Analysis of Electronic Medical Records of Rheumatoid Arthritis Patients on Biological Therapies
A Reuma.pt Study

João Pedro Bento Machado Marques de Freitas

Thesis to obtain the Master of Science Degree in
Biomedical Engineering

Supervisors: Prof. Alexandra Sofia Martins de Carvalho
Dr. Helena Cristina de Matos Canhão

Examination Committee
Chairperson: Prof. João Pedro Estrela Rodrigues Conde
Supervisor: Dr. Helena Cristina de Matos Canhão
Member of the Committee: Dr. Helena Isabel Aidos Lopes

June 2015
Dedico este trabalho à minha família.

Dedico este trabalho à minha família.
Acknowledgments

First of all I would like to express my sincere thanks to my co-supervisor Prof. Dr. Helena Canhão for providing data from Reuma.pt for this thesis. I would also like to thank to my university supervisors, Prof. Dr. Alexandra Carvalho and Prof. Dr. Susana Vinga, who provided me with their opinions, proofread the draft material and with whom I had many meetings during this thesis.

I am also thankful for the opportunity that my supervisors gave me of having a grant to work on the FCT project InteleGen (PTDC/DTP-FTO/1747/2012) and I also want to thank for the availability of Prof. Dr. Conceição Amado.

Finally, I would like to thank to my family for all the support they gave to me.
**Resumo**

**Motivação** A Artrite reumatoide (AR) é uma das principais causas de invalidez no mundo ocidental e está associada à redução da qualidade e esperança de vida. Adicionalmente, cerca de 60% dos doentes não respondem ao tratamento e é sabido que a intervenção precoce pode prevenir, entre outras coisas, danos irreversíveis nas articulações.

**Objectivos** Analisar registos de saúde electrónicos (RSE) do Reuma.pt e aumentar o conhecimento sobre o uso de medicamentos biológicos para AR na prática clínica. Explorar os dados e reconhecer padrões nestes. Estimar as chances de remissão para cada biológico.


**Resultados** O tocilizumab foi a terapia mais eficaz na maioria das análises (Risco relativo (RR), 2.99, Intervalo de confiança (IC)95%, 1.33 to 6.72, P=0.008, ref: etanercept). Não houve remissões com o anakinra. Em comparação com etanercept, remissão foi 73% menos provável com adalimumab e 82% menos provável com rituximab. A duração da AR, idade, peso, e avaliação global do paciente estiveram associadas negativamente com a remissão.

**Conclusões** O tocilizumab foi a terapia mais eficaz, o anakinra não induziu remissão e o adalimumab e rituximab foram os menos eficazes em comparação com a o etanercept. Devem-se fazer mais estudos sobre este assunto.

**Palavras-chave:** Artrite reumatoide, Terapias biológicas, Registos de saúde electrónicos, Regressão de Cox
Abstract

Motivation  Rheumatoid arthritis (RA) is one of the lead causes of disability in the western world and is linked with reduced life quality and expectancy. In addition, about 60% of the patients fail treatment and is known that early intervention might prevent, among other things, irreversible damages to the joints.

Objectives  Investigate Electronic medical records (EMR) from Reuma.pt and increase the base knowledge about the use of biologic drugs for RA in clinical-practice. Explore the data and find meaningful patterns in it. Estimate the chances of remission for the different biologics.

Materials & Methods  EMR from Reuma.pt database. All the 436 patients are diagnosed with RA and are or were treated with biological agents. Descriptive analysis of the data. Summary measures analysis. Incidence rates, incidence density rates and Kaplan-Meier curves for drug "survival", therapy response and remission. Stratified Cox proportional hazards regression for the first remission. Model selection. Recurrent events analysis.

Findings  Tocilizumab was the most efficacious therapy in most of the analysis (Hazard Ratio (HR), 2.99, 95% Confidence interval (CI), 1.33 to 6.72, P=0.008, ref: etanercept). There was no remissions with anakinra. In comparison to etanercept, remission was 73% less likely with adalimumab and 82% less likely with rituximab. RA duration, age, weight and patient global assessment VAS at baseline were negatively associated with remission.

Conclusions  Tocilizumab was the most efficacious therapy and anakinra had no remission and rituximab and adalimumab were less efficacious comparing to etanercept. Further studies need to be done.

Keywords:  Rheumatoid arthritis, Biological therapies, Electronic medical records, Cox regression
# Contents

Acknowledgments ........................................ iii
Resumo ................................................... iv
Abstract ................................................ v
List of Tables ........................................... ix
List of Figures ......................................... xi

## Nomenclature

- Nomenclature ........................................ 1
- Glossary ............................................... 1

## 1 Introduction

1.1 Motivation ......................................... 3
1.2 Data mining ........................................ 5
1.3 Objectives .......................................... 5
1.4 Thesis outline ..................................... 6

## 2 Rheumatoid Arthritis

2.1 Description ......................................... 7
2.2 Disease activity scores ............................. 9
2.3 Measuring functional disability ................. 10
2.4 Remission criteria ................................ 10
2.5 Response criteria .................................. 11
2.6 Treatment of rheumatoid arthritis ............. 11

## 3 Materials and Methodology

3.1 Data .................................................. 15
3.2 Methods ............................................. 18
  3.2.1 Smoothing ...................................... 19
  3.2.2 Summary measures .............................. 19
  3.2.3 Measures of occurrence and association .... 19
  3.2.4 Kaplan-Meier estimator ....................... 21
  3.2.5 Cox Proportional Hazards Model .......... 21
List of Tables

2.1 Disease activity scores formulae table. ......................................................... 9
2.2 ACR/EULAR boolean and index based definition of remission for clinical trials and clinical practice. ................................................................. 10
2.3 EULAR response criteria ........................................................................... 11
2.4 Guidelines for biological therapies approved for Rheumatoid Arthritis .......... 13

3.1 Number of patients per biologic number. ..................................................... 17
3.2 Percentage (number) of patients per drug in cohort dataset. ....................... 18
3.3 Data layouts for different recurrent event models using two hypothetical subjects. . . 24

4.1 Baseline table. .................................................................................................. 30
4.2 Summary measures table. ............................................................................... 30
4.3 Incidence rate by drug per 1000 person-months over the follow-up period of 4 years. ... 37
4.4 IRR and 95% CI of tocilizumab (TCZ) comparing to etanercept (ETN) by age quartile. ... 39
4.5 Cox Proportional Hazards regression results for predictors of remission, Hazard Ratio (95% CI). ........................................................ 42
4.6 Magnitude of confounding table .................................................................... 42
4.7 Best models obtained. Swollen joint count (SJC), rheumatoid factor (RF), extra articular manifestations (EA Manif.). .................................................. 46
4.8 Cox Proportional Hazards models for recurrent events. .................................. 49
4.9 Grambsch-Therneau test p-values for the PWP-CP model covariates ............... 49
List of Figures

2.1 The American College of Rheumatology treatment algorithm after rheumatoid arthritis diagnosis. ................................................................. 12

3.1 Patients time-line. ................................................................. 16

4.1 Body distribution of the swollen and tender joints at baseline. ............... 29
4.2 Hands and feet distribution of the swollen and tender joints at therapy baseline. ............... 31
4.3 DAS28 by drug, over the 4 years follow-up period. ........................... 32
4.4 HAQ, 4 years follow-up. ....................................................... 33
4.5 Disease activity conditional density estimate by therapy. ....................... 34
4.6 Boxplots of the therapy duration. ............................................. 35
4.7 Summary measures plots. ...................................................... 36
4.8 Kaplan Meier estimates stratified by biologic drug. ........................... 37
4.9 Unadjusted incidence rate ratio of the other drugs in comparison with etanercept and stratified by age, concomitant DMARD, descent and time since onset of disease. ............... 39
4.10 Evolution of different scores during backward elimination with different data sets and different criteria to eliminate a variable. ................................................. 45
4.11 Kaplan Meier estimate for recurrent events. ................................... 48

A.1 Kaplan-Meier of remission between naive and non-naive patients. ............ 67
A.2 Kaplan-Meier of drug switch between naive and non-naive patients. ............ 68
A.3 Kaplan-Meier of remission between patients with DMARD and without DMARD at baseline. 68
Nomenclature

$L$  Likelihood
$n$  Number of observations
$p$  Number of parameters
$S(t)$  Survival function
$t$  Time

ABT  Abatacept
ACR  American College for Rheumatology
ADA  Adalimumab
AG  Andersen and Gill counting process independent increment model
AIC  Akaike information criterion
AICc  Akaike information criterion corrected
AKR  Anakinra

bDMARD  Biologic disease modifying anti-rheumatic drugs
BIC  Bayesian information criterion
CCP  C-reactive protein
cyclic citrullinated peptide
CDAI  Clinical disease activity index

Cox PH  Cox Proportional Hazards

CP  Counting processes
CRP  CRP

DAS  Disease activity score

DAS28  Disease activity score for 28 joints
Chapter 1

Introduction

1.1 Motivation

Rheumatoid Arthritis (RA) is one of the leading causes of disability in the western world and is linked with reduce life quality, decreased life expectancy and increased risk of cardiovascular diseases [1]. Moreover, RA affects 0.5% to 1% of the general population worldwide being the most common inflammatory arthritis [2]. Because therapy aims not only to control the symptoms during flares of increased disease activity but also to prevent irreversible damage to the joints, early recognition and intervention can improve function, slow structural progression, mortality and also reduce the financial burden for the patients and society.

Today’s therapies are estimated to work only in 60% of the patients, the therapy selection follows no specific criteria and the causes for RA remains unknown. Furthermore, the time used to find an effective therapy, or therapies, for each patient is of special importance since it may give opportunity for the disease to cause irreversible damage in the joints and bone erosion. However, many studies have helped us understand which factors enhance the disease expression in individuals who have RA [2] and in the future they might help us to find the right therapy for the right patient.

The ultimate goal of epidemiology, biostatistics and other sciences that use statistics or computational methods to deal with biomedical data is to provide evidence that can be used not only for policy making on public health but also on daily clinical practice to support clinical decisions. In the past, the impact of such measures went well beyond public health and had deep reverberation in society and economy.

Nowadays, the developed countries face increasingly aging demographic with rising health care costs consuming increasing portion of the nation’s gross domestic product [3]. This gives data sciences a reinforced role on the organization, strategic planning and optimization of the allocation of resource. Ultimately, the new evidences should translated to new public health policies in a way that the sustainability, universality and equity of the health systems are preserved.

Plataforma Mais Saúde, an association of rheumatic diseases patients in Portugal, states that rheumatic diseases are the first cause for early retirement and that an untreated or late diagnosed patient is two times and a half more costly for the Portuguese State than a early diagnosed [4]. In Portugal, the
rheumatic diseases are responsible for 43% of the cases of work absenteeism and Portugal is one of the countries that more spends per capita with drugs used in the treatment of this group of diseases [5]. In fact, the emergence of biological therapies for RA increased the cost of treatment of rheumatic diseases. However, they yield better results than their synthetic counterparts and have important clinical benefits for the patients.

This thesis focus on studying biological therapies for RA. This relatively new drugs are called this way because they are produced in other organisms by recombining human genes of antibodies against a specific target. This genetically engineered organism are used to produce these molecules because they are too large and complex to be produced synthetically. In the case of RA this antibodies act as inhibitors of a group of molecules responsible for the cell signaling and activation of the immune system cells and the autoimmune response. Because of their production method the price is much higher than the synthetically produced disease modifying anti-rheumatic drugs (DMARDs). For example, a 25 mg tab of methotrexate can cost 600 times more than a etanercept 25 mg vial [2]. For this reason, and also because of the increased risk of infection, they are only used after therapy with DMARDs has failed.

Electronic medical records (EMR) are the technological successor of the paper-based medical records. EMRs centralize and accumulate the medical history data and any other information relevant for the medical domain in databases to provide informed and planned patient care. The data stored in EMRs is more accessible, usable and exists in larger quantities. In addition, EMRs can also guide the decision making process by enabling tools that can instantly help the physician visualize and analyze the data. Another application of EMRs is to monitor the quality of care and for conducting scientific research that deals with important clinical problems.

Since June 2008, the Rheumatic Diseases Portuguese Register (Reuma.pt) is collecting data from rheumatic patients receiving biological therapies or synthetic DMARDs. This register was created by the Portuguese Society of Rheumatology (SPR) and includes patients with RA, ankylosing spondylitis, psoriatic arthritis and juvenile idiopathic arthritis. At each visit demographic and anthropometric data, life style habits, work status, co-morbidities, disease activity and functional assessment scores, previous and current therapies, adverse events, reasons for discontinuation and laboratory measurements are registered. The register includes all Rheumatology departments assigned to the Portuguese National Health Service, a public-private partnership and some private centers [6]. Reuma.pt goal is to keep track of the patients’ disease progression, to support the clinical practice on a daily basis and to provide data for studies about therapies’ efficacy and safety.

The main focus of this thesis is a data set from EMR data extracted from Reuma.pt obtained on mid-June 2014 that includes patients from two health centers with a total of 9305 observations and 426 patients. All the patients are diagnosed as having RA and are being or were treated with biological therapies. Since all subjects were treated with biologics the analysis will be mostly about this new drugs.
1.2 Data mining

Data mining is considered a step of several analysis processes for databases, including the Knowledge Discovery in Databases (KDD) process and the Cross-Industry Standard Process for Data Mining (CRISP-DM) developed by a consortium of European companies and currently used by IBM [7, 8]. The goal of data mining is to find patterns of interest in databases using methods from fields like artificial intelligence, machine learning and statistics. The most striking difference between data mining and statistics is that the first analyzes data without any a priori hypothesis and analyzes mostly observational data from databases without any kind of experimental designed or measurement control. Data mining seems to be a rebranding that comprises many fields that already existed.

The objective of experimental design is to collect data in a way that represents best the underlying population. The main way to accomplish it is by randomization and by diving the experimental units in similar groups (blocks) to reduce sources of heterogeneity and control for confounding variables.

In a statistical study one has a well-defined hypothesis that will be investigated, a specifically chosen population, carefully collected data, control for confounding variables and a set of methods inference tools with well know properties used to generalize the results from a sample to the studied population.

The last decade notoriety of data mining is due in part to the growing quantity of information stored in databases and with how little companies know what to do with it. This phenomenon has even more been pronounced by the promise from some software solutions that the investment will create huge returns.

1.3 Objectives

This study investigates EMR data collected in the Reuma.pt and aims to increase the base knowledge about the use of biologic drugs for RA in clinical-practice. More precisely, it aims to measure the effect of the different biologic treatments and to discover what other factors might influence the progression of the disease activity and functional disability on patients with RA.

This work also aims to study the drug discontinuation, since it is a overall marker of the success of a treatment and depends on the efficacy of the drug, its side effects and doctor and patient preferences.

Ultimately, we also hope to be determine what therapy is more likely to succeed according to the characteristics of the data.

In order to meet this aims, the beginning of this thesis will focus on defining the grounds and contextualizing the rest of the work by reviewing some of the literature about rheumatoid arthritis, its treatments, guidelines, epidemiology, genetics and current criteria used to assessment of disease activity and functional ability.

Then it is paramount to find appropriate ways to analyze Reuma.pt data and to understand the implications and limitations of the results that will be produced. For this reason, it will be reviewed literature about epidemiology, longitudinal data analysis, survival analysis, machine learning and other data sciences or methods.

Next, the data processing and data analysis have to be outlined and prepared. This step consists of
plan and script all steps on the data processing and analysis to produce the results in the form of figures and tables. From the processing of the original data it will be created multiple datasets that are suitable for most of the methods chosen. Thus, R statistical software and other appropriate tools will be used.

Specifically, we want to explore the dataset with a set of summary measures and plots that can be easily understandable and interpretable. In addition, we want to describe the evolution in time of key variables by drug.

Then, we want to measure the effect of the different biologic treatments and other factors on the progression of the disease activity on the patients with RA. Assess the effectiveness of the biologic therapies with and without DMARDs and compare the therapy response between biological therapies’ naïve and non-naïve patients.

After the results are presented clearly and systematically, conclusions can be drawn from the research, methods implications and limitations can be discussed and ideas for future work can be referred.

1.4 Thesis outline

Chapter 1 gives the motivations, aims and objectives of this thesis.

Chapter 2 is about rheumatoid arthritis disease, its treatments and Portuguese and European guidelines. The conditions for introduction in biologic therapies are described and the background for our data is established.

In Chapter 3 we can find the methodology, which is divided in three sections: Data, Method and Data Analysis. On the first section the structure of the original data set and the processing methods to construct the other data sets is described. The second section gives a theoretical review of the methods used in this work and the third section describes how the results in Chapter 4 were obtained.

Chapter 4 starts by representing the data with some descriptive statistics and summaries. Then, it presents the estimates for the effect of the different therapies and variables on remission for the different analysis.

Chapter 5 summarizes and interprets the results in light of the initial objectives, discusses the advantages and limitations of the proposed solutions and also refers possible applications and future work.
Chapter 2

Rheumatoid Arthritis

In this chapter we describe the background inherent with the data used in this work. By other words, the disease, treatments and guidelines are analyzed with the objective of understanding the posed problem and to elicit good solutions to analyze the data.

2.1 Description

RA is a systemic inflammatory disease that causes a chronic inflammatory response of the synovial membrane with consequent cartilage destruction and bone erosion within the joint. The main clinical manifestations of the disease are joint pain, stiffness and excess synovial fluid in symmetrical joints causing swelling of the synovium, mostly, on the hands and feet. The disability in RA patients develops most rapidly during the first 2 years of disease when the constant insult to the synovium leads to the gradual replacement of this tissue by fibrous tissue.

Nonetheless, the cause of RA and why the synovium is the primary target remains unknown. However, it is believed that repeated inflammatory insults in a genetically susceptible individual might contribute to decline of tolerance and subsequent autoimmunity. These insults can be environmental factors, such as smoking or infectious agents that can contribute to the initiation and perpetuation of RA [2].

While RA prevalence ranges from 0.5-1.0% in the general population, in women it is two to three time higher than in men. According to the Portuguese Directorate-General of Health (DGS), in Portugal this proportion is 4:1 [9]. Consequently, the hormonal part in the development of RA have been widely studied to understand what factor might cause RA.

For reasons not yet fully elucidated, the oral contraceptive pill may have a protective effect on the development of RA, even if different levels of female sex hormones does not increase the risk of RA. Moreover, men who have RA present lower testosterone levels and higher incidence of RA development occurs at the beginning of menopause and during the postpartum period after the first pregnancy, specially for women that breastfeed. This suggests that RA may be related with an increase in prolactin or an abnormal response to it [10, 11].

Diet also seems to be relevant in reducing the incidence of RA. One study states that the intake
of vitamin D is inversely associated with RA [12], while another states that a vegetarian diet produced significant general improvement of the patients [13]. However, there are not many studies that associate the incidence of RA with the dietary behavior.

On the other hand, many evidence have been found that support the existence of genetics susceptibility and severity factors of RA. For example, chinese and japanese populations have a lower prevalence of RA (0.2-0.3%) and studies in Africa failed to find any RA cases, while native populations from North America, like the Pima indians (5.3%) and Chippewa Indians (6.8%), have higher prevalences [2].

In addition, the class II major histocompatibility complex (MHC) genes, specifically the genotypes of the hypervariable region of the human leukocyte antigen-DR4 (HLA-DR4) has been associated to an increased risk of RA and also to the severity of the disease. The risk of having RA is 4 to 5 greater in individuals with the HLA-DR4 genotype. However, it is estimated that only 50% of the genetic contribution to RA can be explained by HLA genotype. In addition, the alleles of the tumor necrosis factor (TNF) and many other genes have also been investigated but their contribution to the disease are yet not well defined [2, 10].

RA can also affect other organs and systems with manifestations on the eyes, lungs and cardiovascular system. One of the main characteristics of RA patients is anemia caused by the chronic inflammation [11]. In a 40 years prospective study, a cohort of 100 patients with RA had increased mortality with a median life expectancy reduced by 10 years for men and 11 years for women compared to the general population. Also in this study, the primary cause of death was cardiovascular disease [14]. Other factors that also contribute for the excess mortality among RA patients are infections, renal disease and respiratory failure [2]. For this reasons, early suppression the RA inflammation process can not only have impact on the morbidity of affected population but too on its mortality.

Inflammation, infection or tissues necrosis causes the production by the liver of group of proteins known as the acute-phase proteins that characterize what is known as the acute-phase response. The presence of this acute-phase proteins, as well as increased age or anemia, promotes the erythrocytes aggregation. So, the erythrocytes sedimentation rate (ESR) is an indirect and non-specific way of assessing the patient acute-phase response. The measurement of the level of C-reactive protein, an acute-phase protein, has the advantage that they increase and decrease rapidly compared to ESR in the presence and absence of inflammation [11].

The rheumatoid factors (RF) and anti-cyclic citrullinated peptide antibodies (anti-CCP) are antibodies directed against proteins of the own individual. The RF is found in 70% to 90% of the RA patients and the anti-CCP is found in 75% of the same [11]. Other autoantibodies that recognize either joint antigens or systemic antigens are related with RA, but RF and anti-CCP are the only biomarkers with significant sensitivity and specificity to be considered useful for RA prognostic and diagnostic in clinical practice. It is known that cartilage and bone erosion are more common in RF-positive or anti-CCP-positive patients and that autoimmunity can be present before the unveiling of the first symptoms [2].
2.2 Disease activity scores

A score gives a value to the quantity being measured by gathering quantitative evidence about the patient state that can be later used to monitor the disease activity and decide the best treatment strategy. Scores like the disease activity score (DAS), the simplified disease activity index (SDAI) and the clinical disease activity index (CDAI) try to provide maximum information about the disease activity using a minimal number of variables. The scores on Table 2.1 are of great importance to clinical practice as well as to clinical trials because they provide standard measures and remission criteria that can be used and compared across studies.

<table>
<thead>
<tr>
<th>Score</th>
<th>Formula</th>
<th>Cutpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DAS</strong> (0-12.2)</td>
<td>$0.54 \sqrt{RAI} + 0.065 \sqrt{SJC14} + 0.33 \ln ESR + 0.0072 \times GH$</td>
<td>$&lt; 1.6/ \leq 2.4/ \leq 3.7/ &gt; 3.7$</td>
</tr>
<tr>
<td><strong>DAS28</strong> (0-9.8)</td>
<td>$0.56 \sqrt{TJC28} + 0.28 \sqrt{SJC28} + 0.70 \ln ESR + 0.014 \times GH$</td>
<td>$&lt; 2.6/ \leq 3.2/ \leq 5.1/ &gt; 5.1$</td>
</tr>
<tr>
<td><strong>SDAI</strong> (0-86)</td>
<td>$TJC28 + SJC28 + PtGA + PhGA + CRP$</td>
<td>$\leq 3.3/ \leq 11/ \leq 26/ &gt; 26$</td>
</tr>
<tr>
<td><strong>CDAI</strong> (0-76)</td>
<td>$TJC28 + SJC28 + PtGA + PhGA$</td>
<td>$\leq 2.8/ \leq 10/ \leq 22/ &gt; 22$</td>
</tr>
</tbody>
</table>

Table 2.1: Disease activity score (DAS and DAS28); simplified disease activity index (SDAI); clinical disease activity index (CDAI). Ritchie articular index (RAI); Tender joint count (TJC); swollen joint count (SJC). DAS joint counts is determined using 44 joints and DAS28, SDAI and CDAI using 28 joints. C-reactive protein (CRP) in mg/dL; erythrocyte sedimentation rate (ESR) in mm/h. DAS and DAS28 use the general health (GH) or patient global assessment (PtGA) on a 0 to 100mm visual analog scale (VAS) while SDAI and CDAI use the PtGA and the physician global assessment (PhGA) of disease activity on a 0-10cm VAS. Remission(REM); low disease activity (LDA); moderate disease activity (MDA) and high disease activity (HDA).

While swollen joint count (SJC) reflects the amount of inflamed synovial tissue, the tender joint count (TJC) is associated with the level of pain on the joints.

The Ritchie articular index (RAI) is an articular measurement for the assessment of joint tenderness in patients with RA. The RAI is the sum of the grades of tenderness of 52 evaluated joints (0 for not tender, 1 for tender, 2 for tender and causes wince and 3 for tender, causes wince and effort to withdraw) obtained by applying pressure over the joint margin of articular joints [15].

DAS and DAS28 have also modified formulae that include CRP instead of ESR and a constant that substitute the patient global assessment (PtGA) in case it is missing from the appointment register. Nevertheless, these scores have limitations and can not express all characteristics of the disease. While the mechanisms behind RA are not fully understand, the variables used in the scores will not describe completely the activity of the disease or the treatment effect in RA, too.
2.3 Measuring functional disability

The health assessment questionnaire (HAQ) measures the extent of the functional damage caused by the disease. HAQ has been validated in patients with a wide variety of rheumatic diseases including RA and so, it is not considered disease specific.

The HAQ has two versions, the full HAQ and the 2-page HAQ that contains the HAQ Disability Index (HAQ-DI), pain scale, and global health status scale. The HAQ-DI assesses the patient’s level of functional ability through 8 categories of questions about fine movements of the upper extremity, locomotor activities of the lower extremity, and activities that involve both upper and lower extremities. These categories are dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities. A HAQ-DI score cannot be calculated validly when there are scores for less than six of the eight categories.

The HAQ pain scale and the global health status scale are both visual analog scales (VAS) that assesses the patient condition over the past week in a 15-centimeter, double-anchored horizontal scale that starts at 0 (very well) and goes to 100 (very poor) [16, 17]. Bruce and Fries state in [17] that some researchers consider a difference of 0.10 on the HAQ-DI as clinically important. Others consider 0.22 the minimal clinical important difference.

2.4 Remission criteria

Remission is the state of lessening of disease symptoms without reaching cure and with the possibility of relapse. Remission criteria are fundamental decision boundaries in clinical trials and clinical practice. It is sought from remission criteria that the probability of relapse is minimal but as long as there is not complete understanding of the pathophysiology or a biomarker that describes completely the disease activity, these criteria will not be perfect. However, letting the patient know they are in remission may lead to periods of under treatment by giving the patients and physicians a false sense of security [18].

Some remission criteria were already presented in Table 2.1, but many others exist. For this reason, in 2011 a joint initiative from American College of Rheumatology (ACR), the European League Against Rheumatism (EULAR) and the Outcome Measures in Rheumatology Initiative (OMERACT) met to define a uniform remission criterion for RA in trials and practice. ACR/EULAR committee came up with the criteria presented in Table 2.2. Furthermore, ACR also recommends to include the feet and ankle joints extra to the 28-joint count when evaluating remission [19, 20].

<table>
<thead>
<tr>
<th>ACR/EULAR</th>
<th>Boolean-based</th>
<th>Index-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical trials</td>
<td>SJC, TJS, PtGA, CRP all ≤1</td>
<td>SDAI ≤3.3</td>
</tr>
<tr>
<td>Clinical practice</td>
<td>SJC, TJS, PtGA all ≤1</td>
<td>CDAI ≤2.8</td>
</tr>
</tbody>
</table>

Table 2.2: ACR/EULAR boolean and index based definition of remission for clinical trials and clinical practice. Swollen joint count (SJC) using 28 joints, tender joint count (TJS) using 28 joints, patient global assessment (PtGA) on a 0 to 10 scale, C-reactive protein (CRP) in mg/dL, simplified disease activity index (SDAI), clinical disease activity index (CDAI) [19, 21].

Regarding DAS28 remission criterion defined in Table 2.1 (DAS28<2.6), the EULAR recommenda-
tions consider that it is not enough strict [22]. Additionally, the DAS28 remission definition does not predict radiographic outcomes as well as the other remission criteria [19].

As explained in [23], the boolean-based criteria present a disadvantage because the PtGA is not only influenced by the RA disease activity. For this reason, many patients did not meet the boolean remission criteria even though they did not have signs of disease activity. Consequently, it is better to use index based criteria since they provide more flexible criteria for remission.

Nevertheless, in current clinical practice states of remission are frequently not kept over long periods and a more feasible goal might be to reach and sustain a state of very low disease activity, as defined by the score’s cut-points in Table 2.1. Ultimately there is still no consensus in what is best the definition for remission in RA and these criteria are still being validated.

2.5 Response criteria

The response criteria, validated and developed by the EULAR is present in Table 2.3.

<table>
<thead>
<tr>
<th>Current DAS28</th>
<th>DAS28 change from baseline (ΔDAS28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤3.2</td>
<td>&gt;1.2</td>
</tr>
<tr>
<td>3.2-5.1</td>
<td>moderate response</td>
</tr>
<tr>
<td>&gt;5.1</td>
<td>moderate response</td>
</tr>
</tbody>
</table>

Table 2.3: EULAR response criteria

The developers of the EULAR response criterion considered that a good responder must have a low disease activity, accordingly with the DAS28 cut-points, and that the score must decrease of 1.2 to be statistically significant. The 1.2 value comes from the fact that the variables used to calculate the DAS28 were transformed to have a Gaussian distribution and that this change is two times the measurement error (95% confidence).

2.6 Treatment of rheumatoid arthritis

Early identification and aggressive intervention in RA is fundamental to minimize the irreversible functional damage in the joints. Accordingly with the EULAR recommendations [22], the initial therapy approach for RA patients should consist of conventional synthetic disease-modifying antirheumatic drugs (csDMARDs or simply DMARDs), alone or in combination of two or three, as soon as the disease is diagnosed. On the first 6 months of disease, non-steroidal anti-inflammatory (NSAID) and low-dose glucocorticoids can be considered but they should be reduced the sooner the possible. The treatment target to achieve is remission by the ACR/EULAR criteria or low disease activity as defined by the different scores cut-points on Tables 2.2 and 2.1.

Therapy generally begins with methotrexate (MTX), the most commonly prescribed DMARD. In case of contraindications or intolerance, sulfasalazine (SSZ) or hydroxychloroquine (HCQ) are alternatives.
After initiating a DMARD, patients need to be monitored for potential side effects of the medications. If there is no improvement after three months of therapy the therapy should be adjusted to another DMARD or to a combination of DMARDs. In case remission or low disease activity has not been reached after six months, therapy should be adjusted. In case of poor prognostic the therapy should be adjusted to a biologic DMARD (bDMARD, biological therapy or biologic) in monotherapy or with DMARD(s).

The guidelines from EULAR and ACR are very similar. For this reason we present a simplified version of the ACR algorithm for RA treatment strategy in Figure 2.1.

ACR TREATMENT ALGORITHM

- Establish diagnosis of rheumatoid arthritis early
- Document baseline disease activity and damage
- Estimate Prognosis

Initiate therapy
- Patient education
- Start DMARD(s) within 3 months
- Consider NSAID
- Consider local or low-dose systemic steroids
- Physical therapy/occupational therapy

Periodically assess disease activity

Adequate response with decreased disease activity

Inadequate response (i.e. ongoing active disease after 3 months of maximum therapy)

Change/add DMARDs
- MTX naive
- Other mono R
- Combination R

Suboptimal MTX response
- MTX
- Other mono R
- Combination R

Other mono R

Biologics

Multiple DMARD failure

Symptomatic and/or structural joint damage

Surgery

Figure 2.1: The American College of Rheumatology (ACR) treatment algorithm after rheumatoid arthritis diagnosis. Methotrexate (MTX), prescription (R), disease-modifying antirheumatic drug (DMARD), nonsteroidal anti-inflammatory drug (NSAID). (Figure taken from [2]).

Accordingly with SPR [24], patients with RA are eligible for treatment with biologic agents when they present inadequate response or treatment failure to MTX or any other DMARD. It is considered inadequate response or treatment failure if the disease activity score (DAS28) is equal or superior to 3.2 for more than three months. The same applies for cases of intolerance, toxicity or refusal to take MTX or any other conventional DMARD. Biological therapy also applies for patients with a significant functional (worsing of the HAQ ≥ 0.22) or radiological degradation even if under conventional therapies present a DAS28 between 2.6 and 3.2.

The SPR guidelines also consider that after the first three months of therapy with biologic agents, a reduction of DAS28 equal or greater than 0.6 is considered as a clinical response to the therapy. After six months, the patients are responding if the DAS28 reduced 1.2 or more. In case of inadequate response,
Inadequate response to conventional DMARDS

**In association with MTX**
- abatacept
- adalimumab
- certolizumab
- anakinra
- etanercept
- golimumab
- infliximab
- tocilizumab

**In monotherapy**
- adalimumab
- etanercept
- certolizumab
- tocilizumab

Inadequate response to at least one TNF antagonist

**In association with MTX**
- abatacept
- rituximab
- tocilizumab

**In monotherapy**
- tocilizumab

Table 2.4: Guidelines for biological therapies approved for Rheumatoid Arthritis (Table taken from [24]).

The procedure is to switch to another biologic according with Table 2.4.

Biological therapies are divided in two classes: the tumor necrosis factor (TNF) antagonists (or anti-TNF) and the non-TNF antagonists. Anti-TNF is usually the first class of biologics given to RA patients. Included in the anti-TNF class we have adalimumab (ADA), etanercept (ETN), golimumab (GLM) and infliximab (IFX). As non-TNF antagonists we have abatacept (ABT), anakinra (AKR), rituximab (RTX) and tocilizumab (TCZ). Respectively, AKR and TCZ are an interleukin-1 (IL-1) and IL-6 receptor antagonists. Rituximab targets and destroys B-cells and abatacept prevents the second signal necessary for the T cell activation by inhibiting proteins of the cluster of differentiation (CD) 80 and 86.

Biologic therapies non-responding patients are less likely to respond to a second TNF antagonist. Intolerant patients may respond to a second TNF antagonist but are also less likely to tolerate a second. The same applies to a third TNF antagonist after a second TNF antagonist failure.

The long-term goal of treatment intervention with biologics is remission or at least a low disease activity translated DAS28 below 3.2, without significant functional or radiological degradation [24]. Because of their production method the price is much larger than the synthetically produced DMARDs. For this reason, and also because of the increased risk of infection, they are only used after therapy with DMARDs has failed [25]. The combination of biologics increases the risks of serious infection even more, as reported by Infarmed regarding the combination of anakira and etanercept [26].

To conclude, the ultimate goal of the treatment is to induce remission, or at least a low disease activity level, while structural damage is prevented, symptoms are reduced, function is improved and potential side effects of treatment are carefully monitored.
Chapter 3

Materials and Methodology

The object of study of this thesis is EMR extracted in June of 2014 from the *Reuma.pt* database. The patients records come from observational data from two health centers registered in *Reuma.pt*. There are 9305 observations from a total of 436 patients, 78 of which had already started biological therapies before the follow-up start. All the patients are diagnosed with RA and are or were treated with biological agents.

3.1 Data

The data consists of demographic information, pharmaceutical records, disease activity and disability indicators, response to biological therapy and other clinical data. The data can be described as retrospective observational and as an unbalanced longitudinal, also known as unbalanced panel data. The last description means that the patients have repeated measures during the follow-up period but there is not the same number of observations per patient. In addition, a patient can switch between therapies or be unenrolled on therapy during the follow-up period. As mentioned in Chapter 2, remission in RA, the primary outcome on this work, is not an easy outcome to access since subjects experience periods of disease remission and frequent relapses. The largest follow-up time for a patient in the data was almost 15 years and the smallest was just one appointment. See Figure 3.1 for a real example of the structure of the data.

Since it does not exist a control/reference group on placebo, DMARDs or a defined golden standard for the biologics, etanercept will be chosen as the reference group for comparisons between drugs in this work, due to the fact that it was the most prescribed drug. On all work the DAS28 used was the version with four variables and with the ESR.

In this work, the information about the various DMARDs and GCs was spread across different variables. This variables were collapsed to form a unique variable for the DMARDs and GCs. Furthermore, the dates of the appointments was encoded according with the season of the year to investigate possible seasonality of the disease flare or remission. Variables with a big number of levels were recategorized to ease the analysis and because some groups did not have enough observations for the methods to
converge. This happened, for example, with the variables for race, drinking habits and smoking habits. The 7 groups on the race variable were collapsed to 3 groups: African, Asian and European descent.

Missing data arise often in the context of data analysis and our data is not immune to that problem. Missing data creates some difficulties on the application of the methods and interpretation is based on the assumption that the observations and censoring is at random. It was observed that the farther in time the values are from baseline, higher is the degree of missingness of the data.

Figure 3.1: Time-line of randomly chosen patients since the beginning of biologic therapy. Black areas mean that the patient was in remission during that period. Colored areas mean that the patient has an active biologic therapy. The red vertical dashed line represents median follow-up for this patients. Not all biological therapies appear in this figure.
The original data was used to design a number of studies in order to investigate the association of the different variables to therapy response and disease remission. On a longitudinal dataset, the continuous and categorical variables can also be classified as time dependent and time independent, accordingly if they change or not with time. From a time-depended variable, like the number of days since the start of a particular therapy, and an event variable, like the disease activity or the drug switch, we can extract from the original dataset time-to-event data and analyze it accordingly with the time a subject remained event free or was lost to follow-up.

By observing different estimates for disease activity, there were few subjects at risk and almost no new events after 4 years. Machin et al. [27] suggests to limit the follow-up time to a certain time when the number of patients at risk is less than 15 or the Kaplan-Meier estimate starts to form a “plateau” with relatively big gaps between events. Therefore, the follow-up time was limited to 4 years and it was considered up to four recurrent remissions.

Having in attention that most of the subjects experienced different drugs at different times, only one of the drugs per subject was selected. This was done because the observations of the same subject for different drugs are correlated and the assumption of independent observations that many of the conventional methods in data analysis use needed to be followed. For example, if observations of the same subject is present in different groups being compared this groups are depended of each other.

This choice has the disadvantage that we discard a great part of the data but the advantage that making possible to study the effect of the therapies with a greater variety of methods and less restrictions and assumptions.

<table>
<thead>
<tr>
<th>Biologic Drug</th>
<th>Biologic number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Abatacept</td>
<td>1</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>61</td>
</tr>
<tr>
<td>Anakinra</td>
<td>8</td>
</tr>
<tr>
<td>Etanercept</td>
<td>144</td>
</tr>
<tr>
<td>Golimumab</td>
<td>30</td>
</tr>
<tr>
<td>Infliximab</td>
<td>110</td>
</tr>
<tr>
<td>Rituximab</td>
<td>5</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>34</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>393</td>
</tr>
</tbody>
</table>

Table 3.1: Number of patients per biologic number

Table 3.1 shows the number of patients per biologic number\(^1\). A subject can switch therapy when it fails or there is some side effect. It was observed that etanercept was the most frequent first drug, with 36.6% of the biologic-naïve patients starting on this therapy. We can also observe that if the analysis would use only the patients first drug, many drugs would be excluded because of the small number of patients. This complicates or even makes the analysis impossible given the other variables missing values. Namely abatacept, anakinra and rituximab were not frequently used has the first drug.

To create a dataset with one drug per patient the observations from subjects with no baseline for the first biologic were removed and then the follow-up time limited to 4 years. After this, the procedure

\(^1\) A counter for the number of biologics until the present one
consisted of picking the subjects from the less frequent drug to the more frequent drug without repetition of subjects between drugs.

This dataset, cohort dataset, was used in all comparison between the drugs and consists of 394 subjects distributed by the different drugs accordingly with Table 3.2. It was named cohort dataset because for each subject it was select only one of the therapies experienced by him and then they were divided into groups/cohorts that share a particular drug exposure.

Table 3.2 shows the percentage and number of patients per drug on cohort dataset and we can see that with this procedure the number of patients gets more evenly distributed across the different therapies and makes possible the inclusion of the least frequent first drugs in the study. The number of patients increases in comparison with Table 3.1 because a patient can be included in the analysis even if he does not have baseline for the first drug.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Percentage (Number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abatacept</td>
<td>6.35% (25)</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>16.50% (65)</td>
</tr>
<tr>
<td>Anakinra</td>
<td>2.54% (10)</td>
</tr>
<tr>
<td>Etanercept</td>
<td>23.86% (94)</td>
</tr>
<tr>
<td>Golimumab</td>
<td>8.88% (35)</td>
</tr>
<tr>
<td>Infliximab</td>
<td>14.72% (58)</td>
</tr>
<tr>
<td>Rituximab</td>
<td>14.21% (56)</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>12.94% (51)</td>
</tr>
</tbody>
</table>

Table 3.2: Percentage (number) of patients per drug in cohort dataset.

From this cohort dataset we extracted the information needed for the analysis of time to remission (dataset A and dataset B), time to discontinuation and time to response. On these datasets the follow-up time starts with the beginning of the subject biological therapy and ends when one of these events happens or the subject is lost to follow-up. The current biologic number, a counter for the number of biologic the patient experienced, was included as a covariate to control the effect of previous therapy failures.

The criterion for remission was the index-based ACR/EULAR criterion for clinical trials, Table 2.2 (See page 10). The definitions of event for the time to response data was having at least moderate response or at least good response, according to the EULAR response criteria, Table 2.3 (see page 11).

Dataset B characteristics are in all aspects similar to dataset A but the follow-up period was reduced to just one year. In addition, dataset C was created and consists of time to remission data extracted from the original data. In this dataset, each drug of each patient there is an observation for the time to the first remission.

### 3.2 Methods

The main problems with longitudinal data are: the sample population might differ between the beginning of the follow-up and the end of study, exists within-subject correlation of the observations and the time varying variables might influence the effect of an exposure in time. On this chapter we list some methods appropriate to deal with this problems given the data collected and review part of them.

The bulk of the theory on this section came from the books Fundamentals of Biostatistics [28] and Survival Analysis a Practical Approach [27] and R packages documentation and vignettes. The analysis of recurrent events was based on the theory presented on the book Modeling Survival Data - Extending the Cox Model [29] and several articles [30, 31, 32, 33].
3.2.1 Smoothing

When scatterplots interpretation is troublesome, fitting the data can help visualization and interpretation. Locally weighted scatterplot smoothing (LOWESS or LOESS) are methods that perform linear regressions on local data subsets and then combine them to estimate overall smoothed trend of the dependent variable. Each of this local subsets of data contains neighboring data points that are weighted according to the distance to the "center" of that regression.

Kernel density estimation, is non-parametric way of estimating the probability density function of a random variable sample and can be considered as a smoothed version of histograms. The method consists of estimating a function for each histogram bin, sum all this functions and normalize the overall result to obtain estimated probability density function.

On this work, both the kernel density estimation and LOESS were made using the methods of the `ggplot2` package. The kernels were Gaussian and the bandwidth and other parameters were found using the default methods used in these implementations.

3.2.2 Summary measures

One of the simplest and most effective ways of analyzing longitudinal data is to combine the multiple observations of each subject over time into a summary measure or several summary measures. The data from each subject is this way reduced to a single observation and this can be used to compare the response between groups. Typical summaries are the mean, the slope of a regression line, area under the curve, maximum/minimum value and time to the maximum/minimum response.

For example, in analgesics drugs trials the total pain relief (TOTPAR) and the Sum of Pain Intensity Difference (SPID) are used to summarize the treatment response over time. However, the area under the curve, the minimum value and time to minimum value are no longer comparable if the subjects have different follow-up times, like it happens with our data.

Inside these summaries measures we can also include time-to-event data but an alternative is to use more traditional methods and compare cross-sectional statistics of different time points.

The limitation of this methods is that not all available information per subject is being used. However, that also does not mean that methods that use more information will provide better results.

3.2.3 Measures of occurrence and association

Evaluations of the disease progression and therapy response are made at 3 months and 6 months after the start of therapy, as stated by the EULAR/ACR guidelines for biological drugs and by the SPR as well. A way of comparing the efficacy of each drug is by estimating the prevalence of remission for each therapy at this time points. Prevalence is no more than the proportion of a population with some condition at a particular time point or period. Prevalence is a common measure of occurrence used in cross-sectional studies and can be used to describe prevalence’s odds ratio, absolute risk and relative risks.
Different types of studies have different types of measures of occurrence and association. For example in the case of cohort studies, a type of observational study, the proportion of events or risk is the measures of occurrence used to calculate the relative risk, the ratio of risks between the exposed groups. The problem with risks in our case is that it gives to all persons the same influence when the amount of time at risk of the event to happen during the study period varies from person to person. Incidence rate (or incidence density rate) can be thought of as a risk measure that incorporates differences in subject follow-up times into the comparison. In general, incidence rates (IR) are the number of new cases per number of person per time unit over the follow-up time.

\[
\text{IR} = \frac{\text{Number of events}}{\text{Person-time at risk}}^2
\]  

(3.1)

Equation 3.1 gives the estimator for IR. This equation assumes that the event rate is constant across the entire follow-up time. For this reason, IR may be less meaningful if calculated for long time periods. When a incidence rate varies over time is better to use hazard rates, which are kind of a "instantaneous" incidence rate. The hazard is the instantaneous probability (risk) of having and event among those who are at risk of having the event at a given time in the follow-up time period, so, it can be interpreted as the number of events per time unit.

The confidence interval (CI) is a range of plausible values for a parameter of interested, that depends of a chosen significance level. The significance level determines how often the interval contains the true value of the parameter of interest. In other words, for one hundred random samples of the same size, the 100%(1-\(\alpha\)) CI created will contain the true value of the parameter of interest 100(1-\(\alpha\)) times.

Confidence limits for incidence rates are obtained for the expected number of events based on the Poisson distribution and then divided by the total person-time at risk. For 10 events or more the it was considered that the normal approximation to the Poisson distribution did hold (equation 3.2). For less than 10 events were estimated the exact Poisson confidence limits that explore the relation between the Chi-squared distribution and the Poisson distribution (Equation 3.3) [28, 34].

\[
100\%(1 - \alpha) \text{ CI} = \text{IR} \pm Z_{1-\alpha/2} \frac{\text{IR}}{\sqrt{N}}
\]  

(3.2)

\[
100\%(1 - \alpha) \text{ CI} = \left( \frac{\chi^2_{2N,\alpha/2}}{2T}, \frac{\chi^2_{2(N+1),1-\alpha/2}}{2T} \right)
\]  

(3.3)

By assuming approximate normality for log of the incidence rate ratio (IRR) we can obtain the interval estimates as follows in Equation 3.4. The IRR confidence interval gives a range of possible values for the true incidence ratio for the population being compared by the two samples.

\[
\text{IRR 100\%(1 - \alpha) CI} = \exp \left\{ \ln \text{IRR} \pm Z_{1-\alpha/2} \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \right\}
\]  

(3.4)

Means comparisons of continuous variables that might not follow the Gaussian distribution can be done by using the non-parametric Kruskal-Wallis rank sum test. On this kind of methods, the values are

\[\text{Sum of the amount of at risk time each person contributes.}\]
ordered, a rank is assigned for each value and the mean ranks are compared. Because the ranks are compared and not the values, rank sum methods are robust against outliers and unequal variance between groups (heteroscedasticity). To compare proportions it was used the Pearson's Chi-squared test.

### 3.2.4 Kaplan-Meier estimator

The Kaplan-Meier (KM) method is non-parametric statistic of time-to-event data whose estimates can be interpreted as the cumulative proportion of population at risk of experience the event or as probability of staying event free in function of time. This method has special importance in epidemiology but is also used many different areas, like failure analysis in engineering or econometrics. The KM is estimated accordingly with equation 3.5, where \( N(t) \) is the number of persons at risk at time \( t \) and \( E(t) \) is the number of events at time \( t \).

\[
\hat{S}(t) = \prod_{t} \left( \frac{N(t) - E(t)}{N(t)} \right) = \left( \frac{N(t) - E(t)}{N(t)} \right) \hat{S}(t - 1)
\] (3.5)

To measure association between KM estimates the most used process used is the non-parametric method called log-rank test. Log-rank test compares the distance between two or more KM curves to find a p-value for the null hypothesis that all survival curves are equal. The discrepancies between the observed and expected event count are aggregated across all event times and standardized to form the log-rank statistic \( \chi^2_{\text{logrank}} \) that follows approximately a chi-squared distribution with number of KM curves minus one degrees of freedom. The log-rank test tends to give more weight to discrepancies that occur earlier compared to differences later, because there is more data around at that time. Other tests give different weightings.

### 3.2.5 Cox Proportional Hazards Model

Cox Proportional Hazards (Cox PH) regression is semi-parametric linear model used in the analysis of time-to-event data used for estimation, adjustment and prediction.

The proportional hazards (PH) assumption is better understood by examining the Cox regression model in equation 3.6. On the left hand side of equation 3.6, we have the logarithm of the instantaneous risk of having the event among those who are at risk of having the event at a given time in the follow-up period (or log-hazard). On the right hand side, the baseline log-hazard function, \( \ln \hat{\lambda}_0[t] \), is estimated non-parametrically and can be seen as a time dependent intercept. This function determines the shape of the log-hazard in time.

The covariates of a model can be continuous or categorical and the estimated coefficients can be interpreted as the difference in log-hazard per unit difference in \( x \) at a given time \( t \) in the follow-up period, given that all other \( x \)'s remain constant. By other words, the exponential of the coefficients is the hazard ratio between two groups that differ by a single unit of \( x \). The same applies for categorical variables because they are included in the model as dummy variables, one for each group with exception of the
reference group that normally is not included. For example, the exponentiated coefficient for abatacept is the hazard ratio between it and etanercept, the reference drug.

\[
\ln \hat{\lambda}(t|x_1, x_2, ..., x_p) = \ln \hat{\lambda}_0(t) + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_p x_p
\]  

(3.6)

Thus, the shape of the log-hazard function is similar over time and the effect of coefficients on the baseline log-hazard function is simply a vertical translation. Hence the different groups log-hazard functions will be parallel with a distance determined by the predictors and coefficients values. Since the difference of log-hazards is equivalent to the logarithm of the ratio of the hazards, by exponentiation the hazards ratio between two groups is constant over time.

Cox’s regression maximizes the partial likelihood function to estimate coefficients and their standard errors. This means that the estimated coefficients makes the observed data most likely among all choices for the coefficients. Standard errors are need for hypothesis testing \((H_0 = \beta = 0)\) and to construct CI for the coefficients. The partial likelihood function can be maximized using, for example, the Newton-Raphson algorithm.

The estimated coefficients are asymptotically normality distributed so by computing the \(Z = \hat{\beta}_i / SE(\hat{\beta}_i)\) test statistic we can calculate the p-value for the hypothesis \(H_0 : \beta_1 = 0\) vs. \(H_1 : \beta_1 \neq 0\) and construct the \(100%(1 - \alpha)\) CI for HR, as follows in Equation 3.7.

\[
HR \ 100%(1 - \alpha) \ CI = \exp \left\{ \hat{\beta}_i \pm Z_{1-\alpha/2}SE(\hat{\beta}_i) \right\}
\]  

(3.7)

Normally one experiments different combinations of predicting variables and compares the different models according to some criteria to find the best adjustment. The different models are also used to compare the magnitude of association across the different levels of adjustment. When the goal is to maximize precision of adjusted estimates one should keep in final model only those predictors that are statistically significant.

To evaluate each adjustment the pseudo coefficient of determination (pseudo R-squareds), Akaike information criterion (AIC) and Bayesian information criterion (BIC) were calculated. The R-squared value reflects the improvement of the model with the covariates over a model only with intercept by comparing both likelihoods. The statistical software used calculates the the Cox and Snell pseudo R-squared. The maximum value of the pseudo R-squared is not 1.

The AIC or BIC for a model are also likelihood functions but penalize depending on the number of parameters. In the case of the BIC the penalty depends also of the number of observations. A lower AIC and BIC mean the model is more likely to be the true model. The value AIC can still be corrected for finite sample sizes (AICC) by adding a term that penalizes for extra parameters and depends of the number of observations. Let \(L\) denote likelihood, \(p\) the number of parameters and \(n\) the number of observations, the metrics for model comparison have the form

\[
AIC = -2 \ln(L) + 2p.
\]  

(3.8)
\[ \text{AICc} = \text{AIC} + \frac{2p(p + 1)}{(n - p - 1)}, \]  
\[ \text{BIC} = -2 \ln(L) + p \ln n. \] 
(3.9)  
(3.10)

Non-proportionality occurs when the relationship between the outcome and the continuous predictors is not linear or the HR is not constant over time causing the shape of the hazards curves to be very different or to cross between groups.

A way to assess PH is to plot the complementary log transformation, or \( \log(-\log S(t)) \), against \( \log(t) \). If the hazard rate does not change with time then plot will seem approximately linear. We did not present this results because there was many strata and extract any information about this plots was difficult.

Another way to assess PH is to visualize or fit the scaled Schoenfeld residual [35] for each individual and covariate. Schoenfeld residual "are based on the individual contributions to the derivative of the log partial likelihoo" for a specific covariate [36].

Therefore, a significant regression slope coefficient for the fit of the residuals indicates that the true HR changes over time and the PH assumption does not hold. This method is known as the Grambsch-Therneau test of trend in Schoenfeld residuals and is available using the \texttt{cox.zph} function of the survival package [35]. The \texttt{Cox.zph} function also allows for a global chi-square test for the model and provides a correlation coefficient (r) between transformed survival time and the scaled Schoenfeld residuals to assess if the HR are increasing or decreasing over time.

In case of non-proportionality for a continuous predictor, we can categorize the concerned predictors. Some authors state that when the PH assumption is not met we can interpret the Cox PH results as the average HR over the follow-up time [27, 35].

Cox PH model has different extension to allow for time-dependent variables, stratified analysis and clustered data. Time-dependent variables can be introduced by adding an interaction term between one of the covariates and a function of time, normally \( \ln t \) or by having multiple observations per subject with corresponding time at risk for each value. A stratified Cox model fits a separate baseline hazard functions for each strata. For this reason the the partial likelihood functions for each strata depends only on the observations for each strata. After the maximization of the partial likelihood the effect of the strata is removed without making the assumption of proportional hazards. This is particularly useful if the stratifying variable does not follow this assumption but it is not useful if the effects of this variable are of scientific interest, like the case of the biological therapies.

Statistical independence of the observations is a important assumption that needs to be present for most statistical methods. The independence assumption is violated when some observations are more similar to each other than from other observations. By adding a cluster term to the Cox regression model the correlated observations are identified as belonging to a particular cluster. Each cluster forms a stratum and the model considers that the observations are conditional independent within the clusters. The coefficients are estimated assuming independence within cluster and the variance is estimated using robust sandwich variance estimators. This method corrects the variance estimation assuming correlation between different observations but does not correct the estimate of the coefficients.
Recurrent events models

Simplified statistical models that consider only the time to the first event on recurrent event data, might be discarding important information that can be present on the data. Therefore, there are a variety of methods that extend the Cox model formulation to include multiple events per subjects by using proper data layouts and by correcting the variance of the estimates.

<table>
<thead>
<tr>
<th>ID</th>
<th>Interval</th>
<th>Event</th>
<th>Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>1 (0,90]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AG</td>
<td>1 (110,180]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AG</td>
<td>1 (180,260]</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AG</td>
<td>2 (0,130]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AG</td>
<td>2 (200,365]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PWP-CP</td>
<td>1 (0,90]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PWP-CP</td>
<td>1 (110,180]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PWP-CP</td>
<td>1 (180,260]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>PWP-CP</td>
<td>2 (0,130]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PWP-CP</td>
<td>2 (200,365]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PWP-GP</td>
<td>1 (0,90]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PWP-GP</td>
<td>1 (0,70]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PWP-GP</td>
<td>1 (0,80]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>PWP-GP</td>
<td>2 (0,130]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PWP-GP</td>
<td>2 (0,165]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>WLW</td>
<td>1 (0,90]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WLW</td>
<td>1 (0,180]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>WLW</td>
<td>1 (0,260]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>WLW</td>
<td>2 (0,130]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WLW</td>
<td>2 (0,365]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>WLW</td>
<td>2 (0,365]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>TT-R</td>
<td>1 (0,90]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TT-R</td>
<td>1 (0.180]</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TT-R</td>
<td>1 (0.260]</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>TT-R</td>
<td>2 (0,130]</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TT-R</td>
<td>2 (0.365]</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.3: Data layouts for different recurrent event models using two hypothetical subjects. Andersen and Gill (AG) counting process independent increment model. Prentice, Williams and Peterson (PWP) conditional model with counting process (CP) and gap time (GT) time scales. Wei, Lin and Weissfeld (WLW) marginal model. Total time-restricted (TT-R). Adapted from [37].

When a event occurs multiple times for the same subject, it is known as a recurrent event. Many medical outcomes are recurrent, examples include seizures and migraines and heart attacks. Many also do not have continuous risk intervals, like on the case of recurrent malaria episodes, were exists persistent treatment effect during a period on which a new infection is not possible [32]. Discontinuous time at risk also happens recurrent lung exacerbations in cystic fibrosis patients, were the patient is not at risk of new exacerbation while it is experiencing one [29, 38].

This kind of analysis was also found for the study of sport injuries [31] and functional disability among older persons [33].

There are two types of recurrent events: ordered and unordered. In this work we will only use models for ordered events. The distinction between the models lies on the risk set and on the choice for the baseline hazard function: common or event-specific. A model with a common baseline hazard has the same underlying hazard for all events. An event-specific baseline hazard is a stratified baseline hazard that allows the baseline hazard to be different for each event. For recurrent events the usual
approaches to generalize the Cox’s framework model are on Table 3.3.

Andersen and Gill (AG) [39] model breaks the follow-up time into segments defined by the events and has a common baseline hazard for all events. This model treat the events as being independent because does not differentiate between the first and subsequent events.

Prentice, Williams and Peterson (PWP) [40] model is called conditional because a subject is assumed not to be at risk for the next events until the prior events have occurred. The stratum variable keeps track of the event number. For this model is possible to have two different time scales: the counting process (CP) and gap time (GT) time scales. Gap time scale corresponds to the time since entry or last event.

Wei, Lin, Weissfeld (WLW) [41] model each event is considered as separated process. All subjects are considered at risk for all events, regardless of how many events they actually experience. For this reason the number of observations per subject depends of the maximum of events a subject has experienced in the dataset.

Total Time-Restricted (TT-R) [42] is a model similar to the WLW. This model uses the same intervals as WLW, but deletes the strata for which no additional follow up time is added.

On this models the coefficients are estimated ignoring the existing correlation between the events and the correction for the dependency is done by estimating a corrected variance, similar to the sandwich estimator. As well as in single-event Cox regression, censoring must be independent of the events. In our case time at risk is discontinuous because makes no sense that while on remission periods the subjects are at risk of being in remission.

### 3.3 Data Analysis

A research project normally consists of five phases: the planning/design of the study, data collection, data analysis, presentation, interpretation and discussion of the results. This thesis is concerned with analyzing the data, interpret it and discussion of the results. The role of data analysis or statistics is to learn the characteristics of a population via samples collected and analyzed so that true characteristics of the data are observable.

During this work, the central tendency of continuous data was expressed as means. When is expected that the variables not follow a Gaussian distribution it was used the median. The dispersion tendency was measured by the standard deviation or by the first and third quartile.

Categorical data is expressed by absolute and relative frequencies distributions and Kaplan–Meier survival analysis was used to assess the proportion of subjects that remain event/outcome free.

All comparisons of groups were done using the Kruskal-Wallis rank sum test, a non-parametric test that allows for variables that do not follow a Gaussian distribution. Comparison of proportions was made with Pearson’s Chi-squared test.

Cox multivariate regression models were used to assess potential predictors of remission, adjust for confounding variables and estimate hazards ratios.

There was no data imputation and when comparing models it was always used the dataset defined by the complete cases of covariates of the larger model.
Categorical variables where recategorized when necessary because of the small number counts and continuous variables were categorized when necessary to stratify the analysis or when the proportional hazard assumption was not met.

The significance level ($\alpha$) used in this work was 0.05 (5%) calculated with two-sided tests. This means that for samples of the same size drawn from the same population the chance of finding estimates outside the CI is 5 in 100. When relevant the 95% CI were calculated.

All statistical analysis in this work was conducted with the use of the statistical software R [43]. The dplyr package helped with the data processing and most of the figures and tables were made using ggplot2 and xtable package, respectively [44, 45, 46]. The survival analysis was performed using R’s survival package [47, 29].

3.3.1 Descriptive analysis

The first part of descriptive analysis used the original data set. It was given focus to rare occurrences that would be difficult to study or include in a model. Textual description of those and other observations was made. The different combinations of DMARDs, anatomical distribution of the affected joints at baseline were also some of the results extracted from the original data set.

Cohort dataset was the source for Table 4.1, used to compare across the multiple cohorts baseline values. Mean smooth estimates and kernel density estimates for the evolution of the disease activity were also produced for each drug as well as boxplots for the duration of each therapy.

3.3.2 Summary measures analysis

On the summary analysis, each one of the subjects DAS28’s evolution over time was fitted with a simple linear model and then the mean and 95% CI of the slope and p-values was presented. It were only used observations from the first 6 months to take advantage of the approximate linearity of the period after the start of the drug.

3.3.3 Incidence density rates

IR and IRR with 95% CI were calculated for dataset A, time to drug discontinuation and time to response data. Then, the data was stratified by drug and other factors to investigate for possible interactions with the therapies.

3.3.4 Kaplan-Meier estimates

Kaplan-Meier estimates for the therapies were presented using dataset A along with the result from the log-rank test for this estimates. The same was made for the comparison of biologic agents persistence. In some articles this kind if analysis is also called “drug survival” or medication persistence. Drug persistence is defined as “the duration of time from initiation to discontinuation of therapy” [48] and is an overall marker of the success of a treatment and depends on the efficacy of the drug, its side effects.
and doctor and patient preferences. Remission was defined using the ACR/EULAR remission criteria for clinical trials.

Nevertheless, because there are many different therapies and other factors of interest that should be investigated, a regression method is more appropriate than KM estimates.

### 3.3.5 Cox proportional hazards regression

In this work Cox proportional hazards regressions were used to assess potential predictors of remission and adjust for confounding variables. Another advantage of having one drug per subject (cohort dataset) is that the KM estimates and the Cox regressions come from the same data.

With the regression models were investigated the following variables: biological therapy, sex, age, weight at baseline, height, body mass index, alcohol, smoking, race, years of education, age of onset, DAS28 at baseline, SDAI at baseline, time since onset, season, DMARD therapy, GCs therapy, extra articular manifestations, patient VAS evaluation, physician VAS evaluation, pain VAS evaluation, HAQ, erythrocytes sedimentation rate, reactive C-protein, rheumatoid factor, anti-CCP, number of swollen joints, number of tender joints, erosive and biologic number.

The rest of the variables were excluded because of the number of missing values, were redundant, monotonous in some of the therapies or to keep the number of variables in the models relatively small comparing to the number of observations. For example, the indicator for each of the HAQ questions or for each one the joints was not used, just the resulting values for the score and count.

The estimated hazard ratios and 95% confidence intervals, provide adjustment for potential systematic differences between our therapy groups (confounding factors) and identify risk factors for decrease of therapy efficacy.

**Model selection**

First, the unadjusted HR of each candidate covariates was estimated with on univariate Cox model. Second, the non-statistically significant covariates for a level of 10% were discarded and the remaining were included on a multivariate Cox model. Third, it was used backward elimination, eliminating always the least significant variable. Finally, the models were compared according to the AIC, BIC and \( R^2 \) scores. All the models, univariate and multivariate, were also tested for proportional hazards using Grambsch-Therneau test of trend in Schoenfeld residuals.

The second method for model selection consisted of including 25 of the most relevant variables in the model and subsequently remove the one of the variables according with four different criteria. It was only used 25 variables because of the missing values. The first criterion was the same as previous one: remove the least significant variable. The three other criteria consisted of removing the the variable that will produce the best model according with a measures of quality of the statistical models used for model selection (AIC, AICc and BIC). The therapy variable was forcibly kept in the models even if removing it would produce a better model.

These 25 variables were: therapy, sex, descent, disease duration, weight, height, season, age RA
onset, erosive indicator, ESR, CRP, RF, age, SJC, TJC, extra articular manifestations, patient VAS, pain VAS, doctor VAS, DAS28, SDAI, HAQ, years of education, DMARD(s) indicator and biologic number.

**Recurrent event analysis**

The recurrent event analysis was performed using the Andersen and Gill (AG) counting process independent increment model, the Prentice, Williams and Peterson (PWP) conditional model with counting process (CP) and gap time (GT) time scales, the Wei, Lin and Weissfeld (WLW) marginal model and the total time-restricted (TT-R). The covariates used were the ones from the first model of the second model selection method for the first event Cox regression (Method 2) and the comparison of models was made with the same metrics.
Chapter 4

Results

4.1 Descriptive analysis

The objective of this section is to summarize and generally describe the sample data in a concise and interpretable way. Mostly it pretends to create intuition regarding the data and as aid for the next sections of the data analysis.

The amount of missing values in some variables was a problem throughout the work. A group of important variables for this work were selected and ordered according with the percentage of missing values. The results were: HAQ (44.83%), ΔDAS28 (41.81%), EULAR response (41.81%), SDAI (36.28%), CDAI (30.73%), physician HAQ-VAS (29.79%), BMI (25.94%), years of education (21.3%), Weight (20.73%), erosive (18.4%), DAS28 (18.74%), PCR (13.36%) and SR (11.65%). Figures 3.2a and 3.2b (see page 16) also characterize the structure of missingness in the data.

For example, the original dataset had 436 patients and for RF, anti-CCP and sex variables there was 343 subjects with complete cases. From this patients, 70.3% were RF positive, 69.7% were anti-CCP positive and 85.4% were women. From 366 subjects with reported race, 87.7% were white of European origin, 6% were black, 3.6% mulatto, 1.6% of non-European origin and the rest of Asian origin.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abatacept (n=23)</th>
<th>Adalimumab (n=58)</th>
<th>Anakinra (n=9)</th>
<th>Etanercept (n=67)</th>
<th>Golimumab (n=18)</th>
<th>Infliximab (n=54)</th>
<th>Rituximab (n=47)</th>
<th>Tocilizumab (n=42)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex (F), n (%)</strong></td>
<td>19 (82.61)</td>
<td>51 (87.93)</td>
<td>8 (88.89)</td>
<td>55 (82.09)</td>
<td>16 (88.89)</td>
<td>47 (87.04)</td>
<td>42 (89.36)</td>
<td>33 (78.57)</td>
<td>0.8322</td>
</tr>
<tr>
<td><strong>Age (years), mean ± SD</strong></td>
<td>63.61 ± 14.33</td>
<td>56.83 ± 12.00</td>
<td>63.81 ± 16.16</td>
<td>59.27 ± 13.59</td>
<td>61.59 ± 11.74</td>
<td>60.80 ± 15.25</td>
<td>60.39 ± 12.51</td>
<td>57.74 ± 12.90</td>
<td>0.2277</td>
</tr>
<tr>
<td>Age at onset (years), median [Q1-Q3]</td>
<td>48.07 [36.45-55.18]</td>
<td>39.56 [30.96-50.87]</td>
<td>47.59 [38.78-59.81]</td>
<td>44.99 [32.76-54.52]</td>
<td>43.97 [35.41-54.05]</td>
<td>42.89 [32.63-51.70]</td>
<td>42.55 [35.17-51.12]</td>
<td>47.82 [34.27-54.53]</td>
<td>0.7171</td>
</tr>
<tr>
<td><strong>Weight (kg), mean ± SD</strong></td>
<td>69.46 ± 16.35</td>
<td>72.24 ± 13.92 of 50</td>
<td>76.50 ± 13.23 of 4</td>
<td>71.28 ± 14.43 of 44</td>
<td>72.35 ± 18.48 of 13</td>
<td>71.17 ± 12.42 of 50</td>
<td>70.45 ± 12.80 of 41</td>
<td>75.08 ± 14.78 of 38</td>
<td>0.6265</td>
</tr>
<tr>
<td><strong>Height (m), mean ± SD</strong></td>
<td>1.91 ± 0.08</td>
<td>1.87 ± 0.07 of 50</td>
<td>1.89 ± 0.08 of 44</td>
<td>1.90 ± 0.08 of 44</td>
<td>1.91 ± 0.08 of 41</td>
<td>1.90 ± 0.08 of 40</td>
<td>1.90 ± 0.08 of 46</td>
<td>1.90 ± 0.08 of 46</td>
<td>0.0537</td>
</tr>
<tr>
<td><strong>Rheumatoid factor (yes), n (%)</strong></td>
<td>0.22 (13.04)</td>
<td>0.11 (18.97)</td>
<td>0.11 (11.11)</td>
<td>0.11 (11.11)</td>
<td>0.11 (11.11)</td>
<td>0.11 (11.11)</td>
<td>0.11 (11.11)</td>
<td>0.11 (11.11)</td>
<td>0.0710</td>
</tr>
<tr>
<td><strong>Extra articular manifestations (yes), n (%)</strong></td>
<td>0.03 (14.68)</td>
<td>0.00 (9.47)</td>
<td>0.00 (11.97)</td>
<td>0.00 (11.97)</td>
<td>0.00 (11.97)</td>
<td>0.00 (11.97)</td>
<td>0.00 (11.97)</td>
<td>0.00 (11.97)</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Baseline DAS28 Mean ± SD</strong></td>
<td>5.29 ± 1.63 of 21</td>
<td>5.00 ± 1.42 of 50</td>
<td>6.09 ± 1.26 of 4</td>
<td>5.25 ± 1.48 of 58</td>
<td>5.10 ± 1.74 of 14</td>
<td>5.36 ± 1.32 of 45</td>
<td>5.90 ± 1.31 of 43</td>
<td>5.85 ± 1.28 of 40</td>
<td>0.0458</td>
</tr>
<tr>
<td><strong>Mean ± SD</strong></td>
<td>1.64 ± 0.67 of 39</td>
<td>1.62 ± 0.57 of 39</td>
<td>1.35 ± 0.69 of 49</td>
<td>1.39 ± 0.78 of 12</td>
<td>1.42 ± 0.64 of 37</td>
<td>1.57 ± 0.72 of 25</td>
<td>1.61 ± 0.54 of 36</td>
<td>1.145</td>
<td></td>
</tr>
<tr>
<td><strong>Median [Q1-Q3]</strong></td>
<td>1.50 [1.44-1.94]</td>
<td>1.38 [0.62-1.62]</td>
<td>1.38 [0.75-1.75]</td>
<td>1.25 [1.03-1.75]</td>
<td>1.25 [1.00-1.88]</td>
<td>1.25 [1.03-1.75]</td>
<td>1.25 [1.03-1.75]</td>
<td>1.62 [1.38-2.00]</td>
<td>–</td>
</tr>
<tr>
<td><strong>No. of switches, n (%)</strong></td>
<td>1 (4.35)</td>
<td>2 (22.22)</td>
<td>60 (89.55)</td>
<td>7 (36.89)</td>
<td>33 (61.11)</td>
<td>3 (6.38)</td>
<td>18 (42.86)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td><strong>Mean biologic No. ± SD</strong></td>
<td>1.37 ± 0.94</td>
<td>1.36 ± 0.61</td>
<td>2.00 ± 1.58</td>
<td>1.09 ± 0.29</td>
<td>1.44 ± 0.70</td>
<td>1.09 ± 0.29</td>
<td>2.62 ± 0.85</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td><strong>Concomitant DMARDS (yes), n (%)</strong></td>
<td>22 (95.65)</td>
<td>53 (91.38)</td>
<td>8 (88.89)</td>
<td>61 (91.04)</td>
<td>17 (94.44)</td>
<td>51 (94.44)</td>
<td>43 (91.49)</td>
<td>34 (80.95)</td>
<td>0.4485</td>
</tr>
<tr>
<td><strong>Mean ± SD</strong></td>
<td>5.29 ± 1.63 of 21</td>
<td>5.00 ± 1.42 of 50</td>
<td>6.09 ± 1.26 of 4</td>
<td>5.25 ± 1.48 of 58</td>
<td>5.10 ± 1.74 of 14</td>
<td>5.36 ± 1.32 of 45</td>
<td>5.90 ± 1.31 of 43</td>
<td>5.85 ± 1.28 of 40</td>
<td>0.0458</td>
</tr>
<tr>
<td><strong>Mean ± SD</strong></td>
<td>1.64 ± 0.67 of 39</td>
<td>1.62 ± 0.57 of 39</td>
<td>1.35 ± 0.69 of 49</td>
<td>1.39 ± 0.78 of 12</td>
<td>1.42 ± 0.64 of 37</td>
<td>1.57 ± 0.72 of 25</td>
<td>1.61 ± 0.54 of 36</td>
<td>1.145</td>
<td></td>
</tr>
<tr>
<td><strong>Median [Q1-Q3]</strong></td>
<td>1.50 [1.44-1.94]</td>
<td>1.38 [0.62-1.62]</td>
<td>1.38 [0.75-1.75]</td>
<td>1.25 [1.03-1.75]</td>
<td>1.25 [1.00-1.88]</td>
<td>1.25 [1.03-1.75]</td>
<td>1.25 [1.03-1.75]</td>
<td>1.62 [1.38-2.00]</td>
<td>–</td>
</tr>
<tr>
<td><strong>Concomitant therapies</strong></td>
<td>16 (89.77)</td>
<td>41 (70.59)</td>
<td>57 (85.07)</td>
<td>15 (83.33)</td>
<td>44 (81.48)</td>
<td>42 (89.36)</td>
<td>31 (73.81)</td>
<td>0.1041</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Baseline table.
From the total of patients, it was found that only 3 patients have psoriasis and other 3 had uveitis. One of the uveitis patients also presented serositis while on etanercept and a different subject presented ulcerative colitis while on adalimumab. Another patient had an adverse effect, an abscess, with adalimumab. Two patients presented lymphadenopathy one on etanercept and another with infliximab or rituximab. Unfortunately it was not possible to know with which drugs because the indicator is for all the subject observations. Non-specific colitis was present in one patient during etanercept. All the above patients presented RF.

It was observed a total of different 18 combinations of DMARDs. The most frequent DMARD was MTX, present in 61.68% of the observations (excluding when in combination with other DMARDs). In 13.97% of the observations the patient were without DMARD and 1% of the observations had three DMARDs at the same time. This happened also when the biologic was active. The most common combination of three simultaneous DMARDs reported was MTX + SSZ + HCQ (0.96%). The most common combination of two DMARDs was MTX plus HCQ (6.2%).

It was observed a total of 12 different combinations of GCs, while about 31% of the observations had no active GC. The most common GCs were the prednisolone (30.50%) and prednisone (30.15%). The most common combination of two GCs was deflazacort with prednisolone (0.17%). It was also reported a combination of three simultaneous CGs: deflazacort, prednisolone and betamethasone (0.01%).

The anatomical distribution of the most affect joints in Figures 4.1, 4.2a and 4.2b was made with a total of 348 subjects at the baseline of the first biologic. The light red circles correspond to the swollen joints and the red circle circles to the tender joints.

The figures shows that the joints were more likely to be swollen (or inflamed) than tender (or painful) even if the proportions were sometimes very similar. In proportion, the wrists joints were the most affected with 59.4% and 51.7% of the subjects with swollen joints and 55.9% and 46.4% with tender joints, right and left wrist respectively.

Globally, the hands were the most affected area in the body. The most affected joints in the hand were the second and third metacarpophalangeal. As expected, the results are practically symmetric with
a slight increased prevalence on the right side. The joints more times reported as not assessable were the right and left knees joints, 1.59% and 1.3%.

Figures 4.6, 4.3, 4.5, Table 4.1 and 4.2 came from the cohort dataset (described in the Section 3.1, see page 15).

The baseline table, Table 4.1, comes from a total of 318 subjects. Because the number of complete cases for each drug group was very low and there were many variables, this table tried to find a compromise between these two aspects. The variables were ordered according to the proportion of missing values and eliminated progressively. At the same time the number of complete cases for that sub-dataset was recorded. Then plots like the one in Figure 3.2a (page 16) were used to select the variables to remove. From this set were selected some variables were kept because their importance in the analysis, like DAS28 or HAQ. From the kept variables, some were still removed because they were not of interest or redundant, like in the case of the number of days after the first biologic.

The Kruskal-Wallis rank sum test indicates a significant difference between our therapy groups in relation to the disease duration and also for the current biologic number, according to the guidelines. The overall median, [Q1-Q3], of the disease duration was 13.7 [9.4-20.0]. The Pearson’s Chi-squared test for count data suggests a difference on the proportions between our groups for the RF and for no drug switch and more than two switches.

Figure 4.3: DAS28 by drug, over the 4 years follow-up period. Top left corner: Local Regression (LOESS) estimation of the mean DAS28 in time by biological therapy. Top right and bottom left corner: Estimate of the proportion of each drug by DAS28 value at 3 and 6 months. Bottom right corner: Cumulative distribution function for the 4 years follow-up time.
In the LOESS estimate of Figure 4.3, it was observed that most of the DAS28 variation happened during the first year. On a closer look it seems that the bulk of the change occurs between the third and sixth month and on this interval the observed DAS28 estimate implies, on average, a linear trend.

During this 4 years, the figure shows a similar behavior for all the the anti-TNF biological therapies, with the 95% confidence bands overlapping. Recall that the anti-TNF biological therapies are adalimumab, etanercept, infliximab and golimumab. Excluding the groups with anakinra and rituximab, the intercepts of rest of the drugs are not significantly different and were approximately 5.1 DAS28.

The top right and bottom left corner plots can be interpreted as prevalence estimates for the different DAS28 values given the biologic drug and over a period of two months centered at 3 and 6 months, respectively. We can observe that tocilizumab was more frequent for DAS28 values bellow 2.6 at 3 and 6 months of follow-up, while the DAS28 of the other drugs was more frequent for values above 5.1. The cumulative probability function estimate (bottom right corner) is in respect to the overall 4 years period, so it uses all observations from the patients.

An overall interpretation of Figure 4.3 suggests that abatacept, anakinra and rituximab behave worse than the rest of the drugs and suggests that the anti-TNF have similar responses. Tocilizumab, the IL-6 receptor antagonist, presents the best results on all estimates and it is bellow the threshold for remission of DAS28 score after the 6 months of therapy.

An overall interpretation of Figure 4.3 suggests that abatacept, anakinra and rituximab behave worse than the rest of the drugs and suggests that the anti-TNF have similar responses. Tocilizumab, the IL-6 receptor antagonist, presents the best results on all estimates and it is bellow the threshold for remission of DAS28 score after the 6 months of therapy.

![LOESS smoothing estimate](image)

Figure 4.4: Local Regression (LOESS) estimation of the mean HAQ in time by biological therapy.

The (95% CI) HAQ estimate on figure 4.4 measures the evolution of the functional disability of the patients over time by drug. Anakinra was excluded from this plot because it had only 19 observations from 8 patients.

At the beginning of follow-up, the groups on tocilizumab, abatacept and rituximab had a slightly
higher HAQ than the rest of the other the groups. After the beginning, etanercept's, adalimumab's and tocilizumab's responses were similar, with a fast decrease from around 1.25 to 0.75 in 9 months. After this period, it was evident that the tocilizumab's patients continued to improve slightly while the improvement of the patients on etanercept and adalimumab halted or is not perceptible.

Despite having different intercepts, the shape of the response of the groups on abatacept and golimumab seem to be similar. Initially, they improved during 6 months. Then, the values increased during 6 more months and were followed by improvement until stabilization around the third year. Albeit unsteady decrease and the fluctuation of the estimates, the final mean HAQ of this two drugs was at the level of etanercept's, adalimumab's and tocilizumab's. Regarding infliximab and rituximab, they had a slower rate of improvement than the other drugs.

Figure 4.5: Disease activity (DAS28) conditional density estimate for the period of 4 years after the start of the therapy.

Figure 4.5 shows the evolution of the disease activity for each of the drugs. The estimates are a representation of the probability density estimates for each drug during the period of 4 years after the start of the therapy. They show how the prevalence for the different disease activity cut-points behaves over time for each of the drugs and they seem to be in accordance with the ones in Figure 4.3.

As can be seen, the patients on the anakinra did not achieve remission, while the rituximab plot shows slow progression and slight improvement in comparison with the baseline. The disease activity fell faster on golimumab in comparison with the other anti-TNF (adalimuma, etanercept and infliximab) and the patients that experienced a faster and more expressive improvement were the ones on the
tocilizumab group. For this drug the proportion of high disease activity patients plummet to almost 0% in less than 6 months.

Figure 4.6: Observed therapy duration for each therapy drug in years. Box plots show median values (solid horizontal line), the distance between the first and third quartiles or interquantile range (box), lower and upper extremes (the solid vertical line that extends from the box to the highest or lowest value three halfs of the interquantile range), outliers (values farther than the solid vertical line) and the mean therapy duration (diamonds). The dark red dashed horizontal line is the overall median.

Figure 4.6 show the boxplots for the years of therapy by drug. Infliximab was the biologic with longer duration of therapy (Median [Q1-Q3]: 4.35 [1.56-6.65] years) followed by rituximab and adalimumab. Golimumab had the shortest duration (Median [Q1-Q3]: 1.11 [0.58-1.59] years). The overall median duration of therapy was 2 years and 4 months.

4.2 Summary measures analysis

Regarding to Table 4.2, all subjects with more than two observations were used to estimate the summary measures presented. It were only used observations from the first 6 months to take advantage of the approximate linearity of the period after the start of the drug, evident in Figure 4.3. Anakinra was not considered because of the small number of observations.

First, each one of the subjects DAS28’s evolution over time was fitted with a simple linear model and then the mean and 95% CI of the summary measures was found and presented.

It was found a significant difference between the estimated mean values of the initial DAS28 of the different drugs (P = <0.001). The mean difference in DAS28 after a year was larger for tocilizumab (Mean slope, -9.2, 95% CI, -12.00 to 6.40). The smallest decrease in DAS28 after a year occurred for abatacept. After accounting for sample variability the decrease with abatacept was not significant (Mean
The proportion of significant slopes was larger for tocilizumab for a level of significance of 5%. 31.43% of the subjects on tocilizumab had statistically significant slopes. Rituximab was the next therapy with more significant slopes and had less than half of the number of significant differences compared to tocilizumab (14.29%). The estimates for adalimumab’s slopes were only 5.26% of the times statistically significant. There was a statically difference between all the estimates for the percentage of significant slopes ($P=0.049$).

### 4.3 Incidence density rates

The IR and IRR estimates on this section came from the time to remission data (dataset A), time to drug discontinuation data and time to response data, all datasets extracted from the cohort dataset.

For the time to remission data, patients were followed from enrollment on the therapy until remission or censoring. The remission criteria used was the ACR/EULAR remission criterion for clinical trials (SDAI $\leq 3.3$) and the follow-up period was up to 4 years. On the drug persistence data, or time to drug discontinuation, the subjects were followed until censoring or switch to a new therapy. For time to response the patients were followed until good response, at least moderate response or censoring.

Table 4.3 is a summary of the estimates for the IR and 95% CI for the different events and each of the drugs. The IR results are in event per 1000 person-months and they concern a follow-up period of 4 years after the beginning of the drug.

Regarding the drug persistence, a total of 80 therapy switches were observed (IR, 5.24 per 1000 person-months, 95% CI, 4.1 to 6.4); of these, the rate of switches was smaller for subjects on therapy with etanercept (1 switch, IR, 2.3 per 10000 person-months, 95% CI, 0.1 to 12.7) and higher for anakinra (8 switches, IR, 22.41 per 1000 person-months, 95% CI, 9.68 to 44.16).
Table 4.3: Incidence rate by drug per 1000 person-months over the follow-up period of 4 years.

<table>
<thead>
<tr>
<th>Event</th>
<th>Abatacept</th>
<th>Adalimumab</th>
<th>Anakinra</th>
<th>Etanercept</th>
<th>Golimumab</th>
<th>Infliximab</th>
<th>Rituximab</th>
<th>Tocilizumab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug switch</td>
<td>21.00 (7.98-34.02)</td>
<td>7.18 (4.11-10.25)</td>
<td>22.41 (9.68-44.16)</td>
<td>0.23 (0.01-1.27)</td>
<td>12.09 (4.86-24.92)</td>
<td>5.26 (2.68-7.84)</td>
<td>4.75 (1.81-7.70)</td>
<td>5.14 (2.07-10.59)</td>
</tr>
<tr>
<td>Remission</td>
<td>23.26 (9.35-47.92)</td>
<td>15.34 (8.25-22.42)</td>
<td>0.00 (0.00-64.09)</td>
<td>29.98 (21.32-38.65)</td>
<td>21.04 (8.46-43.35)</td>
<td>24.09 (14.65-33.53)</td>
<td>9.24 (4.01-14.46)</td>
<td>60.56 (38.52-82.60)</td>
</tr>
<tr>
<td>Response GM/L</td>
<td>50.00 (27.52-72.48)</td>
<td>19.93 (13.83-26.03)</td>
<td>72.82 (1.84-405.74)</td>
<td>21.17 (16.18-26.17)</td>
<td>48.46 (27.74-69.19)</td>
<td>17.81 (12.29-23.33)</td>
<td>25.27 (17.80-32.74)</td>
<td>32.20 (22.34-42.06)</td>
</tr>
<tr>
<td>Response G/ML</td>
<td>28.95 (11.84-46.05)</td>
<td>13.18 (8.21-18.16)</td>
<td>0.00 (0.00-268.64)</td>
<td>16.57 (12.15-20.99)</td>
<td>34.62 (17.10-52.14)</td>
<td>13.36 (8.58-18.14)</td>
<td>13.21 (7.81-18.61)</td>
<td>29.06 (19.70-38.42)</td>
</tr>
</tbody>
</table>

Figure 4.8: Kaplan-Meier estimates stratified by biologic drug.
Still regarding Table 4.3, a total of 146 remissions were observed (IR, 23.15 per 1000 person-months, 95% CI, 19.4 to 26.9); of these, the best results occurred for subjects on therapy with tocilizumab (30 remissions, IR, 60.56 per 1000 person-months, 95% CI, 38.52 to 82.60) and the worse results for the anakinra cohort, on which were not observed remissions. Additionally, the remission rate for rituximab, the second worst result, was 9.24 and could be as small as 4.01 and as large as 14.46 per 1000 person-months.

The 51 subjects on tocilizumab had a greater chance of remission when compared to the other 91 subjects on the etanercept (reference group). After accounting for sampling variability, tocilizumab therapy could increase the possibility of remission compared with patients enrolled on the reference drug from 36% to 240%, at the population level (IRR, 2.15, 95% CI, 1.36 to 3.40). Because these results are unadjusted there is the possibility that they are confounded.

The 55 subjects on rituximab had a smaller chance of remission when compared to the other 91 subjects on etanercept. After accounting for sampling variability, rituximab therapy had 83% to 40% smaller chances of remission than the patients enrolled on the reference drug, at the population level (IRR, 0.32, 95%CI, 0.17 to 0.6).

Subjects in the infliximab group had 17% smaller chances of remission in the follow-up period when compared to the subjects in the reference group. After accounting for sampling variability between these two anti-TNF drugs, however, there was no evidence that infliximab was better than etanercept at the population level (IRR, 0.83, 95% CI, 0.51 to 1.35, includes 1).

The IR and IRR CIs for adalimumab and etanercept did not agree. On the first case the 95% CI overlap and on the second it does not include 1 (IRR, 0.55, 95% CI, 0.32 to 0.93). Additionally, these results estimate that this increase in remission rate could be 9.24 and as small as 4.01 and as large as 14.46 per 1000 person-months for rituximab, the second worst result.

In respect to the rates of at least moderate response, there was 291 events (IR, 23.88 per 1000 person-months, 95% CI, 21.13 to 26.61). In this case, the two drugs with highest rate of response was anakinra and abatacept. But coincidentally they were also the estimates with less subjects and events. Anakinra only counted with 4 subjects and 3 events and abatacept with 19 subjects and 19 events.

In contrast, for the rate of good responses, there was 215 events (IR, 17.68 per 1000 person-months, 95% CI, 15.31 to 20.4). The two drugs with highest rate of good response were golimumab (IR, 34.62 per 1000 person-months, 95% CI, 17.10 to 52.14) and tocilizumab (IR, 29.06 per 1000 person-months, 95% CI, 19.70 to 38.42). Subjects on golimumab were 1.1 times more likely of responding good to the therapy than the ones on etanercept. After accounting for sample variability, patients on golimumab had more 20% to 262% chances to respond well to therapy (IRR, 2.08, 95% CI, 1.20 to 3.62).

On the next analysis, the data was stratified by drug and other factors to investigate for possible interactions of this factors with the drugs. On Figure 4.9 we present some of the significant findings. When plotting ratios one should have in attention that the scaling of ratios is not symmetric around the value one. This happens because all values of the ratio bellow one have to fit between the 0 and 1 while all the values of ratio greater than 1 can fit between $1, +\infty[$. With the log-transform this values fit in the interval $]−\infty, 0[$ and $]0, +\infty[$, respectively. For this reason the y-axis of Figure 4.9 was log-transformed.
The age at baseline was discretized according to the quartiles. For patients on tocilizumab, the rate of remission increased slightly with the age in comparison to etanercept. The data show that tocilizumab starts to be more effective than etanercept with increasing age (Table 4.4 and Figure 4.9). However, these values and CI were calculated with small number of events because of the stratification.

<table>
<thead>
<tr>
<th>Quartile (No. of events)</th>
<th>20.8,44.5 (n=21)</th>
<th>44.5,56.5 (n=24)</th>
<th>56.5,64.1 (n=21)</th>
<th>64.1,82.4 (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRR (95% CI) of TCZ vs. ETN</td>
<td>0.94 (0.39, 2.27)</td>
<td>2.31 (1.01, 5.28)</td>
<td>3.39 (1.37, 8.41)</td>
<td>4.18 (1.18, 14.80)</td>
</tr>
</tbody>
</table>

Table 4.4: IRR and 95% CI of tocilizumab (TCZ) comparing to etanercept (ETN) by age quartile.

Patients on tocilizumab had significantly more chances of remission, comparing to etanercept, when also on DMARDs (IRR: 2.76, 95% CI, 1.63 to 4.67).

The data suggests that individuals of African descent have 94% less chances of remission when on adalimumab, compared to etanercept (IRR, 0.06 0.01 to 0.46). The same does not happen for patients of European descent. However, European patients seem to respond better to tocilizumab than African patients, in comparison to etanercept (IRR: 2.29, 95% CI, 1.36 to 3.85). Results for Asian patients were not shown because there were not enough data.

The time since the onset of the disease was also discretized using its quartiles. Remission was significantly less likely for rituximab for the groups with the disease for 2.43 to 8.87 years and 20 to 48 years, comparing to etanercept (IRR, 0.24, 95% CI, 0.07 to 0.81 and IRR, 0.10, 95% CI, 0.01 to 0.77, respectively).

Generally, tocilizumab was as much or more likely than etanercept to lead to remission. In most of the comparison in this figure, rituximab was less likely to achieve remission in comparison with etanercept. The other drugs did not have statistically significant differences to etanercept, excluding the finding...
regarding adalimumab and subjects of African descent.

4.4 Kaplan-Meier estimates

The Kaplan-Meier estimates on this section came from dataset A and time to drug discontinuation data. Figure 4.8a shows a smaller therapy duration for anakinra and abatacept and almost no drug switches for etanercept. There was a significant difference between the Kaplan-Meier estimates ($P < 0.001$). The median persistence for anakinra was one year and two month months.

On Figure 4.8b, it was observed that the curves for abatacept and golimumab crossed the others curves and there were no remissions on anakinra. The intersection of the survival curves is an indication of non-proportionality of the hazards, however, these curves were estimated with few observations and the intersection of the curves could be associated with the larger uncertainty of the estimates for a small sample and not be meaningful of non-proportionality.

The etanercept curve overlaps with the infliximab curve during the first two years and the golimumab curve overlapping with the adalimumab curve during the first year. At the end of the first year, 50% of the patients on tocilizumab already reached remission. In fact, it seems that we could limit the follow-up period even more, to just one year, and avoid the apparent non-proportionality of the hazards. However, this would mean to discard even more data. The log-rank test gave a significant difference between the Kaplan-Meier estimates ($P < 0.001$).

In the appendix A of this work there are the KM estimates for remission and drug switch between patients on DMARD(s) and without DMARD(s) at baseline. There is also the KM for remission between patients naive to biologic drugs and at least in second biologic. None of this estimates was significantly different.

4.5 Cox regression analysis

On this section Cox proportional hazards regression was used for variable selection and estimate the adjusted HR of remission. Likewise, the reference group for the therapies was etanercept.

The HR are obtained by exponentiation of the coefficients of the regression. For categorical variables the HR is the relative risk between considered level and the reference group. For continuous predictors, the HR is the relative risk between two groups who differ by one unit. The estimated HRs are constant regardless of time for any two group of the same variable. The estimated HRs are constant regardless of time for any two group of the same variable and if the HR change over time it is considered that the estimate is the average risk during the follow-up period.

For this analysis the variables investigated using Cox regression were: biologic therapy, sex, age, race, weight at baseline, height, body mass index, alcohol, smoking, years of education, age of onset, DAS28 at baseline, SDAI at baseline, time since onset, season, DMARD therapy, GCs therapy, extra articular manifestations, patient VAS evaluation, physician VAS evaluation, pain VAS evaluation, HAQ, erythrocytes sedimentation rate, reactive C-protein, rheumatoid factor, anti-CCP, number of swollen joints,
number of tender joints, erosive and biologic number.

As seen in Figure 4.8b, the estimates of abatacept and golimumab, cross the other biologics KM estimates and there were no remissions with anakinra. However, these curves were estimated with few observations and the intersection of the curves could be associated with the larger uncertainty of the estimates for a small sample and not be meaningful of non-proportionality.

For this reason the abatacept and golimumab were kept in the data set and the fitted residuals of a univariate Cox model with the therapy as predictor had a non-significant slope for all therapies. The global test also failed to reject the null hypothesis of proportional hazards (P=0.845) confirming the hazards proportionality.

4.5.1 Model selection

Method 1

As referred in the Chapter 3 (see page 27), the first method for model selection consisted of including the significant variables on the initial model and sequentially removing the least significant variable from it. The purpose deleting the least significant variable was to maximize the precision of the model and keep only the variables that are adding information statistically relevant.

Table 4.5, shows the results for the Cox regression for dataset A. It was not found any significant association between the biologic number and remission, however, all the models based on the data from the cohort dataset had the baseline hazard function stratified by the biologic number. Dataset A patients were followed from enrollment on the therapy until remission or censoring, during a follow-up period of 4 years. Remission was achieved according to the index-based ACR/EULAR remission criterion for clinical trials (SDAI ≤ 3.3). On this table, the HR of different therapies were adjusted for confounding factors and the magnitude of the confounding was analyzed.

The interactions between therapy and sex, age, race (descent) and disease duration were this time investigated with Cox regression (see Figure 4.9 for the IRR unadjusted estimates). The interactions were investigated using bivariate models with and interaction term between one of the variables and therapy. It was found an interaction between the age and tocilizumab (HR, 1.04, 95%CI, 1.00 to 1.08, P=0.0456) and between no DMARD and tocilizumab (HR, 0.30, 95%CI, 0.01 to 0.90, P=0.032). The other interactions were not significant and the Asian group had to be removed for the method to converge.

From this variables, therapy, disease duration, weight, age, patient VAS, erosive, pain VAS, HAQ and years of education were statistically significant HR for the univariate regressions, considering a significance level of 10%. The complete cases of the dataset A regarding these variables were then used on the multivariate analysis. Regarding Table 4.5, all models used the sample with 136 observations and 61 remissions.
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Univariate Analysis</th>
<th>Stratified Multivariate Analysis, HR (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted</td>
<td>Model A</td>
</tr>
<tr>
<td>Biologic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etanercept</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Abatacept</td>
<td>0.67 (0.20-2.27)</td>
<td>1.28 (0.24-6.96)</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>0.37 (0.16-0.88)*</td>
<td>0.25 (0.09-0.67)**</td>
</tr>
<tr>
<td>Golimumab</td>
<td>0.66 (0.09-4.98)</td>
<td>1.54 (0.11-22.26)</td>
</tr>
<tr>
<td>Infliximab</td>
<td>0.83 (0.40-1.72)</td>
<td>1.30 (0.57-2.98)</td>
</tr>
<tr>
<td>Rituximab</td>
<td>0.30 (0.10-0.87)*</td>
<td>0.19 (0.04-0.89)*</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>1.90 (0.93-3.88).</td>
<td>3.17 (1.36-7.43)**</td>
</tr>
<tr>
<td>Disease duration</td>
<td>0.93 (0.89-0.96)***</td>
<td>0.93 (0.88-0.97)**</td>
</tr>
<tr>
<td>Weight</td>
<td>0.98 (0.96-1.00)*</td>
<td>0.96 (0.94-0.99)**</td>
</tr>
<tr>
<td>Age</td>
<td>0.96 (0.94-0.98)**</td>
<td>0.97 (0.94-0.99)*</td>
</tr>
<tr>
<td>Patient VAS</td>
<td>0.53 (0.36-0.79)**</td>
<td>0.98 (0.96-1.00).</td>
</tr>
<tr>
<td>Erosive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>1.08 (1.03-1.13)***</td>
<td>0.69 (0.35-1.34)</td>
</tr>
<tr>
<td>Pain VAS</td>
<td>0.61 (0.36-1.03).</td>
<td>1.01 (0.99-1.04)</td>
</tr>
<tr>
<td>HAQ</td>
<td>0.99 (0.98-1.00)*</td>
<td>0.80 (0.43-1.50)</td>
</tr>
<tr>
<td>Years of Education</td>
<td>0.99 (0.97-1.00)*</td>
<td>1.01 (0.95-1.07)</td>
</tr>
<tr>
<td>R²</td>
<td>–</td>
<td>0.41</td>
</tr>
<tr>
<td>AIC/BIC</td>
<td>–</td>
<td>356.97/397.75</td>
</tr>
<tr>
<td>Global PH test</td>
<td>–</td>
<td>0.5417</td>
</tr>
</tbody>
</table>

Table 4.5: Cox Proportional Hazards regression results for predictors of remission, Hazard Ratio (95% CI).1

<table>
<thead>
<tr>
<th>Magnitude of Coounding (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abatacept</td>
</tr>
<tr>
<td>91 %</td>
</tr>
</tbody>
</table>

Table 4.6: Magnitude of Confounding. Percentage change between the unadjusted and the adjusted estimates, according to the estimates of Model A (Table 4.5).

1 Significance codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1
The multivariate models estimation considered different baseline hazard functions for the number of the present biologic (biologic number), for this reason they were stratified multivariate Cox regressions. After accounting for the information of all predictors, just adalimumab, rituximab, tocilizumab, disease duration, weight and age remained significantly associated with remission at the population level.

A simple way of determining if there is confounding factors that may bias the result is by measuring the percentage change of magnitude of measure of association between the unadjusted and the adjusted estimate. If the percentage difference is larger than 10%, confounding is present. Otherwise, there was little confounding and is not meaningful [49].

Table 4.6, shows the confounding magnitudes according to the unadjusted estimates and the adjusted estimates of Model A (Table 4.5). The treatment effect of the majority of the biologics was underestimated in the univariate analysis. The HR for the indicator of erosiveness and HAQ was overestimated while the the patient and pain VAS were overestimated. Years of education, weight, age and disease duration barely changed in magnitude in comparison with the unadjusted counterparts. Therefore, there was no indication that the relationship between remission and therapy was confounded by these variables, nor that the relationship between this group of variables was confounded by treatment.

In Table 4.5, it should be noted that the maximum possible R-squared for this models was 0.947 and that the definition of a good R-square depends of the field one is working on. The Grambsch-Therneau test of trend in Schoenfeld residual failed to reject the null hypothesis in all variables and models, with exception of HAQ in Model B. According to the AIC and the BIC, the model more likely to be the true model was Model E (AIC=352.18). By continuing eliminating the least significant variable, patient VAS would have been eliminated and the resulting model would have an increased AIC equal to 354.12. The next model would have a even higher AIC of 360.84.

Model E, counted with 136 observations like all the other models on this table and was the HR and 95% CI were adjusted for disease duration, weight and age. From 136 observations, 62 remissions occurred and the pseudo coefficient of determination was 0.39.

On this model, there was an increased association of infliximab with remission, comparing to etanercept. However, this finding was not statistically significant (HR, 1.45, 95% CI, 0.68 to 3.10, P=0.341). The same happened for golimumab (HR, 2.06, 95% CI, 0.16 to 27.28, P=0.583) and abatacept (HR, 1.44, 95% CI, 0.27 to 7.77, P=0.672).

Subjects receiving adalimumab had 73% less chances of remission than subjects on etanercept. After accounting for sample variability the result was significant, and patients on adalimumab had between 30% to 90% less chances of remission (HR, 0.27, 95% CI, 0.10 to 0.70, P=0.007). Tocilizumab was again the biologic that provided best chances of remission, this time with adjusted estimates. Subjects on tocilizumab were 2.99 times more likely of achieving remission in comparison the reference group (HR, 2.99, 95% CI, 1.33 to 6.72, P=0.008).

The time since the onset of the disease at baseline (disease duration) was negatively associated with remission (HR, 0.91, 95% CI, 0.87 to 0.95, P<0.001). For each year of RA the chances of remission decrease on average 5% to 13%. The weight was also negatively associated with remission. An increase of a kilogram reduces the chances of remission between 1% and 6% (HR, 0.96, 95% CI, 0.94
to 0.99, P=0.002).

The results also show a significant negative association of age with remission. For every year of age, remission is 3% less likely (HR, 0.97, 95% CI, 0.95 to 0.99, P=0.03). It should be noted that the HR for age and disease duration increase over time (r=0.20).

The value of the patient VAS self-evaluation at baseline, or PtGA, had also a negative effect on achieving remission. A centimeter difference represents a decrease of 12% of the chances of remission (HR, 0.99, 95% CI, 0.98 to 1.00, P=0.496).

The estimates for the remaining variables were not significant on the other models. Nevertheless, each unit increase of HAQ at baseline reduces the chances of remission by circa 20% and the presence of bone erosion at baseline gives the prognosis of less 30% chances of remission.

Method 2

The next method consisted of including 25 of the most relevant variables in the model and subsequently remove the one of the variables according with four different criteria. The number of variables was limited to 25 to increase the number of complete cases in the data. The first criterion was the same as previous one: remove the least significant variable. The three other criteria consisted of removing the the variable that will produce the best model according with a measures of quality of the statistical models used for model selection (AIC, AICc and BIC). The therapy variable was forcibly kept in the models even if removing it would produce a better model.

In addition, this experiment was performed for three datasets extracted from the original but with different characteristics. Dataset A is the dataset already used on the previous analysis and is based on the cohort dataset, where the patients are uniquely divided in groups that share a particular drug and the follow-up period is 4 years. Again on this analysis, it was used only the time to the first remission of each subject.

At the end of the first year of follow-up it was already apparent differences of the KM estimates and to avoid the apparent non-proportionality of golimumab and adalimumab (Figure 4.8b) it was created dataset B. Dataset B characteristics are in all aspects similar to dataset A but the follow-up period was reduce to just one year.

Dataset C consists of time to remission data extracted from the original data. For each drug of each patient there is an observation for the time to the first remission. As result of the repeated measures for each subject the variance models used for this data were corrected to account for within-subject correlation of the observations. Like all other models, the baseline hazard function was also stratified by the biologic number.

Figures 4.10a, 4.10b and 4.10c, show the evolution of the AIC, AICc and BIC values during the backward elimination for each one of the criteria. Subjects on anakinra were excluded from the data because there were no remissions. The p-value criterion was only used for dataset A because it produced mostly the same results than the other criteria and the choice of the least significant variable was ambiguous when comparing the significance of the categorical variables levels with continuous variables.

The figures show that most of the criteria produced the same results. Concerning the scores, it can
Figure 4.10: Evolution of different scores during backward elimination with different data sets and different criteria to eliminate a variable. The x-axes are the number of variables on the model.

be seen that the AICc asymptotically converges for the AIC values when the number of parameters decreases in relation to the number of observations. In addition, the minimum of the BIC score happens for a models with slightly less covariates than for the AIC and AICc score.

It should be recalled that these scores are not comparable for different datasets. The different datasets will be compared to find what covariates are common across models and to contrast the direction and degree of association of the covariates between models.

Table 4.7 presents the best set covariates accordingly with the criteria defined for variable elimination. The models for AIC and AICc are grouped because the minimum for these scores resulted from the same set of covariates. On the left are all the variables that were present in at least one of the best models. The maximum possible R-squared for this models was 0.951 to dataset A, 0.863 to dataset B and 0.928 to dataset C. The global Grambsch-Therneau test of trend in Schoenfeld residual failed to reject the null hypothesis in all models. On second model (AIC/AICc criterion, dataset B) the Schoenfeld residuals for the doctor evaluation of the patient expressed through a VAS (r=0.24, P=0.048), the number of swollen joints (r=0.40, P=0.0123) and the weight (r=0.31, P=0.026) had increased significantly during the follow-up period. On the third model, the HAQ (r=0.2080, P=0.025), the age of RA onset (r=-0.2065, P=0.015) and age at baseline (r=0.2210, P=0.012) residuals changed over time.

On the sixth model, disease duration (r=-0.20, P=0.0356), age of RA onset (r=-0.196, P=0.0404), age (r=0.19, P=0.049), ESR (r=-0.16856, P=0.0604), DAS28 (r=0.16790, P=0.0343) and HAQ (r=0.18183, P=0.0387) also changed over time.

Comparing the best models for the dataset A, the BIC model does not have the ESR and the patient self-assessment VAS. The differences between the models for dataset B are larger and the BIC model has less nine covariates than the AIC/AICc model. BIC’s model from dataset C discarded the doctor VAS, the SDAI at baseline, the indicator for rheumatoid factor, DMARD(s) and erosion.

Comparing the direction of the association with remission, it was observed that the estimates for adalimumab and abatacept were not consistent among the different datasets and models. Rituximab and tocilizumab were the therapies with the most consistent results between the different models. The
Table 4.7: Best models obtained. Swollen joint count (SJC), rheumatoid factor (RF), extra articular manifestations (EA Manif.).

The association of rituximab with remission, in comparison to etanercept, was negative in all models and ranged between 0.12 and 0.57. The HR of tocilizumab and etanercept was always significant except for the first model (AIC/AICc criteria, dataset A, \(P=0.091\)) and ranged between 2.12 and 18.09.

The disease duration was just not kept on the models with dataset B and had significant association with remission in the remaining models it was present. Across the models, the chances of remission decrease between 8% to 18% with each extra year of RA. Accounting for sample variability on dataset A, the chances of remission for the model found using the AIC/AICc criteria decreased between 3% and 12% (HR, 0.92, 95% CI, 0.88 to 0.97, \(P<0.001\)). For dataset C, the chances of remission decreased between 9% and 25% (HR, 0.82, 95% CI, 0.75 to 0.91, \(P<0.001\)).

The weight of the patient was only not present on the BIC model with dataset B and had significant association with remission in the remaining models it was present. Across the models, the chances of remission decrease between 8% to 18% with each extra year of RA. Accounting for sample variability on dataset A, the chances of remission for the model found using the AIC/AICc criteria decreased between 3% and 12% (HR, 0.92, 95% CI, 0.88 to 0.97, \(P<0.001\)). For dataset C, the chances of remission decreased between 9% and 25% (HR, 0.82, 95% CI, 0.75 to 0.91, \(P<0.001\)).

The weight of the patient was only not present on the BIC model with dataset C. The effect of the increase of one kilogram ranged between 0.94 and 0.98 across the models and after accounting for sample variability the chances of remission decreased between 2% and 6% (AIC/AICc criteria, dataset A: HR, 0.96, 95% CI, 0.94 to 0.98, \(P<0.001\)).

The age of the patient was only not present on the BIC model with dataset B and the results, even though significant, were not consistent between datasets. On dataset C the age of the patient is positively associated with remission (AIC/AICc criteria, dataset C: HR, 1.14, 95% CI, 1.03 to 1.27, \(P=0.011\)), while on the other datasets was negatively associated (AIC/AICc criteria, dataset A: HR, 0.96, 95% CI, 0.94 to 0.98, \(P<0.001\)).
The ESR was present at least once in one of the models for each dataset and was positively associated with remission. An increase of one millimeter per hour at baseline makes remission 2% more likely (AIC/AICc criteria, dataset C: HR, 1.02, 95% CI, 1.01 to 1.03, P<0.001).

The patient VAS covariate appeared only on the AIC/AICc models for dataset A and B. In both models the effect of the patient VAS at baseline was of reducing the chances of remission (AIC/AICc criteria, dataset A: HR, 0.98, 95% CI, 0.97 to 1.00, P=0.026).

For the model selection AIC/AICc criteria and dataset B, the analysis for the one year follow-up period gave a significant increase of the rate of remissions for the fall when comparing with the rest of the year seasons. The indicator of bone erosion, the number of swollen joints at baseline, the extra-articular manifestations, the DAS28 and HAQ at baseline were negatively associated with remission. The SDAI and Doctor VAS evaluation at baseline was positively associated with remission.

For the model selection AIC/AICc criteria and dataset C, the year season, SJC, the extra-articular manifestations and patient VAS were no longer select by the AIC and AICc method to be included in the model. In contrast, the of RA onset, the indicators for RF and concomitant DMARD(s) were selected. The results show that every extra year of age of disease onset make achieving remission 13% less likely (AIC/AICc criteria, dataset C: HR, 0.87, 95% CI, 0.78 to 0.96, P=0.006). This finding were very similar for the BIC model for the same data.

The results for this model also show that RF positive patients 84% more likely to achieve (AIC/AICc criteria, dataset C: HR, 1.84, 95% CI, 1.09 to 3.12, P=0.023). The same model also gives 1.7 times more chances for patients without concomitant DMARD(s) (HR, 2.70, 95% CI, 1.44 to 5.08, P=0.002).

The BIC models for dataset B and C had similar results to the AIC/AICc models for the DAS28, HAQ and age of RA onset.

4.5.2 Recurrent events analysis

Models that consider only the first event might be discarding important information that can be present on the data with multiple events per subjects. This results were found using the first four recurrent remissions for each subject and each one only experiences one drug (dataset A).

Figure 4.11, shows KM estimates for the first four remissions. In some cases there was more recurrent remissions for the same subject but the data was limited to the first four due to the short amount of subjects. Assuming that the time for a recurrent remission is independent, there was a significant difference between the KM curves (P<0.001).

From 368 records, 146 achieved remission and 125 relapsed. From these 125 relapses, 62 subjects eventually had a second remission. While the median time for the first remission was almost three years (Median, 2.88, 95%CI, 2.21 to 3.47 years), the median time for the second remission was nine months (Median, 9.41, 95%CI, 8.32 to 16.81 months). It is possible to observe that KM estimates are ordered regarding the time to remission and the remission number. The first remission is the one that takes longer to achieve, and progressively, from the second to the fourth remission, the time to the event gets
shorter.

Table 4.8 presents the results of the Cox regression with the recurrent events models, introduced in Section 3.2.5 (see page 24). The covariates were chosen from the first model of Table 4.7 (Criteria AIC/AICc, dataset A for first remission) and the anakinra cohort was removed because it presented no remissions. The complete cases for this selection of variables counted with 255 subjects and a total of 396 observations. Overall there was 190 remissions and the total time since the beginning of follow-up or relapse until remission (Gap-time) was 478.31 years (Median: 285 days).

The patient identification number was added to the models for robust estimation of the variance and all the estimation were stratified by biologic number. With the exception of the AG model, the baseline hazard function of all other models was stratified by remission number\(^2\). The other differences between models are due to the data layout and time-risk structure.

The global test for trend in residuals only failed to reject hazards proportionality for the PWP-GT model. By analyzing the residuals for each covariate, there was evidence for the change over time of the HR of ESR, which increases \((r=0.15, P=0.021)\). Since the PWP-CP model had the best AIC, Table 4.9 shows the results of the PH tests for each covariate for this model. The p-values suggest that there is enough evidence for the increase over time of the of the HR of RA duration \((r=0.12, P=0.016)\), age \((r=0.13, P=0.028)\) and patient VAS \((r=0.12, P=0.024)\) for a significance level of 5%.

The HR estimates for the biological therapies (Table 4.8), were significant for adalimumab and rit-

2counter of the number of remissions and the number of the risk-interval
Table 4.8: Cox Proportional Hazards models for recurrent events. Prentice, Williams and Peterson (PWP) conditional model with counting process (CP) and gap time (GT) time scales. Wei, Lin and Weissfeld (WLW) marginal model. Total time-restricted (TT-R)

<table>
<thead>
<tr>
<th>Model</th>
<th>AG</th>
<th>PWP-CP</th>
<th>PWP-GT</th>
<th>WLW</th>
<th>TT-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etanercept</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Abatacept</td>
<td>1.22 (0.48-3.13)</td>
<td>0.90 (0.38-2.15)</td>
<td>1.24 (0.49-3.15)</td>
<td>2.24 (0.79-6.40)</td>
<td>1.44 (0.55-3.73)</td>
</tr>
<tr>
<td>Adalimumab</td>
<td>0.40 (0.21-0.77)**</td>
<td>0.39 (0.23-0.68)**</td>
<td>0.41 (0.23-0.73)**</td>
<td>0.42 (0.20-0.89)*</td>
<td>0.46 (0.26-0.82)**</td>
</tr>
<tr>
<td>Golimumab</td>
<td>1.27 (0.56-2.85)</td>
<td>1.10 (0.54-2.22)</td>
<td>1.21 (0.58-2.49)</td>
<td>1.93 (0.74-5.05)</td>
<td>1.44 (0.69-3.02)</td>
</tr>
<tr>
<td>Infliximab</td>
<td>1.00 (0.58-1.69)</td>
<td>0.79 (0.51-1.20)</td>
<td>0.87 (0.58-1.29)</td>
<td>1.07 (0.59-1.95)</td>
<td>1.08 (0.67-1.73)</td>
</tr>
<tr>
<td>Rituximab</td>
<td>0.29 (0.15-0.69)**</td>
<td>0.28 (0.12-0.63)**</td>
<td>0.32 (0.14-0.72)**</td>
<td>0.24 (0.09-0.61)**</td>
<td>0.36 (0.16-0.81)*</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>1.45 (0.86-2.45)</td>
<td>1.01 (0.63-1.62)</td>
<td>1.26 (0.77-2.05)</td>
<td>1.81 (0.98-3.36)</td>
<td>1.53 (0.89-2.60)</td>
</tr>
<tr>
<td>Disease duration</td>
<td>0.97 (0.94-0.99)*</td>
<td>0.98 (0.96-1.01)</td>
<td>0.98 (0.96-1.01)</td>
<td>0.96 (0.93-1.00)*</td>
<td>0.97 (0.94-1.00)*</td>
</tr>
<tr>
<td>Weight</td>
<td>0.98 (0.97-1.00)*</td>
<td>0.98 (0.97-1.00)*</td>
<td>0.98 (0.97-1.00)*</td>
<td>0.98 (0.97-1.00)*</td>
<td>0.98 (0.97-1.00)*</td>
</tr>
<tr>
<td>Age</td>
<td>0.99 (0.98-1.00)</td>
<td>0.99 (0.98-1.00)</td>
<td>0.99 (0.97-1.00)*</td>
<td>0.99 (0.97-1.01)</td>
<td>0.99 (0.98-1.01)</td>
</tr>
<tr>
<td>ESR</td>
<td>1.00 (1.00-1.01)</td>
<td>1.01 (1.00-1.01)*</td>
<td>1.01 (1.00-1.01).</td>
<td>1.00 (0.99-1.01)</td>
<td>1.00 (1.00-1.01)</td>
</tr>
<tr>
<td>Patient VAS</td>
<td>0.99 (0.98-1.00)*</td>
<td>0.99 (0.98-1.00).</td>
<td>0.99 (0.98-1.00).</td>
<td>0.99 (0.98-1.00)*</td>
<td>0.99 (0.99-1.00)</td>
</tr>
</tbody>
</table>

| R² | 0.20 | 0.11 | 0.13 | 0.15 | 0.13 |
| AIC/BIC | 1402.66/1446.46 | 1072.49/1116.29 | 1163.79/1207.58 | 1910.75/1964.76 | 1225.75/1269.54 |
| Global PH test | 0.0730 | 0.0180 | 0.6358 | <0.0001 | 0.0004 |

Table 4.9: P-values for the PWP-CP model covariates on Table 4.8.

uximab. This had already happened for other models with dataset A (Tables 4.5 and 4.7) but on this models the tocilizumab estimates were not significant. However, the direction of association for this therapies and golimumab comparing to etanercept was the same between recurrent remissions models and first events models. In relation to infliximab and abatacept, the estimates for the first were not consistent to the other models and the HR of the second fluctuated around one between models.

For the PWP-GP model, patients on adalimumab were 59% less likely to achieve remission than if they had enrolled on etanercept (HR, 0.41, 95% CI, 0.23 to 0.73, P=0.002). In addition, patients on rituximab were 68% less likely to achieve remission (HR, 0.32, 95% CI, 0.14 to 0.72, P=0.006) and patients on tocilizumab were 26% more likely to achieve remission (HR, 0.32, 95% CI, 0.77 to 2.05, P=0.352) both relative to the effect of etanercept.

The disease duration, weight, age, ESR and patient VAS had results very similar to the previously obtained estimates. An increase of one unit of disease duration or age reduces the chances of remission by 2%. It is also more difficult for older and patients with a higher patient VAS at baseline to achieve remission. Nevertheless, higher ESR at baseline increased the chances of remission (HR, 1.01, 95% CI, 1.00 to 1.01, P=0.064) like it was already seen in estimates of other models.
Chapter 5

Conclusions

This chapter summarizes and interprets the main findings of this thesis in light of the investigation on EMR data from RA patients enrolled in biological therapies and the initial objectives. In addition, it will be discussed the advantages and limitations of the proposed solutions, possible applications and future work.

In this work the problem of analyzing EMR data from RA patients enrolled in biological therapies was addressed with success. First, literature was reviewed about the disease, treatments, guidelines, scores, remission criteria and kind of methods used in this kind of study.

Then, from the original raw dataset were extracted multiple datasets to allow for the different analyzes throughout this work. This was done to overcome the difficulties of dealing with real life data without a specific type of study design during the data collection phase, multiple observations and no specific pattern for the drug switches.

After finding the appropriate ways to analyze Reuma.pt data and of understanding the implications and limitations of the methods, the methodology allowed to represent the data, summarize it and quantify the relative chances of remission (relative risk of remission) for the different therapy, other factors and also to adjust for possible confounding variables.

Every analysis in this work including tables and figures was scripted in R, a free software widely used for statistical computing.

5.1 Descriptive analysis

Despite the simple nature of the descriptive analysis (Section 4.1, page 29) a lot meaningful and easily interpretable results were found and can be discussed.

The data exploration showed that some of the most relevant variables that characterize the disease activity and functional disability had a high degree of missingness and alarmed for the difficulties regarding missing values that could be found in the rest of the work.

The proportion of women on this sample of Reuma.pt was almost 6 times greater than of men. However, the prevalence for women is estimated to be two to three time higher than in men and in
Portugal this proportion is 4:1 [10, 9]. This probably happened because there was more women failing the conventional therapies than men. Regarding RF and anti-CCP, there was about 2 times more patient with it than without. This values are similar to the prevalences found in the general population [11].

It was also found a considerable variety of combinations of DMARDs (n=18) and 1% of the observations had three DMARDs combined. It was also found a combination of three GCs on the 12 different combinations found.

Drugs combination and their interaction is a matter that is difficult to study because of the amount of possible combinations of drugs and the high cost of such an investigations. The drug interaction might cause a exacerbated effect of the drug or decreased effect when the interaction is antagonistic. For this reason, this is still not well studied and more research should be done regarding this area.

The observation of anatomical distribution of the swollen and tender joints at baseline showed that a joint was more likely to be swollen than tender, the distribution is almost symmetric as expected in RA and the wrists and the hands joints are the most affect. In the hands, the second and third metacarpophalangeal joints were the most affected.

On the next part, it was represented the change of disease activity, functional disability over time by therapy groups. It should be noted that this estimates are biased because the windows for the smoothing and kernel estimation might contain more than one observation per patient, but since observations from the same patient are not expected to be found close to one another the effect of this was not relevant.

It is not relevant because this estimates serve only the purpose of data visualization and for aiding on the interpretation of the data, and not for statistical inference of the therapies effects on the population of patients. Descriptive statistics and analysis is concerned only with properties of the sample, while inferential statistics tries to extrapolate the properties found in the data sample to the underlying population.

Also on this analysis, it was of interest to describe each drug effect near the third and sixth month after the therapy start because those are the decision time points.

The results show that most of the disease activity variation occurred for the first 6 to 12 months after the therapy start.

Given the consistence of the results regarding the anti-TNF drug and the lack of significant difference in treatment effects between anti-TNF drugs, the descriptive analysis suggests that these drugs are equivalent.

The overwhelming difference of the estimates and the faster and larger improvements for tocilizumab, comparing to the other drugs, suggests that this drug produces the best results. Rituximab and anakinra had the worst results both for the progression of disease activity and functional disability. It became evident early in this work that the best therapy response came from patients on tocilizumab. This finding turned to be reinforced through out the work for the cohort dataset but also for other datasets with data from all the biologics each patient was exposed.
5.2 Summary measures analysis

The baseline characteristics of each cohort were compared on Table 4.1 and then it was presented the results for the analysis of the data summaries on Table 4.2, both on page 30.

Globally, tocilizumab was the therapy with better results. It had the highest mean rate of decrease of disease activity and the highest percentage significant changes of disease activity. Others summaries like the AUC and the minimum activity or time to minimum activity were not possible to use because the subjects do not have equal follow-up times making this measurement incomparable.

5.3 Incidence density rates

The study of drug persistence was difficult in this work. It should be note that the drugs were not introduced in the market at the same time and some of these drugs were already authorized for some years when others were approved in the Portuguese market. For example, according to Infarmed, tocilizumab was only authorized in 2010. It was not possible to precise when the others drugs were authorized but is mentioned that adalimumab, anakinra, etanercept, infliximab and rituximab did not have a cost-effectiveness analysis because they were authorized to enter in the market in prior to 2006. This was the year that this analysis started to be mandatory for this drugs. For golimumab was made a cost-effectiveness analysis and for abatacept is only mentioned that should be used by RA adults [50].

Since drug persistence does not depend only of the overall success of the therapy or of the doctor and patient preferences. For all this factors and not having enough elements to control for them, it was chosen to limit the analysis to descriptive statistics and unadjusted estimates like the KMs, IRs and IRRs.

It was observed that the therapy duration was greater than the overall median of two years and four months for adalimumab, etanercept, infliximab and rituximab. Etanercept was the most persistence drug with about two drug switches by 10000 person-months and this was also visible on the KM curve. Anakinra and abatacept had highest rates of drug switch.

The IR of remission was best for tocilizumab and worse for anakinra and rituximab. Subjects on tocilizumab had about two times more chances of remission than the ones on etanercept. After accounting for sample variability the subject had 36% to 240% more chances of remission with tocilizumab when with etanercept. This results was statistically significant because the 95% CI did not cross one.

In relation to etanercept, it was observed that the chances of remission increases gradually with age for patients on tocilizumab. In addition, tocilizumab performed better than etanercept when in combination with one or more DMARDs and adalimumab decreased the chances of remission in comparison to etanercept for people of African descent. At last, rituximab was generally worse than etanercept.

There is the possibility that this results are confounded because they are unadjusted estimates.

The therapy response using the EULAR criterion for therapy response was difficult because the criterion has three cut-points (good response, moderate response and low response) and by comparing different levels groupings some of the results were contradictory. In addition, the DAS28 variation was
one of the variable with more missing values and is used to determined the patient response.

5.4 Kaplan-Meier estimates

Tocilizumab’s KM estimate for remission was clearly isolated from the other drugs. The intersection of the survival curves for abatacept and golimumab was probably result of the small number of observations for this drugs and it can not be said with certainty that the hazard are non-proportional. For this reason, further investigation was done when applying the Cox proportional hazards regression.

The median persistence for anakinra was one year and two month months. In contrast, the median persistence for tocilizumab was one year and eight months.

Given the characteristics of the outcomes in the data, it was found that the disease activity is a better outcome than the therapy persistence or the therapy response. DAS28 and SDAI are two good measures of the disease activity. While DAS28 was developed to be normally distributed, SDAI is the current score used to define remission.

5.5 Cox regression analysis

Cox proportional hazards regressions were used to assess potential predictors of remission and adjust for variables correlated both with the outcome and other covariates. This variables, also called confounding variables or confounders, are one of the main problems with non-randomized studies and can cause associations between to the outcome that are not true if it is not performed proper adjustment for them. However, it is not possible to know if all confounders are in the data, we are adjusting for all of them or the method is flexible enough permit for the proper adjustment. So, the best way to achieve maximum strength of causality association between the exposures and the outcome and protect the sample against confounding is to do a randomized study.

5.5.1 Model Selection

Method 1

To find a the most probable to represent the data were compared different models. The first procedure was to fit an univariate Cox model for each candidate predictors of remission. Then the insignificant covariates for a level of 10% were discarded and the remaining were included on a multivariate Cox model. At last, it was used backward elimination, eliminating always the least significant variable.

The aim of deleting the least significant variable was to maximize the precision of the model and keep only the variables that are adding information statistically. Combinations of other covariates were also experimented.

Years of education was statistically significant on the univariate model. This possible relation with remission could be connected with the socioeconomic characteristics of the patients: alimentation, stress
access to complementary therapies, and the care that more informed patient might give to treatment. However, this was not confirmed on the multivariate models.

There was indication that the relationship between therapy and remission was confounded by patient VAS, erosiveness, pain VAS, years of education and HAQ, and that the effect of these variables was also confounded by therapy.

The model most likely to represent this data included therapy, disease duration, weight, age and patient VAS as predictors of remission. Also there was not enough evidence of non-proportionality of the hazards, even including the abatacept and golimumab cohorts.

It should be noted that the HR for age and disease duration increase over time (r=0.20). This results was expected because the HR for one year difference when one is in their thirties does not have the same impact in the HR of when one is in their sixties.

In all the models, subjects tocilizumab had the highest chances of remission in relation to the etanercept. These subject were about 3 times more likely to achieve remission. For the same reference, subjects on adalimumab had 73% less chances of remission and on rituximab 82%. RA duration, weight and age had significant negative effects.

Every year since the beginning of RA and of age decreases the chances of remission 8% and 1%, respectively. One more kilogram of weight makes 4% more likely. It was also found, with exception of adalimumab, that there was not a statistically difference between the anti-TNF.

Method 2

Another criteria for model selection were tried but this time with more variables. Namely, it was again used backward elimination, but this time it was removed the variable that provided the best model accordingly to the AIC, AICc or BIC. It were also used different datasets. Dataset A was again used for this method. Dataset B, like dataset A, comes from cohort dataset but has a 1 year follow-up period. Dataset C has the time to remission or censoring for each biologic therapy of each subject.

It was found that this model selection method does not depend much of the criterion for variable elimination but of the score used to find the best model. For example for dataset A, the model with minimum AIC is equal for both the AIC or BIC criterion. The same happened using the p-value criterion. In addition, the minimum of the BIC score happens for smaller models.

Therapy was the only variable that was present in all models. However, therapy was never a candidate for elimination of the models. The remission predictors present in three or more models were: disease duration, weight, age, ESR and HAQ.

On the best models for dataset C the effect of age was positive regarding remission and the DAS28 and SDAI scores were included and considered statistically significant. One possible explanation for the first case to happen is that every time the patient switches drug the age is correspondent to the moment of beginning of the new drug. So, the age increases for every new drug and also the the likelihood of remission, because the therapy was adjusted. The same happens for the ESR, DAS28 and SDAI values and a better way is to make this variable time-independent and include only the overall baseline value in all the observations from the same subject.
This systematic technique was simpler and easier way of selecting covariates than trying all possible combinations of covariates.

By comparing the results for the best models of the three datasets for the different criteria (minimum AIC for AIC criterion, etc.) it was found that the greatest effect on remission was from tocilizumab. The estimated HR of tocilizumab in relation to etanercept across the different models was between 2.12 and 18.09.

The strikingly higher frequency of remission among patients on tocilizumab in our data and across different models suggests an important effect of this drug on the reduction of the RA activity and remission induction. This results reinforce the relevance of the IL-6 as pro-inflammatory cytokine in RA.

The effect of disease duration, weight and age was also consistent with previous models and of the same magnitude.

5.5.2 Recurrent events analysis

Remission does not last long in RA and relapses and flares of disease activity are common. For this reasons it were used and compared several methodologies for recurrent events.

For recurrent events, it was found that for half the subjects the time for the first remission can be 3.67 times greater than the time for the subsequent remissions. It was also observed that the mean remission-time decreases for every new remission.

For the recurrent events analysis, the effect of tocilizumab was not statistically significant even if positive and consistent with the other results. On other hand, the effects of adalimumab and rituximab was statistically significant and negative. Subject on adalimumab had 60% less chances of remission and the ones on rituximab were 71% less likely than the ones on etanercept. However, there was evidence of violation of the hazards proportionality assumption on this models.

The Cochrane overview on biologics for RA [51] compared the efficacy of abatacept, adalimumab, anakinra, etanercept, infliximab, and rituximab. They found that anakinra was less efficacious than etanercept (RR, 0.44, 95%CI, 0.23 to 0.85, P=0.014), rituximab and adalimumab and that the rest therapies had similar effects between them. This results are consistent with with the findings in this thesis for anakinra and the anti-TNF drugs.

5.6 Limitations

The understanding of the sampling procedure of the data is an important aspect because with it we can determine to whom the results may apply and how well we can extrapolate the results from the sample to the population of RA patients.

One of the factors that might cause a systematic difference between the sample and the population of interest is the non-random assignment of the therapy. The guidelines [24] recommend to prescribe first the anti-TNF class of biologics and on this group there were some drugs more frequent than others. Namely, etanercept (41.5%), infliximab (31.6%), adalimumab (18.3%) and golimumab (6.6%). This can
be due to some drugs only being available recently.

Knowing that non-responders to the first therapy are more likely to not respond for a second drug, this might bias the results in favor of the first therapies. In addition, patients that start a biological therapy already failed the conventional therapies, for this reason the results will only apply to the population that already failed therapy with DMARDs and not to the complete population of RA patients.

Furthermore, it is expected that some patients avoid going to the doctor when they have no complaints. For this reason, it can be expect that the loss to follow-up might not be completely at random, even different between treatment groups and changing these groups characteristics with time. Since the data comes from two health centers that might have populations with different accessibilities and socio-economics aspects, the access and quality of health care might be different between centers.

Consideration has been given to the possibility that the results could have been produce by selection of an unsuitable group of patients, by patients with unusual exacerbation of RA activity and unresponsive to the current therapies. This is supported by the fact that the sampling procedure created a selection bias of both patients and therapy and the data is also convenience sample.

There are superior methodologies to find evidence of association of particular variable to the outcome, establish a stronger causal relationship between the exposure and outcomes and avoid systematic difference between the exposure groups other then the outcome of interest (confounding). Namely, randomization eliminates systematic confounding and and appropriate observational study design can be used to increase the chances of a good representation of the population of interest by the sample. On non-random samples same elements might not reflect the makeup of the population of patients.

Being EMR data there was not a rigorous control of the data collection as in a randomized or typical observational study and, as seen, there was many missing values in the data. Furthermore some therapies, like anakinra, had small groups of subjects and could not be analyzed.

It exists a great number of joints to observe, so the results regarding this matter might be influenced by the 28 joints count for DAS28 or by the personal choice of joints to be examined by the doctor.

Another limitation of this work was the non-existence of a clear reference group for the therapies, like a placebo group or group only not on biological therapies.

One of the biggest drawbacks of the time-to-event analysis is the issue of non-informative censoring. To satisfy this assumption, the design of study must ensure that the mechanisms giving rise to censoring of individual subjects are not related to the probability of an event occurring. For example, the continuation of follow-up must not depend of the patient condition. This assumption can not be guaranteed in our data since there are many factor that can influence this. Violation of this assumption can invalidates both KM estimates and Cox models.

It should be taken into account that lack of statistical significance, is not the same as lack of scientific significance. The scientific significance depends of the context of the study and of the sample size. Small samples can produce non-significant results even though the magnitude of the association at the population is relevant. What happens is that our study can not detected it because does not have enough statistical power. The opposite also holds and not rejecting the null hypothesis is not equivalent to accepting the null hypothesis as the truth.
A limitation of the CI for the IR and IRR estimates is that they tend to work worse for small sample sizes and what would be expected to be a 95% CI has in fact a coverage smaller than 95%. This happens because the central limit theorem only works asymptotically. In addition, the p-values for the Chi-squared approximation may be inaccurate because of small sample size and the null hypothesis can be false, but erroneously fails to be rejected (type II error).

A group of methods suited to analyze longitudinal data are mixed effects models. Mixed models are statistical methods that contain both fixed effects and random effects and can deal with missing values and repeat measures of the same subject on the data. However, the linear mixed effect models and Cox regression with a frailty term experimented during this study had many times difficulty converging. For this reason their results were not included. Probably this happened because many subjects had few observations and that difficult the convergence of the iterative methods used.

**Cohort dataset** was created because most of the subjects start with anti-TNF biologics like it is advised on SPR guidelines [24] and by using only the first therapy the other types of biologics would be under represented on the analysis. In alternative, if it was used all patient's therapies the observations for different drugs would be correlated, like it happened with dataset C.

This choice has the disadvantage that we discard a great part of the data but the advantage that making possible to study the effect of the therapies with a greater variety of methods and less restrictions and assumptions. For example, this makes possible to present correct KM and Cox models from the same data and eliminates the difficulty of comparing the effect of therapies when the same subject is in two different groups.

A limitation of the remission criteria is the amount of relapses and that they just state the criteria must be fulfilled "at any time point". This makes possible to call remission to very short periods of disease inactivity and creates difficulties in defining the remission for prospective studies since the relapse rate is very high.

Linear methods are easy to use and most of all are easily interpretable. However, in the case that the hazard of a predictor changes over time and the hazards proportionality does not hold, some authors considered that the estimate can be interpreted had the average HR over the follow-up time [27, 35]. If the information is not representable by linear combination of variables or features, other methods are more appropriate.

### 5.7 Achievements

As is known, RA is one of the leading causes of disability and is associated with decreased life expectancy. Furthermore, precious time is lost looking for an effective therapy while it is given opportunity for the disease to cause irreversible joint destruction on the RA patients.

Therefore, some of the literature about rheumatoid arthritis, its treatments, guidelines, epidemiology, genetics and current criteria used to assess disease activity and functional ability were reviewed. Then, to construct a methodology for this work, literature about epidemiology, longitudinal data analysis, survival analysis, machine learning and other data sciences or methods were studied, paying attention to
the possible limitation to our case.

It should be mentioned that even though machine learning techniques were studied and tried on this data, the structure of the data limits a lot what kind of analysis can be made and the results would be less meaningful and interpretable than the results of this work.

Finally, this study investigated EMR data from Reuma.pt register and with it was able to extract important information about the sample and contributed the base knowledge about biological therapies for RA in clinical-practice. To achieve that, it was found the best set of predictors for different datasets and the estimates from the different models were compared. The model selection procedure contributed for the development of the best model to fit the this kind data.

More precisely, it was measured the effect on the disease activity of different biologic treatments and of factors like the disease duration, weight and age. In addition, it was found what therapy was more likely to succeed and which one was is more likely to fail according with the characteristics of the data using population based estimates. Other outcomes, like drug persistence and therapy response, were also investigated.

One of the biggest challenges in this work was to adapt real-world data to methods that were developed for dataset with a specific set of characteristics and assumptions.

This thesis gave me the opportunity to work with data on a real world context with all the inherent difficulties of such challenge. Since the beginning of this thesis I could learn to program with R, bash script and work with Weka. I could also deepen my knowledge about many data sciences areas, such as epidemiology, statistics and machine learning.

5.8 Future Work

This work can be used as a guide for future work on Reuma.pt and the estimates might be useful to design future studies and to compute the sample size for a desired statistical power. As is know, preliminary observational studies are first done to develop evidence for the possibility of doing larger and perhaps more costly randomized trials.

In addition, the covariates from the selected models can be later be used for future models and prediction of the hazards in new data.

Some of the therapies in this work did not have a cost-effectiveness study because it was not mandatory at the time they were authorized in the market. Namely, adalimumab, anakinra, etanercept, infliximab and rituximab. The optimization of the resource allocation is an important factor on the context of the sustainability of the health systems. Given the high costs of biologic therapies and the results of this work, future work should be done to analyze the cost-effectiveness of these medications.

In future work can instead use forward selection of covariates to select the best model or be centered on prediction and development of a recommendation system. Interactions between covariates and in time can be further investigated and the results adjusted for confounding variables.

A simple recommendation system that can be easily implemented is based on the K-nearest neighbors. Using a distance measure that uses both categorical, continuous variables, time dependent and
time-independent variables, the subjects can be ranked for each therapy. From this ranks the nearest K subjects are averaged to predict the response of the new subject and the best therapy with the best response is chosen.

Bayesian networks and other recommendation systems also have several applications in health care. Disease diagnosis, selection of optimal therapy, outcome prediction, health care management, construction of diseases models in clinical epidemiology and discovery of functional models through expression data are some of the possible applications in health care [52]. This methods do not need to have explicit representation of time, instead the time-dependent predictor can be represent by creating features that describe the temporal evolution over time, like it was done in the summarized analysis in this work.

Bayesian networks have also been extended to what is called an influence diagram and used in decision theory contexts. An influence diagram is a generalization of a Bayesian network, in which probabilistic inference is used to solve decision making problems by means of maximizing the expected utility. The utility function can depend of multiple factors and reflects the preferences of the agent about decisions that involve uncertainty. Depends on the attitude to risk of the agent and normally is not linear. This kind of system can be used, for example, to select the optimal treatment alternatives.

A group of methods that can be used in the future are the mixed-effects models and an alternative for Cox regression and recurrent events is to use Poisson regression.

To conclude, this study has developed a general methodology for the representation and analysis of Reuma.pt data. It has been demonstrated that some therapies are significantly more efficacious than others and that the time since the onset of the disease, age and weight affect the chances of remission. Further methods should be applied and this data can will be subject of more research.
Bibliography


[22] Josef S Smolen, Robert Landewé, Ferdinand C Breedveld, Maya Buch, Gerd Burmester, Maxime Dougados, Paul Emery, Cécile Gaujoux-Viala, Laure Gossec, Jackie Nam, Sofia Ramiro, Kevin


Appendix A

Kaplan-Meier Estimates

Figure A.1: Kaplan-Meier of the remission between naive and non-naive patients.

p = 0.112

Biological therapy:
- Naive
- Non-naive

Figure A.1: Kaplan-Meier of the remission between naive and non-naive patients.
Figure A.2: Kaplan-Meier of the drug survival between naive and non-naive patients.

Figure A.3: Kaplan-Meier of remission between patients with DMARD and without DMARD at baseline.