the goal of this work to understand how these variables are impacted by static and dynamic ones. Discretization was either made using reference levels based on clinical knowledge or using quantiles, when reference ranges were not available or did not suit the data well.

Table 1: Static variables used in the study, together with the corresponding description and discretization performed.

Variable	Description	Discretization (label: corresponding elements)
Static variables		
Gender	Self-explanatory	0: Male ; 1: Female
Age at biologic onset	Self-explanatory	0:]0,36] ; 1:]36,47] ; 2:]47,+∞[
BMI	Body Mass Index at onset (kg/m2)	0:]0;25[; 1: [25,30[; 2: [30,+∞[
HLA-B27	Whether a patient carries, or not, the HLA-B27 gene	0: Negative ; 1: Positive
Education level	Self-explanatory	0: No education or elementary school; 1: Middle school ; 2: High school ; 3: Higher education
Disease duration	Time in years since the beggining of disease	0: [0,7] ; 1:]7,16] ; 2:]16,+∞[

Table 2: Dynamic variables used in the study, together with the corresponding description and discretization performed.

Variable	Description	Discretization (label: corresponding elements)
Dynamic variables		
DMARD	Whether, or not, a patient is taking a DMARD at the time of the appointment	0: No ; 1: Yes
Corticoid	Whether, or not, a patient is taking a corticoid at the time of the appointment	0: No ; 1: Yes
VAS patient	Patient's global self evaluation (on a scale from 0 to 100)	0: [0,2[; 1: [2,5[; 2: [5,10]
VAS doctor	Doctor's evaluation of the patient's global state (on a scale from 0 to 100)	0: [0,1[; 1: [1,3[; 2: [3,10]
BASDAI Q1	Patient's answer to the 1st BASDAI question (on a scale from 1 to 10)	0: [0,4[; 1: [4,7[; 2: [7,10]
BASDAI Q2	Patient's answer to the 2nd BASDAI question (on a scale from 1 to 10)	0: [0,4[; 1: [4,7[; 2: [7,10]
BASDAI Q3	Patient's answer to the 3rd BASDAI question (on a scale from 1 to 10)	0: [0,4[; 1: [4,7[; 2: [7,10]
BASDAI Q4	Patient's answer to the 4th BASDAI question (on a scale from 1 to 10)	0: [0,4[; 1: [4,7[; 2: [7,10]
BASDAI Q5 & Q6	Average of patient's answers to the 5th and 6th BASDAI questions (on a scale from 1 to 10)	0: [0,4[; 1: [4,7[; 2: [7,10]
BASFI	Patient's BASFI index (average of patient's answers to the 10 questions composing it, on a scale from 1 to 10)	0: [0,2] ; 1:]2,5] ; 2:]5,10]
CRP	C-Reactive Protein's level (mg/L)	0: [0,1] ; 1:]1,3] ; 2:]3,8.5] ; 3:]8.5,+∞[

Next, to input data into the models, the different observations, i.e., the dynamic variables acquired in the different appointments, had to be equally spaced, which did not happen in the original appointments' data. Therefore, in order to have records at the same time points for all patients, it was necessary to choose a time interval between them. It was opted to consider two different intervals, 3 and 6 months, being an algorithm implemented in Python for the purpose of appointments' selection.

Table 3: Outcome variables used in the study, together with the corresponding description and discretization performed.

Variable	Description	Discretization (label: corresponding elements)
Outcome variables		
ASDAS 12M	Composite score to assess disease activity in AS, evaluated at month 12	0: [0,1.3[; 1: [1.3,2.1[; 2: [2.1,3.5] ; 3:]3.5,+∞[
ASDAS improvement 12M	Difference between ASDAS after 12 months of treatment with the biologic and at the begging of the treatment	0:]-∞, 1.1[; 1 : [1.1, 2[; 2 : [2, +∞[
Inefficiency	Whether a patient stopped, or not, the therapy due to inefficiency	0: No ; 1: Yes

In order to have more confidence on the results, different datasets representing the same data were created, with the aim of building several models and being able to compare them. To this end, several parameters were varied, as can be seen in the simplified schematic representation in Figure 1.

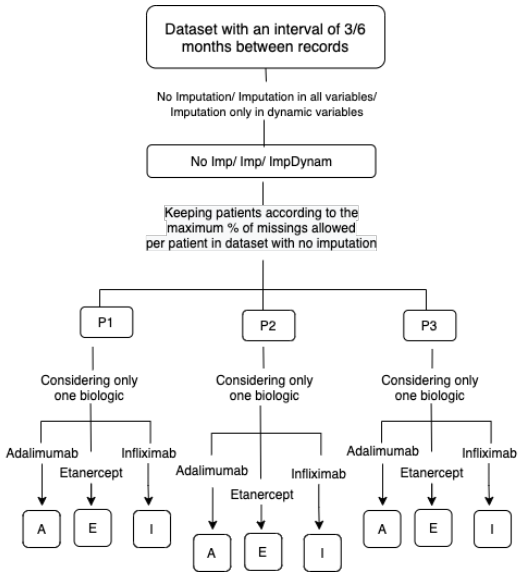

Figure 1: Simplified schematic representation of datasets' creation process.

First, as already mentioned, two different datasets were created by selecting appointments based on different intervals between records, 3 and 6 months, being considered the *initial* datasets. In this step, it was noticed that some patients did not have any records of dynamic variables while others

presented a considerable amount of missing values for these variables, and so, to set an exclusion criteria, patients with more than 70% of missing dynamic data were removed from the database. The next element to be varied was imputation. From each initial dataset, the one without imputation, two more datasets were created, one where imputation was performed in all variables and another one where imputation was performed only in dynamic variables. Imputation was taken into account to deal with the missing values in the data, since training a model with a dataset that has a considerable amount of data missing can significantly impact the quality of the model and, subsequently, the conclusions that can be drawn. Mainly, the techniques used for this task were linear interpolation for dynamic variables and the k-nearest neighbor for static ones. Next, for each dataset, the maximum percentage of missing values allowed per patient was varied, with three different values being considered, being created three datasets from each previously existing one. Finally, from these datasets, which contained patients taking all the different biological therapies, only the 3 most common biologics were considered, adalimumab, etanercept and infliximab. In total, 54 datasets were used in the models.

3.3. Structure and parameter learning

As already mentioned, it was opted to use `bnstruct` R package to learn the networks. This package allows to define constrains to the network structure, which work as a way of inputing domain knowledge by defining layers for the variables, where a variable can only have parents in lower-numbered layers, in order to avoid relations with no clinical or biological sense. This way, variables were divided into 4 general layers, as presented in Table 4, separating static variables from dynamic and outcome ones. Moreover, inside each temporal layer, variables were also separated in 4 layers, according to layering scheme presented in Table 5.

Table 4: General layering of variables in the DBN.

Layer	Variables
1	Gender, HLA-B27, BMI, Education level, Age at biologic onset, Disease duration
2	Variables at time t-1
3	Variables at time t
4	ASDAS improvement at 12 months, ASDAS at 12 months, Inefficiency

Besides defining the layers, it is also possible to specify the influence of upper lower-numbered layers in lower high-numbered layers. In the case of the present work, the possible influences between layers were specified as follows. First, static variables were all placed in the same layer, the upper one and were allowed to influence variables in all other layers, but not each other. Next, variables at each time point (t) were allowed to influence: (1) each other, based on a layering scheme presented in Table 5, (2) variables at the following time point (t+1) and (3) the outcome variables. In turn, variables at the last time point were allowed to influence each other and the outcome variables. Finally, outcome variables are in the lowest layer and edges between them are forbidden.

Table 5: Layering of variables inside each temporal layer.

Layer	Variables
1	DMARD, Corticoid
2	CRP, BASDAI questions, BASFI
3	VAS patient
4	VAS doctor

The DBN structure learning was performed using the Maximum Minimum Hill-Climbing (MMHC) algorithm, an heuristic algorithm that combines the MMPC and HC algorithms. First, the Maximum Minimum Parents Children (MMPC) algorithm infers the skeleton of the network, with non-directed edges, and then Hill Climbing algorithm (HC) returns the edges' orientation by starting the search in the possibilities' space from the configuration provided by MMPC. The use of the MMHC algorithms required the choice of a scoring function to evaluate how good the structure fits the data, in order to find the optimal network. Three scoring functions are provided by `bnstruct`, the Bayesian-Dirichlet equivalent uniform (BDeu), the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). It was chosen to create networks using all three available scoring functions, to then perform a comparison of the different structures obtained, which represent the same underlying data. Finally, to conclude network learning, the set of parameters of the conditional probability distribution at each node was computed by a Maximum-a-Posteriori (MAP) estimation.

4. Results

4.1. DBN models

The structure of the network is represented as an adjacency matrix. An adjacency matrix is a n×n matrix, with n being the number of nodes, where an entry (i, j) takes the value of 1 if there is an edge from node i to node j, and 0 otherwise. As mentioned before, by varying several parameters, different models were built to represent the same data. This way, by comparing the different networks obtained, it was possible to evaluate the confidence on the edges found, with greater confidence being

given to edges that appeared more frequently in the models. To provide an easier and more intuitive summary of data, different adjacency matrices representing the same underlying data were summed, condensing different networks into a single graphical representation. The resultant matrix was then used to create a heatmap, a two-dimensional data visualization technique where the magnitude of a phenomenon, in this case the appearance of a certain edge in the networks, is represented through color, using the `gplots` R package. By varying the interval between appointments, if the dataset was imputated or not, the maximum percentage of missings allowed and the biologic, 18 different heatmaps were created, with 9 different networks each. Indeed, for each heatmap, 3 datasets with different maximum percentages of missings allowed per patient were used. Furthermore, each of the previous datasets resulted in 3 different networks, by varying the scoring function used, AIC, BIC or BDeu. To ease the analysis, each of the heatmaps was named according to its characteristics: Biologic+Interval Between Records+Imputation strategy used.

All correlations identified in the networks with each outcome variable, for each heatmap, together with the number of times that the respective association appeared in the networks are presented for adalimumab, etanercept and infliximab in Tables 6, 7 and 8, respectively.

Table 6: Summary of variables correlating with outcome variables for adalimumab.

	Adalimumab		
	Inefficiency	**ASDAS at 12 months**	**ASDAS change at 12 months**
3MNoImp	VAS doctor 9m [6] BASFI 9m [5]	BMI [3] VAS doctor 0m [2] BASDAI q2 3m [2] BASFI 6m [2] CRP 6m [3] VAS doctor 12m [6]	BMI [9] VAS patient 0m [2] VAS doctor 0m [4] CRP 9m [1] BASDAI q3 9m [1] BASDAI q3 12m [1]
3MImp	VAS doctor 3m [3] VAS patient 9m [3] BASFI 12m [6]	VAS doctor 9m [2] VAS doctor 12m [7] BASDAI q2 12m [1] BASDAI q4 12m [5]	VAS patient 0m [7] CRP 0m [1] BASDAI q2 12m [3] BASDAI q3 12m [1] BASDAI q5&q6 12m [1]
3MImpDynam	VAS doctor 3m [3] VAS patient 9m [3] BASFI 12m [6]	VAS doctor 9m [2] VAS doctor 12m [7] BASDAI q2 12m [1] BASDAI q4 12m [5]	BMI [6] VAS patient 0m [2] BASDAI q2 12m [3] BASDAI q5&q6 12m [2]
6MNoImp	VAS doctor 0m [5] VAS doctor 6m [3] BASFI 6m [2]	BMI [9] VAS doctor 12m [9] BASFI 12m [1]	Education [3] VAS patient 0m [2] CRP 0m [6] BASDAI q5&q6 12m [1]
6MImp	BASFI 6m [3] VAS doctor 12m [3]	VAS doctor 12m [9] BASDAI q2 12m [3] BASDAI q4 12m [1] BASFI 12m [3]	VAS patient 0m [6] CRP 0m [6] BASDAI q2 12m [2] BASDAI q5&q6 12m [1]
6MImpDynam	BASFI 6m [3] VAS doctor 12m [3]	BMI [3] VAS doctor 12m [8] BASDAI q2 12m [2] BASDAI q4 12m [2] BASFI 12m [2]	BMI [3] VAS patient 0m [4] CRP 0m [5] BASDAI q5&q6 12m [1]

Starting with inefficiency, it is possible to note that only adalimumab's networks present significant correlations between inefficiency and other variables. Indeed, etanercept's networks do not show any variable correlating with this outcome, while networks for infliximab present a single correlation with BASDAI questions 5 and 6 after 12 months of therapy initiation. Among associations with inefficiency observed for adalimumab, a significant amount are with variables recorded at more advanced time points, which is expected since the greater the time the patient is subject to a therapy, the more clearly its impact on the patient can be seen. On the other hand, correlation between inefficiency and VAS doctor at 3 months was identified, where patients with worse doctor's evaluation after 3 months of therapy have greater probability of therapy inefficiency, compared to patients with lower VAS doctor at that time, leading to the conclusion that this variable can act as an early in-

Table 7: Summary of variables correlating with outcome variables for etanercept.

	Etanercept		
	Inefficiency	**ASDAS at 12 months**	**ASDAS change at 12 months**
3MNoImp	—	VAS doctor 3m [1] BASFI 6m [4] VAS doctor 9m [3] BASFI 9m [6] VAS doctor 12m [4]	VAS doctor 3m [6] BASDAI q4 6m [1] VAS doctor 12m [6]
3MImp	—	Gender [1] VAS doctor 6m [8] VAS patient 12m [2] VAS doctor 12m [1] BASDAI q3 12m [3]	CRP 0m [7] BASDAI q3 0m [1] VAS patient 12m [1] BASDAI q2 12m [1] BASDAI q4 12m [2]
3MImpDynam	—	Disease duration [1] VAS doctor 6m [8] VAS patient 12m [2] VAS doctor 12m [1] BASDAI q3 12m [3]	CRP 0m [7] BASDAI q3 0m [1] VAS patient 12m [1] BASDAI q2 12m [1] BASDAI q4 12m [2]
6MNoImp	—	BASFI 6m [4] VAS patient 12m [1] VAS doctor 12m [9] BASFI 12m [2]	CRP 0m [5] BASDAI q4 6m [5] VAS patient 12m [2]
6MImp	—	VAS patient 12m [1] VAS doctor 12m [9] BASDAI q2 12m [2] BASDAI q3 12m [1] BASFI 12m [2]	CRP 0m [7] VAS patient 12m [1] BASDAI q1 12m [1] BASDAI q2 12m [2] BASDAI q3 12m [1]
6MImpDynam	—	VAS patient 12m [1] VAS doctor 12m [9] BASDAI q2 12m [2] BASDAI q3 12m [1] BASFI 12m [2]	CRP 0m [7] VAS patient 12m [1] BASDAI q1 12m [1] BASDAI q2 12m [2] BASDAI q3 12m [1]

Table 8: Summary of variables correlating with outcome variables for infliximab.

	Infliximab		
	Inefficiency	**ASDAS at 12 months**	**ASDAS change at 12 months**
3MNoImp	—	BASDAI q5&q6 3m [3] VAS doctor 9m [9] VAS patient 12m [3]	VAS doctor 0m [9] CRP 3m [1] BASDAI q1 9m [2] BASFI 9m [1] VAS patient 12m [2]
3MImp	BASDAI q5&q6 12m [1]	VAS doctor 12m [9] BASDAI q2 12m [4] BASDAI q4 12m [1]	CRP 0m [1] BASFI 0m [1] VAS doctor 3m [3] VAS patient 12m [6] BASFI 12m [1]
3MImpDynam	BASDAI q5&q6 12m [1]	Education [1] VAS doctor 12m [9] BASDAI q2 12m [4] BASDAI q4 12m [1]	CRP 0m [1] BASFI 0m [1] VAS doctor 3m [3] VAS patient 12m [6] BASFI 12m [1]
6MNoImp	—	VAS doctor 12m [9] BASDAI q2 12m [2]	BASDAI q1 6m [8] BASDAI q5&q6 12m [1]
6MImp	—	VAS doctor 12m [9] BASDAI q2 12m [6]	BASDAI q5&q6 0m [1] VAS patient 6m [1] BASDAI q1 6m [2] BASDAI q3 12m [3] BASDAI q5&q6 12m [5]
6MImpDynam	—	HLA-B27 [3] BMI [2] VAS doctor 12m [9] BASDAI q2 12m [4]	BASDAI q5&q6 0m [1] VAS patient 6m [1] BASDAI q1 6m [2] BASDAI q3 12m [3] BASDAI q5&q6 12m [5]

dicator of therapy response and if after 3 months of treatment initiation the doctor's evaluation on the patient's condition is still high, there are more chances that the therapy will fail in the future. Finally, for VAS doctor at baseline, it seems that, for patients with higher VAS doctor at therapy beginning, there is a higher proportion of people in which therapy is inefficient, than in patients where VAS doctor was smaller, being hypothesized that having higher VAS doctor at baseline can in some situations lead to therapy inefficiency.

Regarding ASDAS at 12 months, it is possible to note that it correlates mainly with variables recorded 12 months after therapy initiation. Indeed, since ASDAS is a score that measures disease activity, it is expected that other dynamic variables reflecting the patient's condition evaluated at this time may evolve in the same direction, leading to the existence of correlation between them. Moreover, CRP and BASDAI questions 2,3 and 6 are part of the ASDAS calculation formula, being the correlation between them expected. Concerning static variables associated with ASDAS at 12 months, a strong correlation between this variable and BMI was found for biologic adalimumab. Indeed, according to the literature, patients with greater BMI respond worse to therapy, which was confirmed in the obtained networks by having smaller BMI more associated with smaller ASDAS at 12 months and greater BMI more associated with greater ASDAS at 12 months. Nevertheless, this correlation appeared mainly in patients being treated with adalimumab, which may indicate that the BMI impact is stronger for patients being treated with this therapy, compared to patients being treated with etanercept or infliximab. In literature, there was no evidence about this fact, however, when discussing this finding with medical professionals, they mentioned that although no studies have yet been done on this field, it was something that they had already noticed. The proposed motive for this was the fact that for both etanercept and infliximab the amount given to the patients is based on their body weight, which doesn't happen with adalimumab, where instead a fixed amount is given. Therefore, this may lead to obese patients being treated with adalimumab not receiving enough amount of medication needed to produce good therapy response, and therefore, this may lead to the conclusion that this therapy should be avoided for obese patients. On the other hand, for etanercept, two correlations with static variables were found, being those variables gender and disease duration. Nevertheless, each correlation was only found in a single heatmap, in a single network, indicating that this association may not be very strong. Finally, for infliximab, three static variables were found to correlate with ASDAS at 12 months, education, HLA-B27 and BMI. Regarding association with HLA-B27, it seems that HLA-B27 positive patients may respond better to therapy, however, this is not clear. For associations with education and BMI, no clear pattern was identified.

Finally, concerning ASDAS change at 12 months, several correlations were obtained between this outcome variable and other variables, for the different biologics. ASDAS improvement is obtained by subtracting the ASDAS value at 12 months to the ASDAS value at baseline, therefore depending on ASDAS in these two time points. When looking into the identified correlations, a considerable amount was indeed with variables at baseline and 12 months after therapy. It could be expected that the identified correlations were with variables at these two time points simultaneously, nevertheless it was not always the case, since several single correlations were found. Regarding variables at baseline, some that presented a greater amount of correlation were CRP, VAS patient and VAS doctor, where a greater value of these variables was associated with a greater improvement. These variables can be seen as reflecting the level of disease activity of the patient, like ASDAS does, and so these findings should be evaluated carefully, since patients with higher disease activity at baseline, and subsequently a greater ASDAS, have more room for improvement, i.e. for decreasing their ASDAS score. This hypothesis gains strength when noticing that these variables did not appear correlated with ASDAS at 12 months, possibly meaning that the association is more with ASDAS improvement than with the disease activity level itself. The same logic can be applied for dynamic variables after 12 months of therapy, where greater values of VAS doctor/ patient, which reflect a greater disease activity, are associated with smaller improvements, as expected, since the activity is still elevated after 12 months of therapy, while smaller values have a greater probability of being associated with greater improvements. Regarding correlation with static variables, which appear more interesting for the purpose of this dissertation, for adalimumab, a great correlation between ASDAS change and BMI was identified, with greater BMI being associated with a smaller ASDAS improvement. This way, the hypothesis that adalimumab might not be the best therapeutic option for obese patients is strengthened. Another correlation for adalimumab's patients also identified was with education, however, no defined influence pattern was found. On the other hand, for etanercept and infliximab, no correlation with static variables was found.

4.2. **Chi-square tests of independence**

To verify some previously obtained associations between variables, Chi-square (χ^2) tests of independence were performed. A Chi-square test is a statistical hypothesis test performed between a pair of categorical variables that determines whether the two variables are independent or related, by assuming that no relationship exists between the two categorical variables, i.e., that they are independent (null hypothesis). The results from this test are used to calculate a p-value, which allows to evaluate how likely the observed frequencies would be if the null hypothesis is true, by representing the probability of obtaining results 'as extreme' or 'more extreme' than the ones obtained.

Since the main goal of this study is to help predict therapy outcome, and the ideal is to obtain associations between outcomes and variables that can be measured before therapy initiation, these tests were performed only for static variables and for VAS patient, VAS doctor and CRP at baseline, which were the dynamic variables measured at the beginning of therapy that more commonly appeared correlated with outcome variables in the networks, and for which the identification of a pattern was possible. The tests were performed for all different biologic therapies and the obtained p-values are presented in Tables 9, 10 and 11, for adalimumab, etanercet and infliximab, respectively. A significance level of α=0.05 (5%) was chosen, meaning that for p-values smaller than this value the null hypothesis should be rejected, i.e., the variables are not independent of each other.

Table 9: P-values obtained from Chi-square tests for adalimumab.

	Adalimumab		
	Inefficiency	**ASDAS at 12 months**	**ASDAS change at 12 months**
Gender	0.104	0.000	0.002
HLA-B27	0.112	0.399	0.442
BMI	0.922	0.004	0.008
Age at biologic onset	0.686	0.003	0.102
Disease duration	0.821	0.001	0.855
Education	0.552	0.277	0.064
VAS patient at 0M	0.064	0.090	0.000
VAS doctor at 0M	0.093	0.531	0.060
CRP at 0M	0.805	0.825	0.000

Table 10: P-values obtained from Chi-square tests for etanercept.

	Etanercept		
	Inefficiency	**ASDAS at 12 months**	**ASDAS change at 12 months**
Gender	0.338	0.001	0.028
HLA-B27	0.315	0.471	0.403
BMI	0.126	0.809	0.748
Age at biologic onset	0.881	0.002	0.966
Disease duration	0.370	0.000	0.211
Education	0.661	0.180	0.085
VAS patient at 0M	0.140	0.002	0.001
VAS doctor at 0M	0.814	0.129	0.043
CRP at 0M	0.758	0.733	0.000

Table 11: P-values obtained from Chi-square tests for infliximab.

	Infliximab		
	Inefficiency	**ASDAS at 12 months**	**ASDAS change at 12 months**
Gender	0.741	0.033	0.042
HLA-B27	0.014	0.038	0.068
BMI	0.904	0.131	0.105
Age at biologic onset	0.231	0.522	0.118
Disease duration	0.859	0.185	0.794
Education	0.424	0.018	0.259
VAS patient at 0M	0.351	0.130	0.707
VAS doctor at 0M	0.213	0.570	0.601
CRP at 0M	0.860	0.769	0.002

Starting with inefficiency, looking into the obtained p-values, the null hypothesis can only be rejected for the pair inefficiency & HLA-B27, for infliximab's patients. For HLA-B27 negative patients, the number of patients that experienced therapy inefficiency was higher than expected if variables were independent, while the number of patients that didn't experience therapy inefficiency was smaller. On the other hand, for HLA-B27 positive patients, the contrary was identified, which seems to indicate that infliximab might not be the best therapeutic option for HLA-B27 negative patients.

On the other hand, for ASDAS and ASDAS change at 12 months, several associations with other variables were found to be statistically significant. Proceeding with HLA-B27, this variable also appears to be correlated with ASDAS at 12 months for patients taking infliximab, in line with an association previously found in the networks. Again, it seems that HLA-B27 positive patients respond better to infliximab's therapy than HLA-B27 negative patients. Moving into gender, it is possible reject the null hypothesis for gender and both ASDAS and ASDAS improvement at 12 months for all three biologics studied. It was noted that the number of male patients with smaller ASDAS is greater than expected, while with greater ASDAS is smaller than expected. On the other hand, for female patients, the opposite was noticed. In turn, for ASDAS change, male patients also seem to experience better improvements, which is in line with the literature, where male patients are reported to respond better to anti-TNF therapies. Next, regarding BMI, correlation between this variable and ASDAS/ASDAS improvement in patients treated with adalimumab was one of the strongest found in the networks. Indeed, looking into the p-values obtained, only for adalimumab's patients it's possible to conclude that the pairs BMI & ASDAS at 12 months and BMI & ASDAS change at 12 months are related. It was noted that for patients with higher BMI, the number of patients with higher ASDAS after 12 months is greater than expected, as well as the number of patients that experience smaller improvements, confirming the results previ-

ously found. On the other hand, concerning age at biologic onset/ disease duration, these variables did not present any clear association with any outcome in the networks, nevertheless, according to the literature, younger age and smaller disease duration are predictors of good clinical response to anti-TNF therapy. Through the Chi-square tests, these variables appeared related with ASDAS at 12 months for adalimumab and etanercept, being noticed that the number of younger patients/ with smaller disease duration presenting smaller ASDAS after 12 months was higher than expected and presenting higher ASDAS at 12 months was smaller than expected, which seems to be in line with the literature. The fact that this association was not identified for infliximab's patients might indicate that infliximab is a good therapeutic option for older patients/ patients with higher disease duration, since these patients don't seem to respond worse with age. Finally, regarding education, ASDAS at 12 months and education were found to be related for infliximab's data, nevertheless, it was not possible to identify any pattern in the data. Regarding VAS patient at baseline, the results from the Chi-square tests presented a statistically significant relationship with ASDAS improvement for patients being treated with adalimumab and etanercept, where patients with higher VAS patient were associated with higher improvements and patients with smaller VAS patient with smaller improvements, For etanercept, VAS patient at baseline also appeared related with ASDAS at 12 months, where a smaller VAS patient at baseline is more associated with smaller ASDAS after 12 months, while greater VAS patient at baseline seems to be more associated with higher ASDAS values. Finally, for infliximab, the Chi-square tests did not reveal any statistically significant association between VAS patient at baseline and any other variable. These findings seem to indicate that a higher VAS patient at baseline can be translate into a greater room for improvement, but not as a predictor of smaller disease activity after 12 months of therapy. Concerning VAS doctor, the results from the Chi-square tests only found a statistically significant relationship between this variable and ASDAS improvement 12 months after therapy initiation for patients being treated with etanercept, where the number of patients with a smaller VAS doctor at baseline that experience smaller improvements is greater than expected and the number of patients with a greater VAS doctor at baseline that experience greater improvements is greater than expected. Finally, regarding CRP at baseline, a statistically significant relationship was identified for all three biologics with ASDAS improvement , but not with ASDAS: for smaller levels of CRP at baseline, the number of patients experiencing smaller improvements is higher than expected and the number of patients experiencing higher improvements is smaller than expected. For higher levels of CRP at baseline, the opposite was identified. Nevertheless, as mentioned, these findings should be evaluated carefully, since patients with higher disease activity at baseline, and subsequently a greater ASDAS, have more room for improvement.

5. Conclusions

Due to the need of improving decision-making in ankylosing spondylitis, it is essential to improve the prognosis process, by being able to predict the most adequate treatment option according to the patient's characteristics. In this work, the approach used allowed to identify potential risk factors associated with the different AS therapies, working as a preliminary study in trying to distinguish between different anti-TNF agents.

DBNs are graphical representations of probabilistic dependencies among random variables over time, being a major strength of this approach the explicit representation of the relations between the variables. Additionally, the R package chosen to build the DBNs allows the presence of missing data, which was also a great advantage due to the considerable amount of unobserved values in the dataset. Still, several processes were conducted in order to try to reduce them, as removal of patients with a great amount of missing values and imputation, since drawing conclusions from datasets with a great amount of missing values might not lead to the best conclusions. However, both these methods can also significantly impact the conclusions drawn from data. On the other hand, to build the DBN models, the variables were required to be discretized, which also impacts the obtained results. Finally, to achieve more reliable conclusions, Chi-square tests were performed as a complement of BNs.

Next, it is noteworthy the different results obtained for each therapy studied. Greater BMI was found to be a predictor of worse response for adalimumab's patients, with greater BMI values being associated with a poorer response to therapy and smaller values associated with a better response. Higher age at therapy onset/ disease duration were also identified as predictors of worse response for adalimumab's and etarnecept's patients. Finally, being HLA-B27 positive appeared as a good predictor of response for infliximab's patients. Moreover, literature findings were also corroborated with this study, as it was the case of gender, where male patients were found to respond better to all three therapies. Additionally, it was also noticed that the therapy's effect is already visible after 3 months, being hypothesized that an absence of response, or a smaller response after this period might be an indi-

cator of therapy inefficiency.

Regarding suggestions for future work, first, it could be interesting to study not only the first therapy received by patients, but also the subsequent ones, since patients respond differently depending on whether they received previously another biological therapy or not. Second, concerning the outcomes studied, the occurrence of adverse events is also one of the greatest reasons for withdrawing therapy, and could also be studied and evaluated. Furthermore, the drug survival time can also be taken into account in future studies. On the other hand, to learn the networks, different algorithms besides `bnstruct` R package can be used and compared. Additionally, DBNs can be used to predict the patient's temporal evolution, by predicting how the dynamic variables evolve over time, and thus perform simulation analyses about the potential effectiveness of different therapies.

Acknowledgements

This document was written and made publically available as an institutional academic requirement and as a part of the evaluation of the MSc thesis in Biomedical Engineering of the author at Instituto Superior Técnico. The work described herein was performed at INESC-ID and IT, in collaboration with the EpiDoC Unit, CEDOC, during the period of March 2020-June 2021, under the supervision of Prof. Alexandra Carvalho and Prof. Susana Vinga, and within the scope of the Project PREDICT (PTDC/CCI- CIF/29877/2017), funded by Fundo Europeu de Desenvolvimento Regional (FEDER), through Programa Operacional Regional LISBOA (LISBOA2020), and by national funds, through Fundação para a Ciência e Tecnologia (FCT). The author would like to thank Prof. Susana Vinga and Prof. Alexandra Carvalho, for their usefull insights and suggestions, as well as everyone involved in the PREDICT project and in the Systems Biomedicine group, for the insightful discussions.

References

[1] S. Arends, E. Brouwer, E. van der Veer, H. Groen, M. K. Leijsma, P. M. Houtman, T. L. Th A Jansen, C. G. Kallenberg, and A. Spoorenberg. Baseline predictors of response and discontinuation of tumor necrosis factor-alpha blocking therapy in ankylosing spondylitis: a prospective longitudinal observational cohort study . *RMD Open*, 2011.

[2] Arthritis Society. Ankylosing Spondylitis. `https://arthritis.ca/about-arthritis/arthritis-types-(a-z)/types/ankylosing-spondylitis`. Accessed: 2020-04-24.

[3] Assessment of SpondyloArthritis International Society. Physiology, acute phase reactants. `https://www.ncbi.nlm.nih.gov/books/NBK519570/`. Accessed: 2020-04-30.

[4] H. Canhão, A. Faustino, F. Martins, and J. E. Fonseca. Reuma. PT - the rheumatic diseases portuguese register. *Acta Reumatologica Portuguesa*, 36:45–56, 2011.

[5] A. M. Carvalho. Scoring functions for learning bayesian networks. *Inesc-id Tec. Rep*, 2009.

[6] A. R. Cravo, V. Tavares, and J. Canas Da Silva. Terapêutica anti-TNF alfa na espondilite anquilosante. *Acta Medica Portuguesa*, 19(2):141–150, 2006.

[7] M. N. de Almeida Rodrigues de Sousa. Advances in probabilistic graphic models, 2017.

[8] L. E. Dean, G. T. Jones, A. G. Macdonald, C. Downham, R. D. Sturrock, and G. J. Macfarlane. Global prevalence of ankylosing spondylitis. *Rheumatology (United Kingdom)*, 53(4):650–657, 2013.

[9] EULAR. Horizon 2020 framework programme eular's position and recommendations horizon 2020 framework programme eular's position and recommendations. 2011.

[10] A. Franzin, F. Sambo, and B. Di Camillo. Bnstruct: An R package for Bayesian Network structure learning in the presence of missing data. *Bioinformatics*, 2016.

[11] E. Gremese, S. Bernardi, S. Bonazza, M. Nowik, G. Peluso, A. Massara, B. Tolusso, L. Messuti, M. C. Miceli, A. Zoli, F. Trotta, M. Govoni, and G. Ferraccioli. Body weight, gender and response to TNF- blockers in axial spondyloarthritis. *Rheumatology*, 53:875–881, 2014.

[12] J. R. Maneiro, A. Souto, E. Salgado, A. Mera, J. Gomez-Reino. Predictors of response to TNF antagonists in patients with ankylosing spondylitis and psoriatic arthritis: systematic review and meta-analysis. *RMD Open*, 2014.

[13] J. Sieper, J. Braun, M. Rudwaleit, A. Boonen, A. Zink. Ankylosing spondylitis: an overview. *Annals of the Rheumatic Diseases*, 2002.

[14] G. Layh-Schmitt and R. A. Colbert. The interleukin-23/interleukin-17 axis in spondyloarthritis. *Current Opinion in Rheumatology*, 20:392–397, 2008.

[15] P. A. Lord, T. M. Farragher, M. Lunt, K. D. Watson, D. P. Symmons, and K. L. Hyrich. Predictors of response to anti-TNF therapy in ankylosing spondylitis: Results from the British Society for Rheumatology Biologics Register. *Rheumatology*, 49(3):563–570, 2009.

[16] M. Lorenzin, A. Ortolan, P. Frallonardo, F. Oliviero, L. Punzi, and R. Ramonda. Predictors of response and drug survival in ankylosing spondylitis patients treated with infliximab. *BMC Musculoskeletal Disorders*, 16(1):10–17, 2015.

[17] G. J. MacFarlane, E. Pathan, G. T. Jones, and L. E. Dean. Predicting response to anti-TNFα therapy among patients with axial spondyloarthritis (axSpA): Results from BSRBR-AS. *Rheumatology (United Kingdom)*, 59(9):2481–2490, 2020.

[18] F. Mahmood and P. Helliwell. Ankylosing spondylitis: A review. *European Medical Journal*, 2017.

[19] A. Meyer-Baese and V. Schmid. *Statistical and Syntactic Pattern Recognition*. Elsevier, 2014.

[20] K. Nas, K. Yildirim, R. Cevik, S. Karatay, A. Erdal, O. Baysal, Z. Altay, A. Kamanli, Y. Ersoy, A. Kaya, B. Durmus, O. Ardicoglu, I. Tekeoglu, M. Ugur, A. J. Sarac, K. Senel, A. Gur, and S. Ozgocmen. Discrimination ability of ASDAS estimating disease activity status in patients with ankylosing spondylitis. *International Journal of Rheumatic Diseases*, 13:240–245, 2010.

[21] National Health Service England. Treatment ankylosing spondylitis. `https://www.nhs.uk/conditions/ankylosing-spondylitis/treatment/`. Accessed: 2020-04-25.

[22] Spondylitis Association of America. Overview of Ankylosing Spondylitis. `https://spondylitis.org/about-spondylitis/types-of-spondylitis/ankylosing-spondylitis/`. Accessed: 2020-04-24.

[23] C. Stolwijk, M. van Onna, A. Boonen, and A. van Tubergen. The Global Prevalence of Spondyloarthritis: A Systematic Review and Meta-Regression Analysis. *Arthritis Care Research*, 2015.

[24] L. E. Sucar. Dynamic and temporal bayesian networks. `https://ccc.inaoep.mx/~esucar/Clases-mgp/Notes/c9-dbn.pdf`. Accessed: 2020-05-02.

[25] N. Vastesaeger, D. Van Der Heijde, R. D. Inman, Y. Wang, A. Deodhar, B. Hsu, M. U. Rahman, B. Dijkmans, P. Geusens, B. V. Cruyssen, E. Collantes, J. Sieper, and J. Braun. Predicting the outcome of ankylosing spondylitis therapy. *Annals of the Rheumatic Diseases*, 70(6):973–981, 2011.

[26] A. Zandonà, R. Vasta, A. Chiò, and B. D. Camillo. A Dynamic Bayesian Network model for simulation of disease progression in Amyotrophic Lateral Sclerosis patients. *BMC Bioinformatics*, 2017.

[27] W. Zhu, X. He, K. Cheng, L. Zhang, D. Chen, X. Wang, G. Qiu, X. Cao, and X. Weng. Ankylosing spondylitis: etiology, pathogenesis, and treatments . *Nature*, 2019.

[28] J. Zochling. Measures of symptoms and disease status in ankylosing spondylitis. *Arthritis Care and Research*, 2011.

[29] J. Zochling, J. Braun, and D. van der Heijde. Assessments in ankylosing spondylitis. *Best Practice and Research: Clinical Rheumatology*.