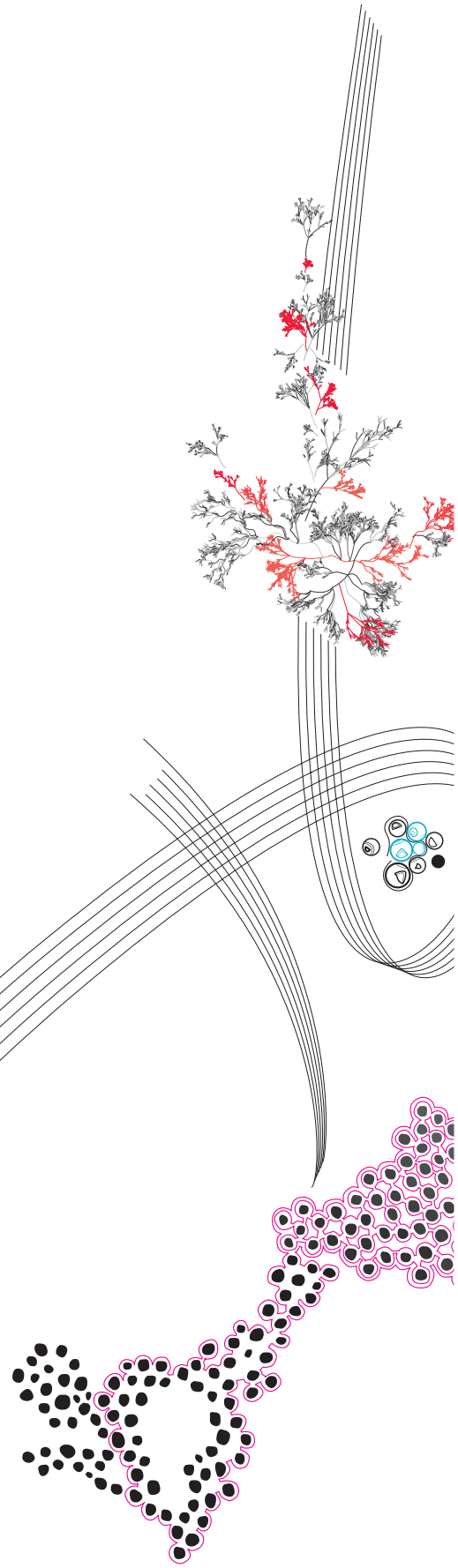


Masters Thesis Technical Computer Science

Modeling the effectiveness of treatment for Rheumatoid Arthritis with UPPAAL

T. Stouten

Primary Supervisor: DR.IR. R. LANGERAK
Secondary Supervisor: DR. M. POEL
External Supervisor: PROF.DR. H.E. VONKEMAN

February, 2024

Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Modeling the effectiveness and side effects of treatment for Rheumatoid Arthritis with UPPAAL

T. Stouten*

February, 2024

Abstract

Clinical trials are a necessity in order to measure the effectiveness of newly developed treatment protocols. This is no different for the treatment of Rheumatoid Arthritis. To ease the process of developing a new protocol, we aimed to make a credible patient model in UPPAAL that is able to react on newly developed treatment protocols. A tool to easily create these protocols in a format that UPPAAL supports is developed alongside the models, as well as a "run-time application" in which a medical professional can load the patient and the protocol model, resulting in a prediction of the effectiveness of said protocol.

In this research, we dived into the literature available for the treatment process of Rheumatoid Arthritis (RA). Using the publicly available literature, models which imitate the progression of RA symptoms when a treatment protocol is in use have been constructed. These models are purely meant as a proof of concept to see if such models could reasonably predict the outcome of a protocol, which can be implemented by a medical professional. This research goes no further than creating a discussion tool to show that creating and using such models are a viable option for the general patient population, but that it will require much more research before actual usage for specific patients is possible and credible. This means that applying those models to real patients in order to decide the best course of action is not the goal of this research and will not be endorsed by the developer.

Keywords: Rheumatoid Arthritis, UPPAAL, real-time modeling

*Email: t.stouten@alumnus.utwente.nl

Contents

1 Introduction

It is in the human nature to discover and invent as much as possible. This includes discovering and producing cures for illnesses and diseases. One of such diseases is Rheumatoid Arthritis. This is a disease that affects and damages the human joints, resulting in chronic pain and deformities. The prototype treatment protocol is able to get the disease under control, but is unable to fully cure it. New protocols that are designed must go through a clinical trial in order to verify the effectiveness, however, much work, proof of concept and legalities are required to conduct such a clinical trial.

Due to this, we aim to create digital tools to simulate a patient following a new protocol in order to estimate the effectiveness. The most important and difficult part of this research is getting the patient model just right, such that the effectiveness predictions for a protocol are sufficiently credible.

In this research we have developed models that are able to simulate the evolution of the disease activity of Rheumatoid Arthritis in patients. We start explaining what Rheumatoid Arthritis is in section 2, followed research questions in section 3 and the goals of this research in section 4. The stochastic model checker that we use as foundation, UPPAAL, is be discussed in section 5. Section 6 discusses existing research about medical examples in UPPAAL and section 7 describes the creation of the proof of concept models that we have constructed of which the predictive results are discussed in section 8. The infrastructure of our tool in order to reach our goals is discussed in section 9. After that, we discuss the prototype models we have created once the proof-of-concept models have proven that this research is possible in section 10, and the new results and the created tool are discussed in section 11. Finally, we discuss our findings in section 12, the conclusion of our research in section 13, and what requires more research in the future works section 14.

2 Rheumatoid Arthritis

Rheumatism is a group of conditions causing chronic (often intermittent) pain affecting the joint, muscles or connective tissue[2]. In this research, we take a better look at one specific form, Rheumatoid Arthritis. RA is an autoimmune and inflammatory disease of the joints. This means that your immune system wrongly attacks healthy cells, initially causing inflammation in the lining of small joints, progressing to larger joints, for example, the joints of the hands, wrists and knees. This lining becomes inflamed, causing damage to the bones and cartilage and weakening the tendons and ligaments. This damage to the joints causes deformities and bone erosion, which usually leads to severe chronic pain and unsteadiness.

2.1 Causes

Rheumatoid Arthritis is a multifactorial disease, meaning that a multitude of risk factors have been discovered that increases the probability of developing RA and or increases its severity[9][10][1]. These factors are grouped together by the type.

2.1.1 Epidemiological factors

Epidemiological factors include age, sex, family history with RA and ethnicity. The three most important factors are described why and how they could effect the development of RA.

Age: Although RA can occur at any age, including childhood, the likelihood of developing RA increases with age, with the highest likelihood among adults in their sixties[3].

Sex: Of all new cases, the probability of the patient being female is about 2 to 3 times higher than being male. According to Sparks et al, there is a cumulative risk of developing RA of 3.6% for woman and 1.7% for men[12]. It is uncertain why this is the case. A controversial topic is the role of hormones in the development of RA, as the stimulatory effects of oestrogen on the immune system may play a role in the development in RA.

Factors that increases the risk:

- Early menopause
- the presence of polycystic ovary syndrome, the most common endocrine disorder in woman of reproductive age
- pre-eclampsia, a pregnancy complication that includes symptoms like high blood pressure, high levels of protein in urine, nausea and severe headaches.
- post-partum periods.

Factors that decrease the risk:

- Breastfeeding
- Hormone replacement therapy
- Oral contraception

Family history: People with specific genes, like the HLA class II genotype, have an increased risk of developing RA and have an increased risk of RA being more severe compared to people without the risk genes[5].

2.1.2 Environmental factors

While genetics have long been considered due to the increased risk of developing RA when they are first degree relative of a RA patient[5], the overall risk due to genetics remains limited. Due to this relatively low risk factor, research has hinted towards the fact that environmental factors play a relatively bigger role in increasing the risk of developing RA.

Gut microbiota The gut microbiota co-develop with the host from birth. The composition of this microbiota play a major role in regulating the balance of Th17 cells, influencing the host immune responses, tolerances and susceptibility to autoimmune diseases such as RA

Dietary factors . Diets have been brought in connection with the development of RA as either a disease trigger or moderator, as food might exacerbate or reduce inflammation[11]. For example, excessive consumption of red meat and coffee increases the risk of developing RA, while consuming a healthy amount of fruit, vitamine c and olive oil decreases this risk. The exact data here in inconclusive at most, but noteworthy nonetheless.

Obesity White Adipose Tissue (WAT), the tissue where humans store energy, is suspected to be capable of storing and releasing cytokines, which may make individuals more susceptible to autoimmune inflammatory diseases by exacerbating inflammation[8]. While the data proving this is conflicting, evidence of the importance of obesity seems to become increasingly common.

Lung Mucosa It is suggested that mucosal inductive sites play a role potential role as RA-associated autoimmunity sites, meaning that this potentially is where RA is initiated[6].

Periodontal Disease Periodontitis (A serious gum infection) is a disease that is more frequently associated with subjects with RA than with the healthy population. This disease might be the trigger for a mucosal-based immune response which ultimately could lead to the development of RA.

Smoking Smoking is estimated to account for 20-30% of the environmental risk for RA. By smoking, an individual produces more of the PAD2 enzyme which leads to increased citrullination, the reaction for converting the amino acid arginine into the amino acid citrulline, which may lead to anti-citrullinated-proteine antibodies (ACPA). Therefore, smoking is seen as a risk factor for developing RA, mainly in ACPA-positive individuals. Furthermore, there is increasing evidence that smoking may also cause a more serious course of the disease.

Silica/dust inhalation Multiple studies have shown that firefighters and other people who are often exposed to silica or dust have shown a higher number of ACPA-positive RA cases compared to people who are not often exposed to it.

Virus infections Infectious agents may play a role in the enhanced development of autoimmune diseases by increasing the rate of this development.

2.1.3 Serological and immunological factors

Auto-antibodies The presence of anti-citrullinated-proteine antibodies is associated with the development of RA, however, this is a controversial claim until more clinical research is done; yet it indicates a potential risk factor.

Acute phase reactants

Many of these factors on their own introduce an almost negligible risk increase for the development of RA and are controversial to say the least. However, by combining a multitude of studies and many of these risks does create a good overview of what we currently know about the causes of Rheumatoid Arthritis.

2.2 Diagnosis

Rheumatoid Arthritis has many ways to present itself. The most common way to start suspecting RA is when a patient shows symptoms like morning stiffness lasting longer than 30 minutes paired with symmetrically swollen joints, especially of the hands and feet. Once there is a suspicion of RA being at play, a patient should get referred to a specialist, where the clinical diagnosis can begin[21].

For the diagnosis of Rheumatoid Arthritis, the first step usually is generating an EULAR score according to the ACR/EULAR criteria[21]. Those criteria are published in 2010 and can be read in table 1. This table must be filled in resulting in a score for every category, which combined form the EULAR criteria score. In order to correctly diagnose RA, a EULAR score of at least 6 in combination with evidence of synovitis (synovial inflammation) in at least one joint is required. However, different criteria may be present at different times, creating a cumulative score. The strictness of this varies per case, as the rheumatologists are aware of the fact that the later the patient starts treatment, the worse the outcome will be if the disease is indeed RA.

Once this combination of factors has been diagnosed, a disease activity score should be constructed that measures RA activity, in this study we focus on the Disease Activity Score in 28 different joints (DAS28). The DAS28 score is calculated by combining 4 different types of clinical data, Tender Joint Count 28 (TJC28), Swollen Joint Count 28 (SJC28, Erythrocyte Sedimentation Rate (ESR) and General Health (GH)[18].

TJC28 Tender joint count. This is the number of joints out of the 28 joints shown in figure 1 that are tender and painful at the time of examination.

SJC28 Swollen joint count. This is the number of joints out of the 28 joints shown in figure 1 that are swollen at the time of examination .

ESR Erythrocyte sedimentation rate (ESR) is a blood test that can show the presence of inflammation in your body. During an ESR test, a sample of the patient their blood is collected in a tall, thin test tube, after which it is left to settle for an hour. Once the hour has passed, the number of erythrocytes, red blood cells, that have settled at or sunk to the bottom of the test tube is measured. Erythrocytes always sink slowly in a blood sample; however, inflammation causes the red blood cells to clump together, making them heavier such that they sink faster. The number extracted from the ESR test that is used in the DAS28 score is the amount of Erythrocyte sedimentation after an hour in mm.

Classification criteria	Score
Joint involvement (0-5)	
1 large	0
2-10 large joints	1
1-3 small joints (large joints not counted)	2
4-10 small joints (large joints not counted)	3
>10 joints including at least one small joint	5
serology (0-3)	
Negative RF and negative anti-CCP	0
Low positive RF or low positive anti-CCP	2
high positive RF or high positive anti-CCP	3
Acute-phase reactants (0-1)	
Normal CRP and normal ESR	0
Abnormal CRP or abnormal ESR	1
Duration of symptoms (0-1)	
<6 weeks	0
≥6 weeks	1
RF = rheumatoid factor; anti-CCP = anti-cyclic citrullinated peptide antibodies; CRP = C-reactive protein; ESR = erythrocyte sedimination rate	

TABLE 1: Eular criteria[21]

GH General health. A patient is asked to mark on a Visual Analogue Scale, a 10cm line, how well they can deal with the pain and current annoyance of RA. 0 means that they do not experience any inconvenience, while 100 means that they can't live in an acceptable state.

The DAS28 score is then calculated with the following formula[18]:

$$DAS28 = 0.56 * sqrt(TJC28) + 0.28 * sqrt(SJC28) + 0.70 * ln(ESR) + 0.014 * GH$$

The DAS28 score is a number between 0 and 10 and indicates the disease activity in a patient. The definition of the score is as follows:

DAS28 > 5.1	High disease activity
3.2 < DAS28 ≤ 5.1	Moderate disease activity
2.6 ≤ DAS28 ≤ 3.2	Low disease activity
DAS28 < 2.6	Remission

2.3 Treatment

Curing Rheumatoid Arthritis is a long-term goal that is currently unachievable. Getting RA under control to give the patients a life without excessive joint pain is not. To achieve remission of RA, a multitude of protocols have been developed and tested over the years. In this section, we discuss a few protocols that are currently widely used.

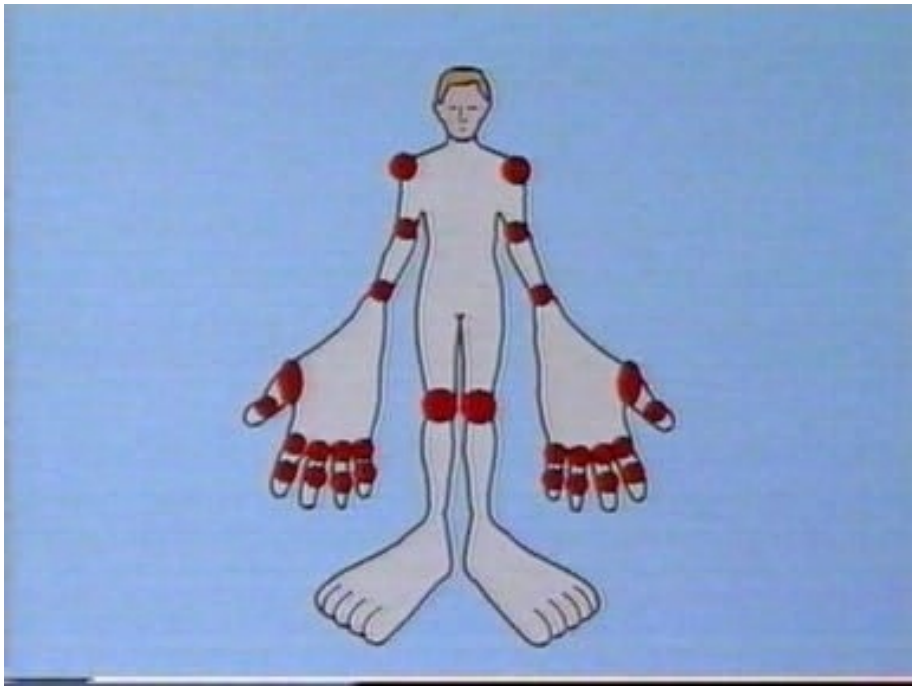


FIGURE 1: DAS28 joint form[19]

2.3.1 T2T

Two very well reviewed protocols are based on the treat to target (T2T) Principle[15]. The basis of T2T is setting a target and constructing a set of rules and restrictions to reach this target. For RA, this means that we require protocol with the target of reaching remission, meaning that the symptoms and pain are barely noticeable which results in a life without noticeable inconveniences from RA and the progression of joint damage is halted. The standard chosen for remission is the DAS28 score of below 2.6. This is a huge improvement over the routine care that was the original treatment strategy for RA. This routine care was based on the individual clinical judgement of a physician looking at the current state of a patient, adding or removing medication in terms to improve the circumstances as quickly as possible. While this was effective for some patients, the general effectiveness of T2T is much higher.

Two T2T protocols that represent state-of-the-art treatment protocols are discussed in the next section that are widely used and that form the basis of our research.

Step-up approach starting with methotrexate monotherapy

The first protocol we look at is a step-up approach starting with methotrexate monotherapy[13]. There are set evaluation moments on which the medication will be adjusted depending on the current DAS28 score with the goal of reducing this score to a value below 2.6. These evaluation moments are in week 0, 8, 12, 20, 24, 36, 52 and every 3 months thereafter.

The treatment is started in week 0 with 15mg methotrexate (MTX) per week with folic acid on the second day after MTX. In case of insufficient response, DAS28 is still >=2.6, during the next evaluation moment 8 weeks later, the MTX dosage will be increased to 25mg/week. During the next evaluation moment sulfasalazine (SSZ) 2000 mg/day will be added, which is increased to 3000 mg/day in week 20. Patients who have a Das28 score of 3.2 or higher in week 24 are given an anti-tumor necrosis factor-α treatment and

the SSZ is replaced by subcutaneous administration of 40mg adalimumab every 2 weeks, which is increased to 40 mg/week in week 36. If the DAS28 score is 3.2 or above in week 52, adalimumab will be exchanged for 50mg of etanercept each week. After this, every 3 months the patients will return for a check-up to evaluate how the treatment is affecting the DAS28 score and how it is affecting the patient.

If at any stage remission (DAS28<=2.6) is reached, the current medication is held constant. In case of a sustained remission of 6 months, the medication will gradually be reduced and will eventually be discontinued. If a flare happens, DAS28>=2.6, the last effective medication will be restarted, and the protocol will continue from that point onwards.

Deviations from this protocol are allowed where necessary, for example when the patient is allergic to or shows contraindications for a specific type of medication.

Initial disease-modifying antirheumatic drug combination approach

The second T2T protocol is an initial disease-modifying antirheumatic drug combination approach[13]. The set evaluation moments for this protocol are in months 0, 2, 4, 6 and once every three months after that.

Treatment is started with a combination therapy of 20 mg MTX each week and 200 mg HCQ twice a day with a single optional intramuscular triamcinolone injection with a maximum dosage of 120 mg. After one month, the MTX dosage will be increased to 25 mg/week. Like the first protocol, each patient is given folic acid on the second day after MTX usage.

If the DAS28 score has not reached the remission rate during the first evaluation moment, the dosage of MTX will be increased to 30 mg/week with a single optional intramuscular triamcinolone injection. If the DAS28 score shows moderate to high disease activity (DAS28>=3.2) a TNF inhibitor (adalimumab, etanercept or infliximab) will be added. In case of low disease activity (2.6<=DAS28<3.2), a decision between 2000-3000 mg of SSZ a day or an intramuscular triamcinolone injection must be made. This last step will repeat every subsequent evaluation moment.

Like the first protocol, if at any stage remission has been reached, the current medications are held constant. If the DAS28 score stays below 2.6 for 6 months, the medication will gradually be decreased until eventually they are discontinued. If at any stage a RA flare-up happens (the pain and thus the DAS28 score rises to above remission rate), the last effective medication will be restarted according to the protocol.

Deviations from this protocol are allowed where necessary, for example when the patient is allergic to or shows contraindications for a specific type of medication.

2.3.2 Assessment

During an evaluation moment, the current well-being of the patient is checked to see if medication should be increased, decreased or immediately discontinued. A full DAS 28 score is constructed by checking for swollenness and tenderness in the 28 joints, performing an ESR test and doing a VAS questionnaire. Potential side effects and possible allergic reactions are also checked for. Based on this information, the protocol will either be continued or altered as the rheumatologist deems it necessary.

2.3.3 Side effects

As is the case with every type of medication, there is a non-zero chance of developing side effects. The known side effects of every type of medication used in the protocols with a

probability of 1% or more of showing up are collected in table 2. Most common side effects are inconvenient at best and are potentially life-threatening at worst. At population level, most side effects are usually outweighed by the effectiveness of the medication against RA symptoms.

Medicine	>10%	1-10%
MTX	Swelling or inflammation of the mouth; Loss of appetite; Hairloss; Vomiting	Diarrhea; Headaches; Nausea; Pneumonia; Itching, rash, reddening of the skin
SSZ	Gastric distress; vomiting;	Loss of appetite; Headaches; Diarrhea;
Adalimumab	Cold symptoms; Headaches; Rash; Joint and muscle pain; Liver problems;	Systematic infections; Intestine infections; Psoriasis; Oral infections; Ear infections; Genital infections;
Etanercept	Headaches; Respiratory tract infections; Psoriasis; Skin infections; Cystitis;	Allergic reactions; Auto-antibody formation; Rash; Itchiness;
HCQ	-	Loss of appetite;
Triamcinolone	-	Infections; Headaches; Cataract;
Folic Acid	-	-

Table 2: Side-effects[7]

3 Research Questions

Developing a new treatment protocol for an illness requires extensive work, discussion, and guesswork, which is then finalized by writing up a treatment study that must be committed to and approved by the ethics commissions. Only then, the real study can be conducted to see if the proposed treatment are successful or should be adapted further. To improve this workflow, this research aims to develop a tool that could predict the outcome of newly constructed protocols before clinical trials is conducted. This result could be taken into consideration during the discussion stage of a protocol and should not be used as deciding factor.

3.1 Is it possible to model the treatment of RA?

3.1.1 Motivation

Conducting a clinical trial to test the effectiveness of a new protocol requires careful planning, ethical approval and a cohort of patients that all agree to participate in the trial. This does take time, even if this trial only changes an existing trial by changing the order in which medication is used. Experimenting with protocols in the development stage could show strengths and weaknesses such that the protocol could be improved on before a real trial is conducted. Another reason for this tool is to experiment with medication that is currently seen as unethical. Using too much of a specific medicine or medicine combinations that are currently frowned upon can be tested and used in further discussions in order to develop potential protocols that would otherwise not be created. Resulting in potential protocols that are perfectly ethical, while the separate steps originally were not.

3.1.2 Patient model

A model of a patient is required to test a protocol on. This model should contain all variables that are important for the illness, from the BMI to the RA factor and from the compliance to allergies. In order to create a patient model we have to answer a few questions.

- How do we represent the course of and the current severity of the illness?
  This research aims to simulate patients who have RA. To simulate this correctly, we have to keep track of the severity of the illness and how this progresses.
- How do we show the positive and negative effects of medication?
  I order to show the effectiveness of a medicine, positive and negative factors should be taken into account, and these factors should be visualised to the user.
- How do we represent side-effects caused by the medicine?
  Most types of medication have some kind of side effect. These side effects could vary from mild inconvenience, like headaches, to life-threatening complications that ultimately results in death. It is required that these side effects are visualised one way or another.
- How can we make the model expandable?
  Medical sciences are ever evolving and new illness factors could be discovered any moment. To prepare for new discoveries, we aim to create models that can be adapted and extended to include these new factors such that they are future-proof.

3.1.3 Treatment protocol model

In order to test the effectiveness of a treatment protocol, the protocol in question should be transformed into a model. This model must be able to interact with the patient model and could be used to dial in the workings of the patient model. In order to create a protocol model we have to answer a few questions.

- How is the medication represented?
  In order to allow for protocol model creation, medication representation must be designed, most importantly, how these types of medication should be described, and what notation should used.

- How are the steps of a protocol represented?
  Each treatment protocol consists of steps. How each step is represented is important for the design and consistency of each protocol model.

- How do we model the time that passes during simulations?
  UPPAAL has clocks build in and is a major deciding factor for using this tool as basis. How the time will be divided has yet to be decided, as hours, days, weeks, and month are all viable options.

- How do we implement transitions of protocol steps?
  Once the time and steps have been implemented, we require a way to transition from step to step during a simulation based on the condition of the patient and the time.

- How can we personalize each protocol?
  Initially, each protocol is created for a general patient. A way to personalise each protocol would be nice to have. This personalization could be done by hand or by setting some flags in the patient model.

3.1.4 Evaluation treatment protocol model

Evaluating the workings of a protocol is the reason why the patient and protocol models are developed. In order to correctly evaluate a protocol model we have to answer a few questions.

- What data will be evaluated by the simulation?
  In order to evaluate the severity and course of the RA, the most important factors that will be compared to the literature must be decided.

- How will this data be extracted from these models?
  The data that we require for the evaluation must be extracted from our models at a specific time. We must decide a way to get the data from our models at specified intervals.

- How will the data be represented?
  In order to efficiently compare the data from our models with data that is available in the literature, this data must be represented in a comparable format.

3.2 Is it possible for medical professionals to create and evaluate models?

3.2.1 Motivation

The goal of this research is to make it possible to experiment with protocols without the involvement of patients to rapidly test new and more effective protocols. This is only possible if the professionals developing such protocols can transform these protocols into models and evaluate the results.

3.2.2 Prototype tool for creating a protocol model

In order to create a tool that generates the protocol models we have to answer a few questions.

- What information is required to create a protocol?
  Once we have created a protocol model, we know what information is required to create protocol models from scratch.
- How is a protocol model structured?
  Once we have crafted a model by hand and discovered what data is required to generate new steps, we could reformat the protocol models such that removing, editing, and adding protocol steps will be possible.
- How can we load existing protocols?
  The tool should be able to load protocols created manually and created by the tool.
- How do we edit these protocols?
  Once we have taken care of the structure and required data, and we know how to load models, we have to develop methods to edit the model.
- How do we create queries to evaluate the protocol?
  Not every protocol requires the same manner of evaluation. Some might require the severity of the illness after a number of weeks, others might require the number of weeks until a specified severity has been reached. We must develop a way to add and edit the queries in a protocol.
- Can we export these protocols?
  The models created and evaluated by our tool should be exportable to save for later discussion.
- What commands are required?
  In order to comply to all the questions, commands to perform these tasks must be implemented.

3.2.3 Linking the tool to UPPAAL

While creating a protocol model is important, it is also important that a professional can run the simulations. A few questions to keep in mind are as follows.

- How do we connect our tool with UPPAAL?
  The verification tool in UPPAAL will be used to evaluate every query present in our models. This tool must be connected to our tool by using the API that is provided by UPPAAL.

- How is the length of a simulation decided?
  A simulation should not run an infinite amount of time. A way to limit the simulation would be a nice to have. This could be in the form of a maximum duration in protocol in terms of days or weeks, or in terms of runtime limitation.

- How many runtime variables should be changeable by the professional?
  In theory it would be possible to edit the entire model via a tool. However, it would be unwise to let the user edit the workings of the human model, because unexpected and untested results might show up that is the opposite of the potential verified human model developed in this and future research. Thus, it is important to decide which variables are safe to edit and which variables should be set in stone.

3.2.4 Making the tool user friendly

In order to make the tool user-friendly, we must conduct user tests. The result of such a test shows what should be changed and what works.

4 Goals and approach

In this research, we aim to create a tool which can predict the effectiveness of existing and newly created Rheumatoid Arthritis protocols. This should be done by creating a protocol model, and loading this into a program where you can set prediction queries. This program aims to return the probabilities of a query in the form of a range, and it displays the general course of specific values, for example, the DAS28 score.

4.1 models

The first step towards reaching our goal is by manually creating the required models in the UPPAAL language, which is explained in detail in section 5. First, we research which factors play a crucial role in the development of RA, as well as which result data is important to generate with simulations. Based on this data, we create a input-output analysis and decide which data should be contained in the patient model and which data should be contained in the protocol model.

We developed a patient model combined with a protocol model that contains no medication first. This patient model contains the important factors that play a key role in the development of RA, for example, BMI, General health, susceptibility to side effects, age, sex, and compliance to follow a protocol.

Once a patient model has been created, a model for a protocol must be created. This model depicts the rules for when to use which medications. This is the type of model that can be created by a medical professional, in order to test new protocols.

The details of this implementation can be read in sections 7.1 and 7.2.

4.2 proof of concept

Nothing is gained from purely creating these models without verifying if the probabilities and statistical data generated by these models and queries are according to data available in the literature. Once a model is created, we construct queries to generate data that is available in the literature. For example, what is the probability for a patient to reach remission after 6 months. UPPAAL then runs hundreds of instances, after which it returns an interval in which the probability lies. This probability could be compared to the numbers collected by testing a protocol on a real cohort, however, this is outside the scope of this research.

4.3 Runtime application

Once we have models that are generating a plausible result, the next step is creating an app which makes it easy for the user to run queries. This application requires three components: a patient model, a protocol model, and queries. The models are loaded into the application by selecting them. The queries can be created by the user by selecting a basic query for drop-down menu and supplying some input, for example selecting "reaching remission within X weeks", followed by the input "26" to create the query "what is the probability that the patient reach remission at least once within 26 weeks?". There should also be an option to create your own queries, but most used queries should be easily accessible without the need for knowing the UPPAAL-specific logic language.

4.4 Model creation environment

The models we have used for our other goals are all manually constructed within UPPAAL. We aim to create an environment in which a medical professional must be able to create a protocol model without much or any knowledge about UPPAAL. This environment should be able to load existing models, edit them and export them, as well as create new models from scratch. These generated models should be compliant with the UPPAAL language and must be usable with the runtime application.

4.5 Usability

The applications we have created must be understandable for every medical professional with a small number of instructions and little documentation. Creating a model must be intuitive and running queries on these should only require a few logical steps. The usability can be ensured by having a test-panel use our tool, requesting feedback, and lastly, use the feedback to improve the tool.

5 UPPAAL

UPPAAL is an integrated tool environment for modeling, simulation and verification of real-time systems, developed jointly by the Uppsala university in Sweden and Aalborg University in Denmark; The combination of these university names UPP and AAL form the name of this tool, UPPAAL.

5.1 What is UPPAAL?

UPPAAL is a tool in which one can create models for systems as a collection of nondeterministic processes with finite control structures and real-valued clocks, communicating through channels or shared variables. This means it is possible to create multiple systems dependent on each other while they all evolve over time. The evolution of variables in these systems can be constructed in the form of differential equations or the evolution can be modeled as a series of uncertainties. The individual systems can communicate with each other due to the channels, shared variables and synchronized clocks. The clocks and synchronization are typical examples why UPPAAL is chosen for time-critical applications, like communication protocols and real-time controllers.
UPPAAL consists of three main parts: A description language, a simulator and a model checker.

Descriptive language UPPAAL uses a descriptive language to represent the models in a non-deterministic way. The user is able to define data types, and to describe system behaviour as networks of automaton extended with clocks and data variables. UPPAAL contains a graphical editor that makes it easy to create this language in the form of a drag and drop system. This reduces the amount of errors created by incorrect written code and makes the model better understandable with the aid of visuals, compared to just text.

Simulator A tool which enables the user to dynamically execute the models. This can be done manually or by letting the tool make random choices. This enables the user to test models for deadlocks and other mistakes before actually using the model checker.

Model checker UPPAAL contains a stochastic model checker that is able to solve queries containing invariants and reachability problems by exploring the state-space of a system. The result of each query is either given as a single number if it stays constant, or, when a stochastic choice is available, as a range or percentage after running hundreds and thousand of runs. It is important for our research that this model checker is able to handle stochastic choices, as life and medication are not definable in static variables. There is a probability that a medicine works as intended, as well as a probability that it does not work and many side effects are introduced. This SMC is also available as a standalone tool called Verifyta. This is what is later used by our tool.

5.2 Why UPPAAL?

A multitude of different model checkers exist that we could have used for this research, but we have chosen UPPAAL. There are three reasons for why we have chosen this environment over others.

Editor The graphical model builder is an environment that makes it easy to create and edit models and test them with the use of the simulator. This makes it great to build, update and renew the models on the fly, resulting in a easy to use environment to development, test and improve the perfect model. This ease of use is one of the main reasons why we decided to use UPPAAL over other potential environments.

Modularity Model checkers usually accept one system with models and multiple queries to evaluate for that specific system. The models within this system interact with a mechanism like a shared variable or gates. How well this works is dependent on the model checker. UPPAAL is different in this regard. In UPPAAL, models can communicate and synchronise by either shared variables or broadcast channels. Due to this, it is fairly easy to replace side models with new models, in our case different protocols, that interact with the main model, in our case the Patient model. Rapidly changing the protocol model while keeping the main model intact is a main feature for this research.

Clocks Real world protocols are defined in terms of time, often in days or weeks. During this time, the severeness of RA changes. In UPPAAL, time can be modeled by real-time clocks with continuous values that can be compared to constants to form constraints and can be reset when deemed necessary. Due to this mechanism, we are often able to use differential equations that update the severeness (DAS28) for during of multiple clock-units. Because UPPAAL uses continuous clocks instead of clock ticks, this evolution happens in a continuous fashion.

Uncertainties Real world systems are rarely completely deterministic and uncertainties are almost always part of the full picture. UPPAAL supports this by adding branch points with probability weights, in addition to the use of exponential and uniform distributions.

6 Existing work

No research has been found on modeling the course of RA in UPPAAL or in other modelling frameworks during the execution of this thesis. Other projects involving UPPAAL an the medical field do exist.

6.1 A methodology for evaluating tooth wear monitoring using timed automata modelling

The reduction of dental hard tissue is what we call tooth wear. Multiple techniques with varying costs can be used to reduce this wear, but which technique is the best for specific cases? This research aimed to answer the question if a yearly or if once-in-five-years counselling/monitoring protocol yield better outcome results in terms of tooth wear. To answer this question, UPPAAL is used to represent the course of the tooth wear including only the important factors. With the models created, queries for the number of "good years", and costs are constructed. After which the variables of how often counselling occurs are changed between simulations of these models. The models have predicted that yearly counselling and monitoring yields the best outcomes which can be used to make decisions on which protocol to use in the near future[22].

6.2 Modeling Diagnostic Strategies to Manage Toxic Adverse Events following Cancer Immunotherapy

Although immunotherapy is a viable treatment against certain types of cancer, approximately 10% of all patients experience immune-related adverse events (irAEs). Early detection and treatment of these irEAs prevents progression to sever stages. This can be achieved by routine testing for irAEs and, when it is detected, the immunotherapy might be required to be put on halt. However, when false-positive test outcomes result in a halt on medication, it affects the treatment outcomes negatively. This research has used UPPAAL to model differences in test accuracy for irAEs by varying the specificity and sensitivity of these tests. The models are callibrated with the use of 248 non-small-cell lung cancer patients. This research concluded that increasing the specificity of testing decreases the discontinuation of treatment due to suspicion of irAEs by reducing false-positive test outcomes[16].

6.3 Useful takeaways

The most important takeaways from these papers is the understanding that it should be possible to simulate medical situations in such a way that the outcomes can be used to increase the effectiveness of treatments. The UPPAAL models created for these studies are used as base ideas and the concepts in these models have inspired us where possible.

7 Proof-Of-Concept models

The development of the models started with discovering which data is important for the development of RA, how the severity of the illness is represented and how treatment protocols in general are constructed.

7.1 Creation of Patient model

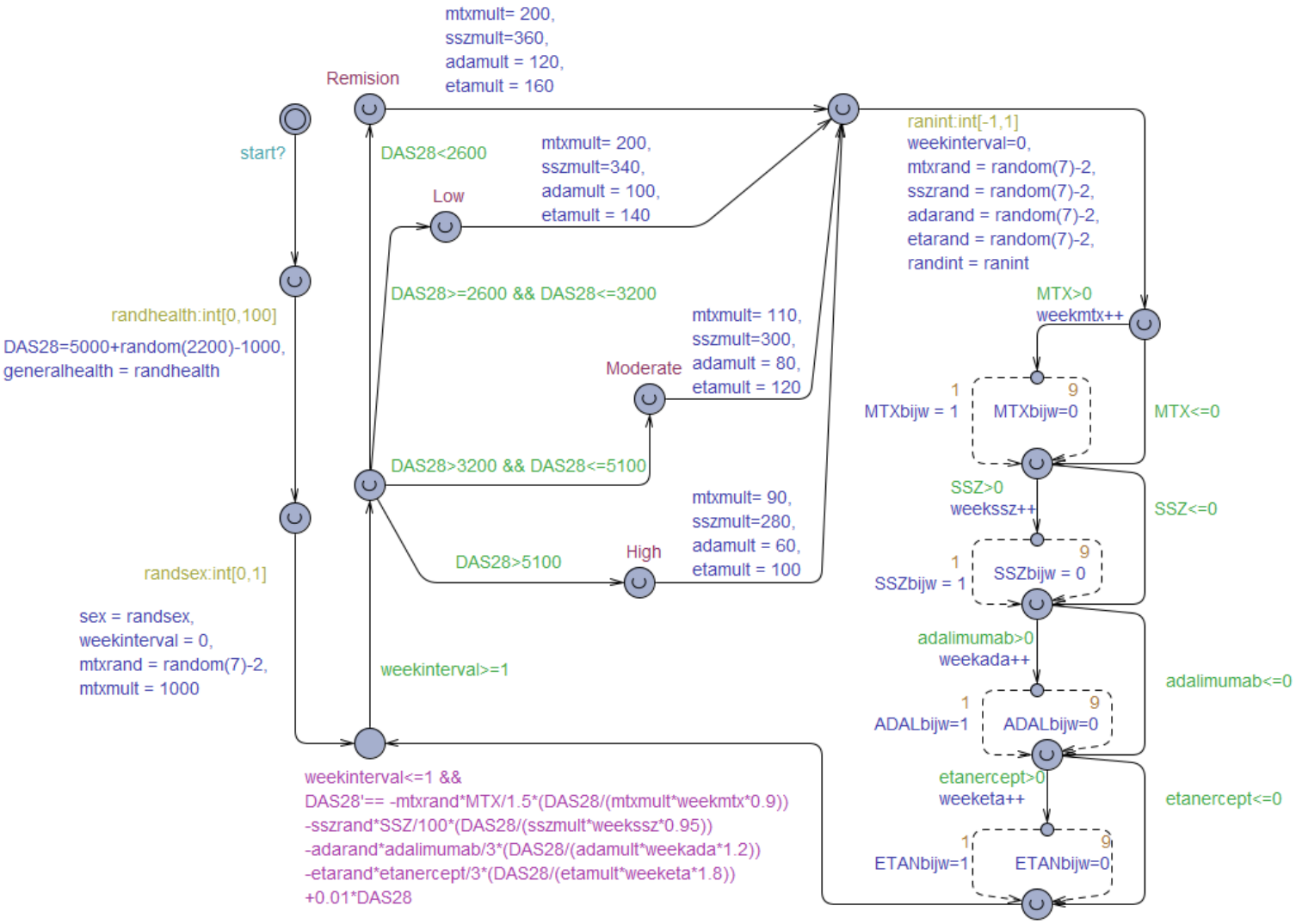


FIGURE 2: Patient model, see Appendix A for further explanation.

The most important model of this research is the patient model. This patient model simulates the evolution of data required to calculate the severity of RA and, at later stages, variables in order to generate the probabilities of developing side effects and how well the medication works.

7.1.1 Requirements

Das28 The severity of RA is depicted by the Disease Activity Score with the evaluation of 28 joints, Das28. This is the value on which most decisions in a protocol are based and is the most important variable in this model.

Sex It has been indicated that the sex of a patient plays a major role in the development of RA and the evolution of the Das28 score.

General health The general health is an important factor to keep in mind when evaluating the evolution of RA. If a patient has a multitude of health conditions, RA might not evolve in the same way compared to when no other health conditions are present.

Medication effect In order to visualise the effect that medicine has on the DAS28 score, we require the medication effect. Each type of medication has their own healing effect and side effect which are required to create a good patient model.

7.1.2 Design choices

The Model can be divided in 4 stages. Initialization, evolution of the DAS28 score, the effectiveness of medication, and calculating side-effects. We first discuss the important variables and why these formats have been chosen. After this we discuss the four main parts of the patient model. Finally, the entire model is shown.

Das28 Das28 is initialized as a clock in our model to be able to use in a differential equation. Because doubles are not fully supported by UPPAAL, Das28 is multiplied by 1000 such that a score of 4.5 is initialized as 4500.

General health General health is represented by an integer value between 0 and 100. Because this value is not yet used by the model, this range is subject to change when needed.

Sex The sex of a patient is 0 or 1, where 0 defines the male sex and 1 defines the female sex.

Weekinterval This is a clock that can be reset to 0 whenever a specific amount of time has passed.

Medicinerand In order to have some randomness in the DAS28 calculations, we require a randomness factor for the effectiveness of every type of medication we use. This multiplier range is at currently arbitrarily chosen between -2 and 5. A negative value means that the medication increases the DAS28 score, while a positive number decreases the DAS28 score

Medicinemult This is the multiplier of how effective a type of medicine is right at the beginning. The lower this value, the more effect it has. The initial value is high in order to make sure the first iteration passes without unexpected behaviour that could get introduced by a protocol model.

7.1.3 Initialisation

In this step, we set the variables for a patient.

Das28 Tthe Das28 score is initialized to a random moderate to high score, 4.0 to 6.2, which corresponds with 4000 and 6200. In order to generate a value between 4000 and 6200 in a fair way, we generate a value between 0 and 2200 which we add to 5000, after which we subtract 1000.

Generalhealth We generate a random value between 0 and 100.

Sex This value is randomly set to 0 or 1.

Weekinterval This is a clock that we initialize to 0.

Mtxrand This multiplier range is at currently arbitrarily chosen between -2 and 5.

Mtxmult The initial value is set to 1000 in order to make sure the first iteration passes without unexpected behaviour that could get introduced by a protocol model.

7.1.4 Evolution

After the initialization of all variables and clocks, we end up in the node where the DAS28 score gets recalculated with a large formula. This formula takes the DAS28 score and reduces this value in small intervals over the span of 1 time unit by each type of medication. The MTX part of the formula is as follows:
$-mtxrand \times MTX \div 1.5 \times (DAS28 \div (mtxmult \times weekmtx \times 0.9))$

- Multiply the MTX randomness factor with the current amount of MTX that is used by the patient and divide it by 1.5. This gives the effectiveness value of the MTX medication when the Das28 score is a high number.
- The assumption has been made that the higher the Das28 score is, the better the medication works, however, the longer you use a medication, the lesser the effect of the medication is. This is why we divide the Das28 score by the MTX multiplier by the number of weeks the medication is in use. We also multiply this divider by 0.9 after tuning this formula to the results found in the Step-up approach starting with methotrexate monotherapy paper.
- These two values are multiplied and this result is subtracted to the current Das28 score.
- This formula is applied to each medication used. The values 1.5 and 0.9 are specific for MTX and had to be tuned for each type of medication. This way, it is possible to tune the effect each type of medication to represent their real world counterpart.

7.1.5 Effectiveness of medication

Once the week is over and the DAS28 score is updated, the multiplier for each type of medication is set. This is done by choosing one of four different paths based on the severeness of the disease. Here we assume that medication will have a larger effect the higher the DAS28 score is, because that is a trend we see in the literature. During the first weeks of a protocol, the DAS28 score drops relatively fast, while at later weeks this score changes less drastically[13]. We then generate the random medicine effectiveness value between -2 and 5.

7.1.6 Side effects

The last step is calculating the probability of side effects showing up. There is a 10% chance of a side effect showing up and a 90% chance of nothing happening at all for MTX. This is done for every type of medication to set a flag if a side effect has occurred and the probabilities of side effects are taken from farmacotherapeutischkompas[7]. This is a feature that has not been further implemented yet and should be improved in further versions.

7.2 Protocol in UPPAAL

For the proof of concept protocol model we have adapted the step-up with MTX protocol[13]. We start by setting the initial medication, in this case 15mg of MTX for 8 number of weeks. During this time, the DAS28 score is updated by the patient model. Once the specified number of weeks have passed and the DAS28 score is still above the remission rate of 2.6, the medication is changed according to the protocol, in this case to 25mg of MTX for another 4 weeks. The same loop is invoked once the time has passed, and the medication is changed if the DAS28 score remains above 2.6.

If the DAS28 score is less than 2.6, we keep the medication stable for 8 weeks. If, during these 8 weeks, the DAS28 score rises to above 2.7, we resume the original protocol, while we slowly decrease the medication once the 8 weeks are over and the DAS28 score is still below 2.7.

Every week we remove some of the medication. We start by removing some etanercept until the medication is no longer in use, then we decrease the amount of Adalimumab, SSZ and finally MTX. If at any point the DAS28score rises to values of above 2.6, the last working medication is resumed for 8 weeks. If the patient is stable, we resume decreasing medicine, while the original protocol is resumed when the DAS28 score rises above 2.7 during these 8 weeks.

7.2.1 UPPAAL model

- The weekdeadline variable is initialized as 8 in the model declaration.
- A method called medUpdate is defined in the model declaration that changed the medication and new week deadline according to the step-up MTX protocol.
- broadcast a start signal to all models and set the initial medication to 15mg of MTX. This start also resets all week clocks to 0.
- Once a clock called weekmed reaches the value of the weekdeadline, in this case 8, a transition is forced. If the Das28 score is above 2600, the function medUpdate is called. Else, the protocol goes to wait state.
- The model stays in this sate for 8 weeks while the Das28 score remains under the 2700. If the Das28 score rises above 2700, the medUpdate method is called, else the medication is reduced.
- The last added medication should be reduced by a set amount until noting remains. For example a patient uses 25 MTX and 4000 SSZ, the method reduces SSZ by 100 each week until there is 0 SSZ in use. Then MTX is reduced by 1 each week. Once the Das28 score rises, the last changed medication should be reverted. Once a score

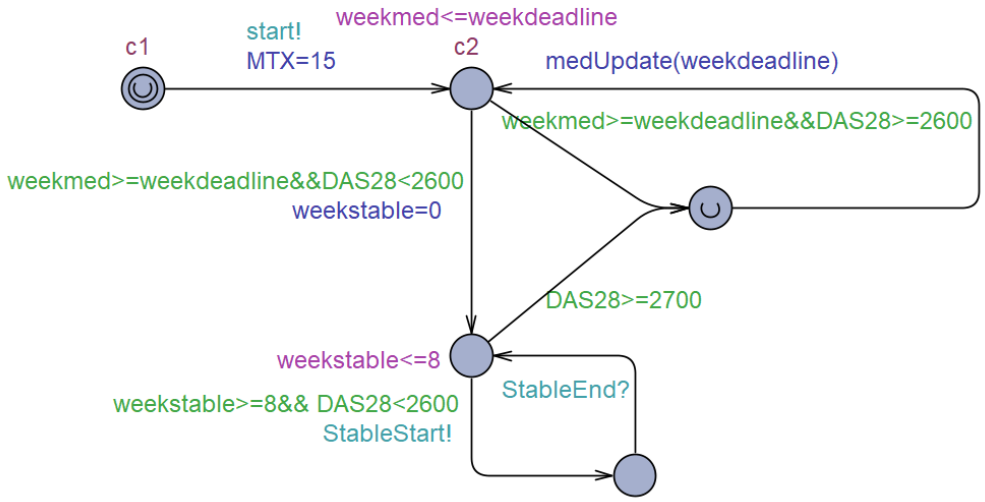


FIGURE 3: Protocol model

of 2600 or higher is achieved, no changes are made for an additional 8 week, except for when the score rises above 2700, then the medUpdate method is called.

The medUpdate method works as follows:

```
void medUpdate(int weekdl)
{
if (weekdl ==8){
weekmed = 8;
MTX=25;
weekdeadline = 12;
weekstable=0;
}
else if(weekdl==12){
...
```

7.3 Queries

The queries currently made for this model are quite simple, but they give a good overview of how good our models represent reality.

```
Pr[<=26] (<> Patient.Remission)
Pr[<=52] (<> Patient.Remission)
Pr[<=26] (<> Patient.Low)
simulate [<=50;1] {DAS28, MTX*100}
```

- Pr[<=weeknum] calculates the probability of the remaining query after the number of weeks specified as weeknum have passed

- <> returns true if the statement after it is eventually evaluated as true. For example <> Patient.Remission. This return true if the path walked by the simulation eventually reaches the state Remission.

- simulate [<=time;runs] is a query that makes UPPAAL run simulations. This simulation walks through the models for the duration of time, and the simulation happens as often as the runs depict.

- {var1, var2} Depicts which variables are recorded during the duration of a simulation.

The query "Pr[<=26] (<> Patient.Remission)" calculates the probability of a patient reaching remission at least once during the first 26 weeks. UPPAAL runs this query multiple time in succession until the confidence level reaches 0.95. This usually happens after about 400 runs, but the number of runs can be manually set and the confidence level can be increased if so desired at a later stage.

The query starting with "simulate [<=50;1] {DAS28, MTX*100}" means that there will be 10 simulations that run for 70 time units with the values of the variables DAS28 and MTX stored in a graph.

These two types of queries are the basis of verifying these models against the real-world values, which is explained further in section 8.

8 Proof-of-concept results

To be able to validate the models we have generated, we must compare the data that is obtained by our prototype patient and protocol model with the data that is available in the literature. We start by discussing the results that are documented, followed by what our model returns as result and, finally, comparing these two results to verify if our model works as intended.

8.1 Literature

Data for the progress of the DAS28 of the DREAM cohort that followed the "Step-up approach starting with methotrexate monotherapy" protocol is recorded in detail by Steunebrink et.all.[13]. Figure 4 shows the evolution of the mean DAS28 over time. This graph shows that the DAS28 decreases over time, but the less sever the disease activity is, the slower it declines. The second graph from Steunebrink et.all. that is important is shown in figure 5. This graph shows the proportion of patients that have reached their first remission at any given point. At 26 weeks, 62 out of 128 patients have reached their first remission, which is 48.4%. After 52 weeks, 92 out of 128 patients have reached remission, which is 71.9%. Both of these graphs shown also include a second cohort as was present in the original research, which could be used for future reference when this second protocol is modelled with our tools.

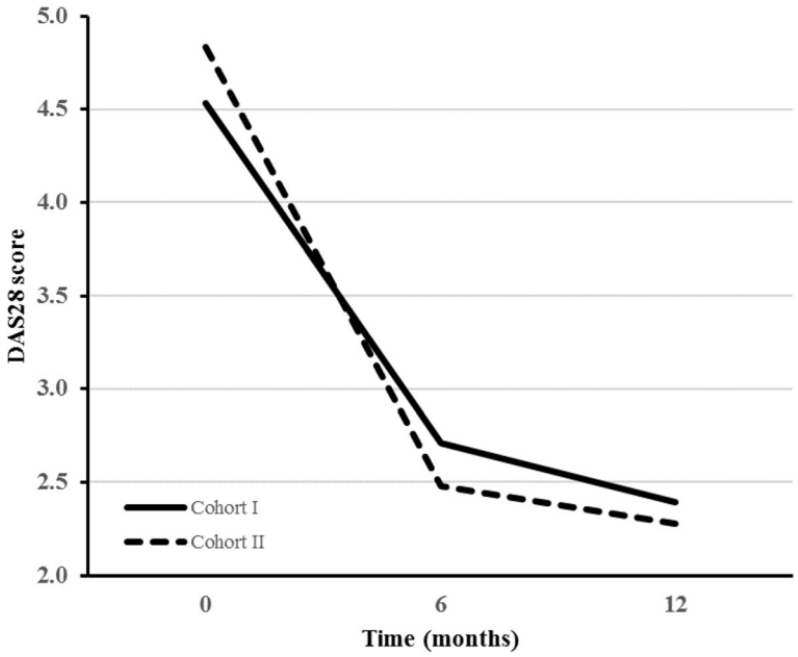


FIGURE 4: Decrease in mean 28-joint Disease Activity Score, the solid line is for the cohort that uses the protocol we have modeled[13]

.

8.2 UPPAAL

The proof of concept model I have explained in detail in the sections 7.1 and 7.2 is used in combination with the queries as shown in section 7.3. The data we retrieved from the

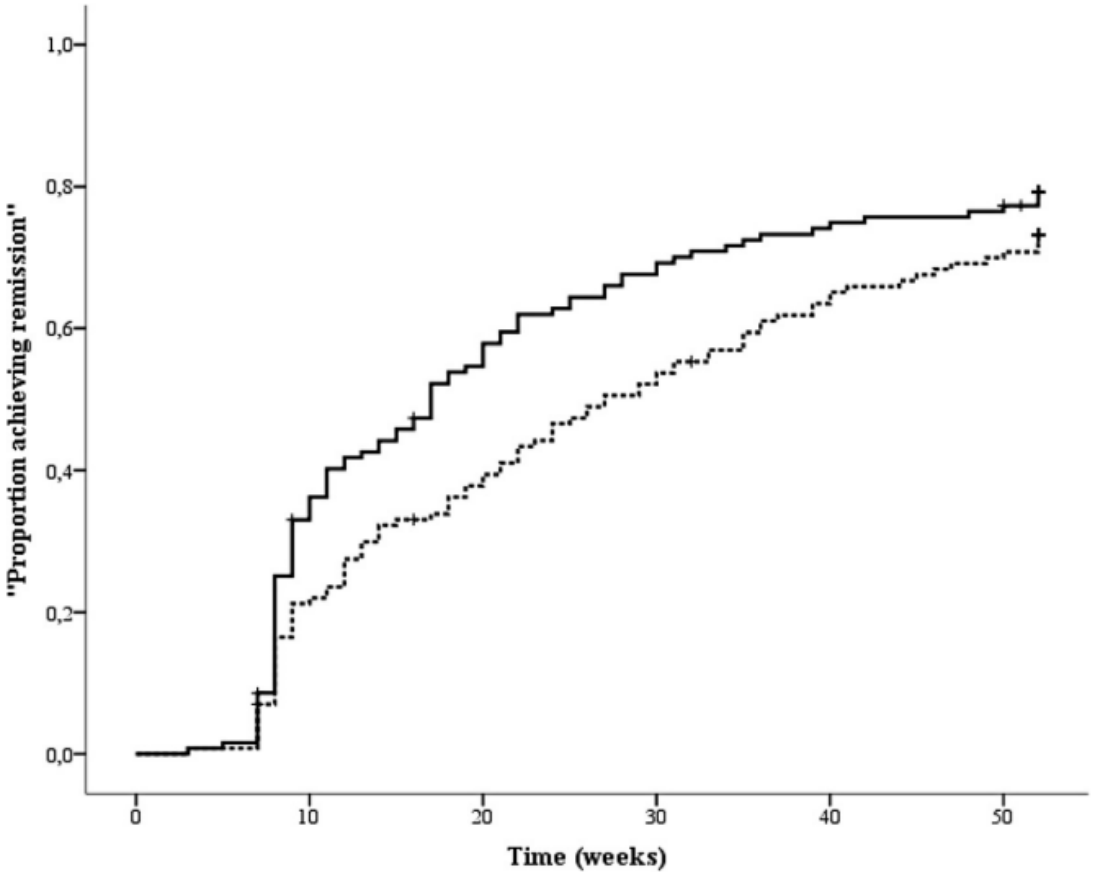


FIGURE 5: Kaplan-Meier curves for time until first remission was reached. The dashed line represents the cohort that used the same protocol as our protocol model.[13]

literature contains percentages of patients that reached their first remission at latest in X weeks, we have picked two of these moments, namely 26 and 52 weeks, to generate predictive data with our model. This is done to support this proof of concept and in the continuation of this research we generate the entire Kaplan-Meier curves. The proportion of patients that reached remission in 26 and 52 weeks is shown in Figure 6 and Figure 7 respectively with a 95% confidence interval. A simulation of the evolution of the DAS28 for one random patient is generated and shown in Figure 8.

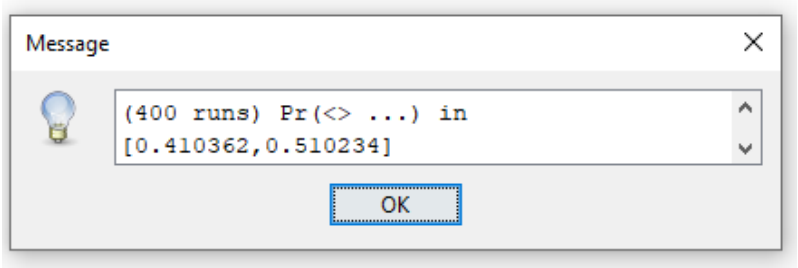


FIGURE 6: MTX protocol first remission in 26 weeks

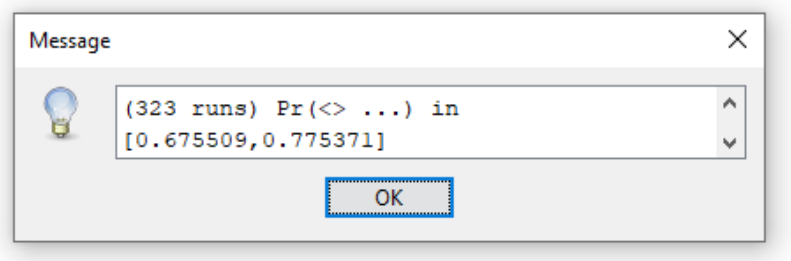


FIGURE 7: MTX protocol first remission in 52 weeks

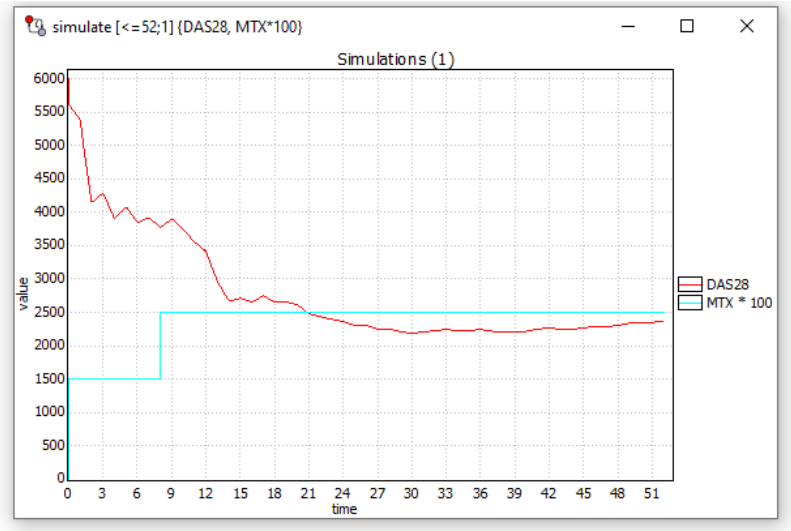


FIGURE 8: Das28 evolution

8.3 When is it verified?

In order to verify that the generated data is in line with reality, we must define our verification criteria. For this proof of concept, we have used the default settings in UPPAAL where a 95% confidence interval is generated after 300 to 500 runs. We call the generated data verified when the real values are contained within these intervals. In the continuation of this research, we will adjust the variables and the number of runs to generate a smaller interval, resulting in a more credible prediction.

8.4 Comparison with literature

Comparing the data that is available in the literature shows that our results are not that far off from the values documented in the literature, as can be seen in figure 5[13]. In our simulations, we have a 95% confidence interval that tells us that between 41.0% and 51.0% of all patients will reach their first remission in 26 weeks. The literature tells us that in a cohort of 128 patients, 62 reached remission within those 26 weeks, which is 48.4% of all patients. This number lies within our confidence interval meaning that our model is able to represent a somewhat credible result. The same can be said about the probability of patients having reached their first remission within 52 weeks. Our models generates the 95% confidence interval 67.6% to 77.5% while the literature shows that 92 out of 128 patients have reached remission, which is 71.9%. This number lies withing the confidence interval, meaning that our model is able to represent this result in a credible manner.

While our model can reasonably predict the proportion of patients that have reached their first remission in 26 and 52 weeks, they must be tweaked and improved before these models are ready for future research, as the evolution of the DAS28 score currently evolves rapidly in a random manner without much logic. This can be seen when you look at what happens within one week in figure 8. In one week going from a DAS28 score of almost 5.5 to 4.2 is not likely to happen and further research is required to get the formula right. Besides this point, our model is currently only able to generate correct data for these two data points. We can only call the model credible if it is able to generate Kaplan-Meier curves that are in line with the ones in the literature, instead of purely these two moments.

9 Infrastructure

Our goal for this research is to end up with a tool which can be used by medical professionals to predict the effectiveness of newly constructed protocols. To achieve this, we have created a tool in which new protocols can be designed and tested. This tool consists of 3 main parts, as discussed earlier. The first part is the core, which is a predefined patient model created by the developers of the tool. The second part is a toolbox that allows medical professionals to create digital models of protocols, that can load existing models to change some parts and save these models to be used in the third part. This final part is the run-time too. This tool accepts a core and a protocol model and does calculations to predict the effectiveness of the protocol.

9.1 Core

The core of this entire research is the patient model. This model consists of a loop in which the DAS28 score, effectiveness of medication, side-effects and other health related variables evolve. This loop is currently defined as one week, but this can be changed to what the medical specialists deem necessary. Because this model consist of many variables and formulas, this model is provided by the developers and should not be edited by the user, however, variables such as the BMI, age, and sex should be configurable by the user. In order to update and improve this model over time, feedback given by medical professionals will be taken into account for future development and new models will be published when available. The newest stable version of the core is provided with the tools.

9.2 Toolbox

UPPAAL is a program that has a steep learning curve in order to use it efficiently. In order to eliminate this, a tool that has predefined building blocks is created. Protocols can be constructed with these simple building blocks, while the model that the toolbox generates is compatible with UPPAAL.

9.3 Runtime

This tool is where the calculations and simulations take place. When this tool is booted, the user must select a core and a protocol model. Once these are loaded, the tool shows a list of predefined queries that can be calculated. While the predefined queries should be sufficient to calculate most results, the user is also able to write their own queries. Once the user starts the simulations, the tool connects to UPPAAL and let it run the simulations. The results of these simulations are shown in the run-time tool. While this process can be done in UPPAAL, it makes sense to streamline this process in order to make this tool more user friendly. If, at any stage, more functionality is required, it can be added in the next revision of these tools.

10 Prototype models

As can be read in section 7.1, the variable at the basis of every component is the DAS28 score. This score is constructed from 4 different components, the TJC28, SJC28, ESR and GH. In the proof-of-concept models, we have crafted a formula that calculates the new DAS28 score based on the medications used. While this resulted in a model that generated results that are comparable with real world results, this formula was difficult to understand and adding new medications to it would not be a trivial task.

Besides this fact, calculating the exact DAS28 score is unfeasible with this formula as it would require in depth understanding of the evolution these four components. The exact effect of each type of medication is not documented based on each of these factors. One type of medication, for example, could decrease stiff and aching joints, while it increases the ESR value. This means that the DAS28 score will go down, however, the general health might become worse due to side effects.

Lastly, the DAS28 score is a relatively new construct, meaning that early RA research uses a different disease activity scale that is incompatible with the DAS28 score. For example, one such research is using a disability score and joint corrosion scale, which is not directly translatable to a DAS28 score.

10.1 Different way of thinking

These realizations mean that the proof-of-concept models are no longer usable in their current state and must be simplified and divided in a more systematic way. As stated earlier, calculating the exact DAS28 score is unrealistic and could be research of its own. To simplify this, we have created four different severity states in the same way as the DAS28 score is usually divided in the literature[13][4], High, Moderate, Low and Remission. The DAS28 score is often already divided in these four states when discussed, meaning that protocols using the DAS28 score should be easily adaptable to this new form. While this approach is less refined compared to using the real DAS28 score, it has the advantage that results of older protocols can be adapted to be represented in this form.

10.2 Patient model

This model has been reworked to incorporate the new design decisions of having severity states rather than the DAS28 score. The model has 5 main states, high, moderate, low, remission and stopped. A simplified version of the model can be seen in figure 9.

The disease activity states are divided as follows:

High: DAS28>5.1
Moderate: 3.2<DAS29<=5.1
Low: 2.6<=DAS28<=3.2
Remission: DAS28<2.6

- A patient is placed in one of the 4 severity states at the beginning of the simulation based on their Das28 score or different notation of severity.
- From each severity state, a patient could remain in that state, move one state up or down every set time span of a week. A patient with high disease activity could remain in High or could transition to moderate disease activity. Similarly, a patient in low disease activity could remain in low or could transition to either moderate disease activity or remission.

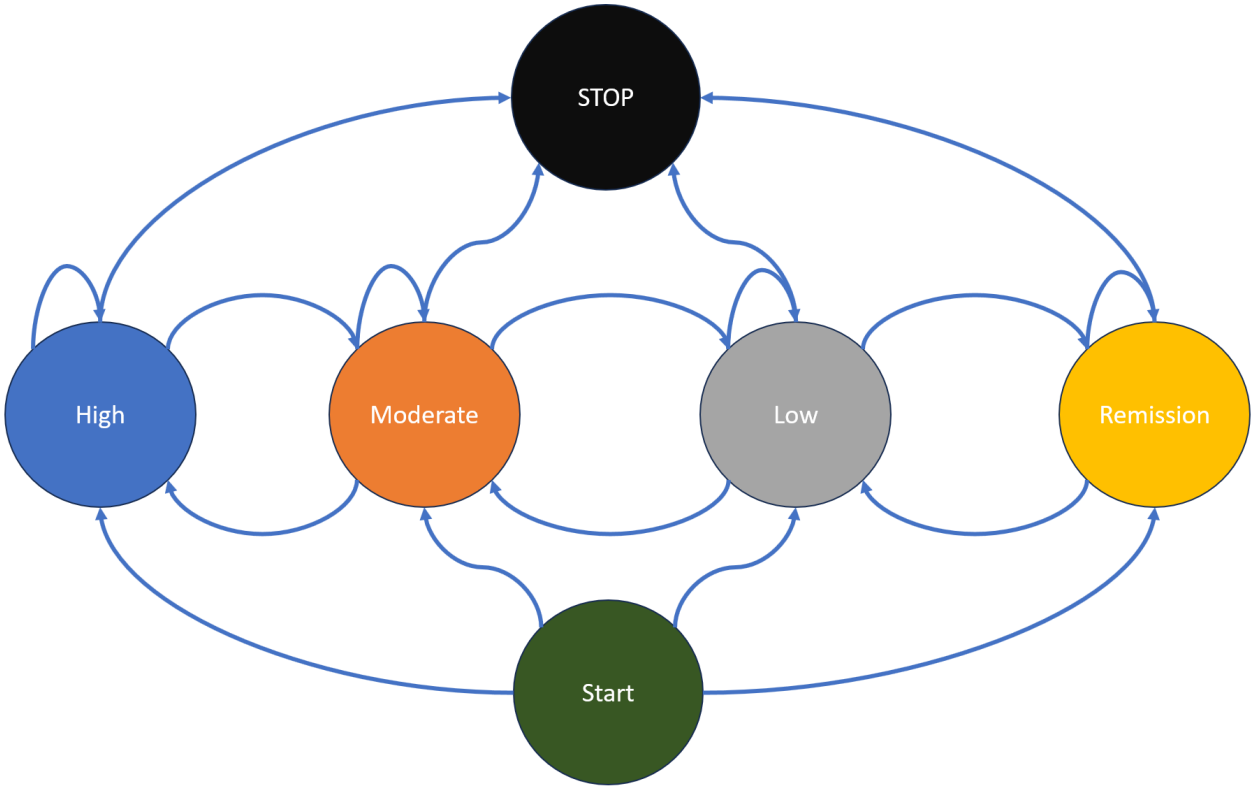


FIGURE 9: Concept patient model

- The calculations of the weights for going to a new state are no longer done by the models, rather, these calculations must be done by a professional who knows the ins and outs of a type of medication.
- The weight for each transition must be given for each medical package. This is done by the developer for the known medication packages, however, the user is responsible for setting these weights for newly added medication packages. For example, a patient with moderate severity, three values are needed. One with the weight of going to high severity, one with the weight of staying moderate severity and one with the weight of going to low severity.
- We have added a stop state for when a patient stops with the protocol due to ineffectiveness of the medication, severe side effects or other reasons. This requires another weight that must be added by the developer or user, which is only applied during the first week of using this medication.

This new setup means that adding a new type of medication is trivial once the twelve required values are known, which makes it possible for a plug and play system to be created.

These requirements result in a vastly different model compared to the proof-of-concept model 2, as the new version has all the calculations in methods. This new model is shown in an abstract form in figure 9, the full version can be found here [14].

10.3 Protocol models

The proof-of-concept protocols were used to edit the amount of medicine that a patient uses. For example, it could add 2000mg of Adalimumab each week and was used in the formula.

This approach meant that a variable must be added for every type of medicine, making it difficult to add new medicine. The new protocols are based on medicine packages. Instead of having 15mg of Methotrexate and 2000mg of Adalimumab, a patient uses medicine package 4. This allows for protocol generation and on the fly editing of these protocols.

The model used for Every protocol has the same idea as originally, the change in this model is mainly in the method and with the addition of a stop-state. this new model can be seen in figure 10

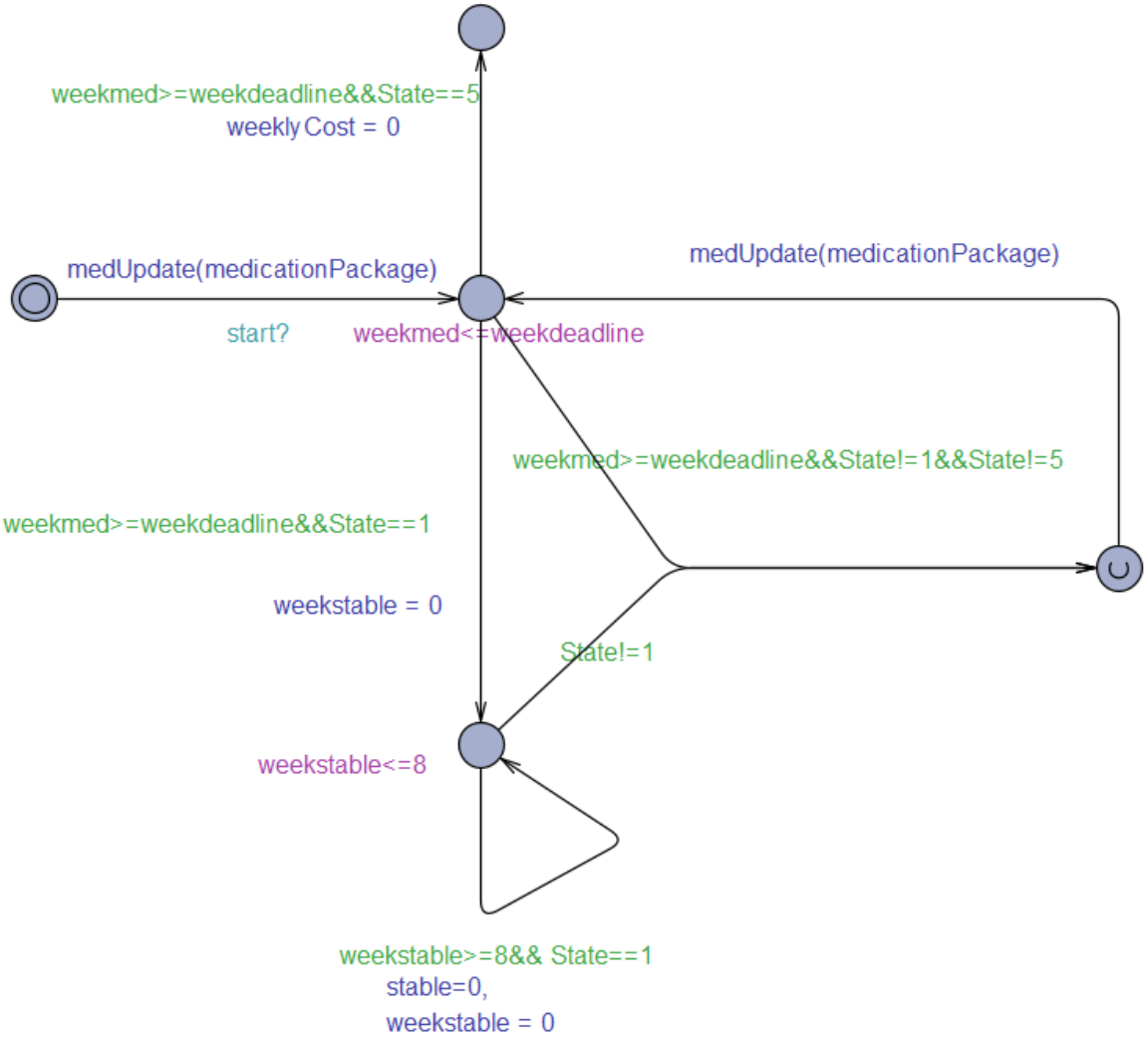


FIGURE 10: Protocol model

10.4 values

The weights for each medication combination must be manually added to our models. In table 3 the weights for every transition are shown for each medication package that is used by our models. H is high, M is moderate, L is low and R is remission. The transitions are read from left to right, so M-L 21 means that the weight of transitioning from the state Moderate to the state Low is 21.

	H-H	H-M	M-H	M-M	M-L	L-M	L-L	L-R	R-L	L-L
NONE	99	1	19	80	1	14	85	1	10	90
NON-SAARD	95	5	4	95	1	14	85	1	10	90
MTX15	69	31	17	80	19	12	85	12	9	90
MTX25	67	33	16	80	21	11	85	14	9	90
MTX25-SSZ20	65	35	15	80	23	10	85	16	9	90
MTX25-SSZ30	63	37	14	80	25	9	85	18	8	90
MTX25-ADAL20	61	39	13	80	27	8	85	20	8	90
MTX25-ADAL40	59	41	12	80	29	7	85	22	8	90
MTX-ETAN50	57	13	11	80	31	6	85	24	8	90
MTXGC	69	31	17	80	19	12	85	12	9	90
HCQ	69	31	17	80	19	12	85	12	9	90
HCQGC	69	31	17	80	19	12	85	12	9	90
TDT	62	38	13	80	26	9	85	19	8	90
TDTGC	62	38	13	80	26	9	85	19	8	90
MTXFIL	61	43	11	80	31	6	85	26	8	90
MTXTNFi	58	46	10	80	34	5	85	28	8	90
MTXTNFi2	59	45	11	80	33	6	85	27	8	90

TABLE 3: Table with weights per medication package

10.4.1 No medication

As there is no research readily available that reports the progression of RA, we have made the assumption that the chances of recovery are very slim without the use of medication.

10.4.2 non-SAARD

This medicine package is based on the research conducted by van der Heide et al[17].

This research shows that the Disability score does not improve in six months, from 1.3 to 1.3 and barely improves after 12 months, 1.2. Pain improves from 45 to 30 to 34, joint score improves from 136 to 102 to 86 and the ESR improves from 42 to 37 to 35. While these improvements seem promising, it is not a substantial improvement. Therefore, we have assumed that the initial disease activity would have been high. After six months, 35% of the patients would have gone to moderate and after twelve months 50% of all patients would have reached moderate. Low and remission are never reached.

Weeks	High	Mod	Low	Rem
0	100	0	0	0
26	65	35	0	0
52	50	50	0	0

TABLE 4: Disease Activity state assumption of NON-SAARD in percentage

10.4.3 Step-up approach starting with methotrexate monotherapy

Most of the medication packages that are present in our model are based on this step-up approach starting with methotrexate monotherapy protocol. The assumption has been

made that every next step in the protocol should have more effect than the last, meaning that each type of medication is slightly stronger than the last.

The values we had available were that after six months 45.9% of all patients have reached remission. After twelve months the percentage has grown to 60.3%[13]. The mean DAS28 score is also available in the form of a graph, at 6 months this is 2.7, at 12 months this is 2.4. This results are documented in table 5 where only the remission percentage is factual. After much experimentation, we have created the weights for each type of medication used in this protocol: MTX15, MTX25, MTX25-SSZ20, MTX25-SSZ30, MTX25-ADAL20, MTX25-ADAL40 and MTX25-ETAN50.3

Weeks	High	Mod	Low	Rem
0	100	0	0	0
26	0	4.1	50	45.9
52	0	1.7	38	60.3

TABLE 5: Disease Activity state assumption of T2T step-up with MTX in percentage

10.4.4 PRIMERA

The PRIMERA protocol is a state-of-the-art protocol that is currently in trial phase, therefore no data exists yet. The medication in the protocol is comparable with the step-up protocol10.4.3, meaning that we can make assumptions on the effectiveness of the medication.

The assumption is made that a glucocorticoid (GC) in this protocol is used as a kickstart for the medication. For example, methotrexate (MTX) takes a few weeks to have full effect, with the use of a GC this is almost instantaneous.

A second assumption is that hydroxychloroquine (HCQ) has a similar effect to MTX, as it is used as an alternative step to MTX.

A third assumption is that the Triple DMARD therapy (TDT) that consists of a mixture of MTX, HCQ and Sulfasalazine (SSZ) has a similar effect to MTX-SSZ3000.

The fourth assumption is that MTX and Filgotinib (FIB) is slightly more effective than a TDT as it is later in the protocol, but not as effective that the next step in the protocol, which is MTX and a TNF inhibitor (TNFI).

The fifth assumption is the effectiveness of MTX and a TNFI, which we assumed to be average of MTX+ADAL40 and MTX+ETAN50, as both these medicines could be used in this step.

The sixth and last assumption we made is that the effectiveness of MTX and a second TNFI would be slightly lower than the effectiveness of MTX and the first TNFI.

10.4.5 Stop

It is possible for a patient to stop using the prescribed medication and drop out of the protocol at any stage, for any reason. These reasons include people who are non-compliant, no noticeable effect, the occurrence of side effects, people no longer showing up for counselling or even death. These stop criteria are based on the research with Sankey diagrams from The research of Velthuis et al[20].

After every medication period, it is visible which medication strand is chosen next. The strand we focus on is the “complete stop” strand. This means that we have created a table

that shows how many people using a specific type of medication have stopped medication all together. We must keep in mind that after each change, only a smaller number of people continue, meaning that the data gathered from the 4th and final medication switch includes such a small number of people that the significance and trustworthiness of these results are uncertain. The difference between twenty out of a hundred people and one out of five is a significant difference, nonetheless, due to this lack of data, we have used these numbers as the basis of the stop weight for each type and can be seen in table 6.

Medicationn	Combined	Male	Female
All patients	372 (100%)	126 (33.9%)	246 (66.1%)
stopped	84 (22.6%)	28 (22.2%)	56 (22.6%)
MTX	50/302 (16.56%)	20/98 (20.41%)	30/204 (14.71%)
MTX	20/119 (16.81%)	8/41 (19.5%)	13/78 (16.7%)
MTX+SSZ	3/108 (2.78%))	0/36 (0%)	3/72 (4.17%)
HCQ	7/22 (31.82%)	5/9 (55.6%)	2/13 (15.4%)
SSZ	3/31 (9.68%)	2/13 (15.4%)	1/18 (5.6%)

TABLE 6: Number of people stopping with medication

10.4.6 Costs

The price of the medication can be seen in the table 7. These prices are collected from the farmacotherapeutischkompas[7] and rounded to the closest integer.

Medication package	Cost per week
MTX15	1
MTX25	2
MTX25+SSZ20	6
MTX25+SSZ30	7
MTX25+ADAL20	157
MTX25+ADAL40	311
MTX25+ETAN50	160
MTX+GC	3
HCQ	3
HCQ+GC	4
TDT	9
TDT+GC	10
MTX+FIL	217
MTX+TNFI	236

TABLE 7: Costs of each medication package per week in euro

11 tool

To make use of the models we created, a tool is developed in which a medical professional can load a patient model, a protocol model and a list of queries. The protocol model and queries can be fully edited within this tool, the patient model can be altered by editing the medicine package list. A manual for using the tool can be found in the GitHub accompanying this research[14].

11.1 required components

To be able to edit and create parts of these models in a tool, three base models must be loaded, the patient model, the protocol model and the queries. The patient model is provided with a list of basic medication packages that can be extended at any point. The protocol model that is required before this tool can function is a basic protocol with a single step where the next steps are itself.

11.1.1 Patient

The patient model can be edited by changing the medication packages. The tool is able to extract the medication package names and edit these accordingly. The first medication in the list has package number 0, the second 1 and so on. A new package can be added by giving a medication name that is not yet used, after which the costs, the transition weights and the stop weight must be inserted. Existing packages can be edited using a similar method where the name of the package exists in the package list. Because the protocol model depends on these packages, it is easy to create critical errors by deleting a single package, which is the reason why only adding and editing packages is supported.

11.1.2 Protocol

The protocol model can be fully edited in terms of steps, duration and medication.

A numbered list of steps is shown, after which the user can decide if they want to add, remove or edit a step. By choosing an option with a number, the tool asks for the new duration, which medication package are to be used and the next steps when the patient is in the high, moderate, low or remission state.

Lastly, the user is able to set the initial step by selecting that option.

11.1.3 Queries

The tool requires a file of queries. This list contains four queries: What is the disease severity state after 26 weeks, what is the total cost after 26 weeks, what is the last protocol step after 26 weeks and show 3 simulations of 26 weeks. The number of weeks and simulations can be edited by our tool to create the queries required by the user.

11.2 workings

This research has the UPPAAL program at its core, our tool is no different. Instead of using the whole UPPAAL suite, the tool itself only uses the UPPAAL model checker called Verifyta. This checker is a command line tool that accepts a single UPPAAL file and a query file and returns the result of the queries directly to the command line.

Our tool exports the newly edited patient and protocol models and merges the potentially edited patient and protocol model into a single file which is also exported. This merged model and the query file is passed on to Verifyta, after which the results of each query are printed on the command line.

Once Verifyta is done, the program stops and must be restarted to rerun the queries or make edits to these protocols.

11.3 Development cycle

This tool has been rapidly developed as a proof of concept prototype. The most important requirements though were implemented in the first version of the tool. The tool and functionality was later discussed with H. Vonkeman and a an additional requirement was implemented. Namely the query to see the protocol state a patient was following after a certain number of weeks.

Because of the rapid development of a prototype, the tool has not been tested by a test-panel and no more than a textual user interface has been developed.

12 Prototype results

The development of the models and tool is only worth so much if they are unable to provide trustworthy results. This section discusses the collected results and the verification of them. The results generated by our models can be found in Appendix B.

12.1 Results

To evaluate if our patient model is trustworthy, we have applied a multitude of different queries to our models. In figure 8 the mean severity states of each protocol are shown. 1 means that the patients is in remission, 4 means that the patient has a high disease activity score. This indicates that the PRIMERA protocol is the most effective of the four protocols tested, but more information is required to make such a claim.

	26 weeks	52 weeks
No medication	3.94	3.94
non-SAARD	3.56	3.44
T2T step-up MTX	1.83	1.53
PRIMERA	1.32	1.29

TABLE 8: Mean severity state of all protocols

12.1.1 no medication and non-SAARD

To start with no medication or a protocol that solely consists of a patient using a non-saard medication. We have used a query that runs 100000 simulations and reports the disease activity state after 26 and 52 weeks. The result can be seen in figures 11 and 12 respectively. These results do not include a chance of stopping, because no data exists depicting this and a patient is unable to stop following a protocol if there is no protocol to follow.

12.1.2 severity MTX and PRIMERA

For the T2T step up approach with MTX and the PRIMERA protocol we started with the same query as before. What is the disease activity after 26 and 52 weeks. Data of how manye people stop the medication used these protocols exists which is why we have simulated two different scenarios. One hypothetical scenario where no one stops and one scenario where stop criteria are implemented. These results can be seen in figures 13 and 14 for the T2T protocol and in figures 19 and 20 for the PRIMERA protocol. These models show that there is no real difference in the severity spread due to patients stopping the protocol.

12.1.3 Final protocol steps

A third query reports the final protocol step that has been used at that point in time for both the T2T protocol and the PRIMERA protocol. These results can be seen for the T2T MTX protocol at 28 weeks and 52 weeks in figures 17 and 18, for PRIMERA this is shown in figures 23 and 18. For the T2T protocol, the figures show that more patients have the first two protocol steps when the stop criteria happens. This leads to the conclusion that

most people who stop stop during those protocol steps. For the PRIMERA protocol, most patience seem to drop out at protocol step 5. This graph also shows that only a small number of patients reach protocol steps 7 or higher, meaning that developing aditional steps for this protocol might not be worth the investment.

12.1.4 costs

The final query that is evaluated by our models is the costs. The graphs with the costs for eacht patient using the T2T protocol can be seen in figure 15 and 16. The results for the PRIMERA protocol can be seen in figures 21 and 22.

The graphs for MTX shows that the costs for a large number of patients remains under the 120 euro in total, while there are patients for who the costs could rise up to 7200 euros. This costs distribution for the PRIMERA protocol differs vastly. The treatment costs of the majority of the patients remains under the 150 euros, while only a small number of patients reached a cost of at most 8500 euros.

12.2 verification

Verification of our models is done by comparing the results of our models with the results available to the public and by discussing our results with a medical professional.

12.2.1 T2T step-up MTX

The T2T protocol shows that after 26 weeks 45.1% of all patients reached remission when no one stopped, 32.6% reached remission while 28.6% of the patients stopped. After 52 weeks 60.1% of all patients reached remission when no one stopped, 41.4% reached remission while 29.8% of the patients stopped. This is close to the numbers we have found in the literature. However, this comparison is not our only verification.

Weeks	literature	tool	tool stop	total stop
0	0	0	0	0
26	45.9	45.1	32.6	28.6
52	60.3	60.4	41.4	29.8

Table 9: Remission rates T2T step-up MTX protocol

12.2.2 No med and non-SAARD

For every other protocol that we have described earlier is no such data available to the public in a directly comparable state. This is why we have discussed our findings and results with a medical professional who specializes in Rheumatoid Arthritis, prof. dr. H.E. Vonkeman. The results of not using medication or using a NON-SAARD seem to correspond with the expectations of Vonkeman.

12.2.3 PRIMERA

The predictions of the PRIMERA protocol are optimistic. The rate at which the patients reach low disease activity and remissions are faster in the beginning than expected. After 26 weeks, 75.1% of all patients have reached remission, 57.3 if we keep the stopping patients

in mind. After 52 weeks, 77.2% of all patients will have reached remission, which is 55.2 if we keep track of all stopped patients. The overall course of our prediction of the PRIMERA protocol is according to what it is hoped to achieve. Thus, while optimistic, the results are plausibly in line with the real world that can only be verified once the case study of this protocol has concluded.

13 conclusion

Over the course of this thesis the research question have been taken into mind and most of them have been answered.

13.1 Is it possible to model the treatment of RA?

The outcomes of the simulations indicate that the models are able to represent the severity of RA based on different protocols.

13.1.1 patient model

- We have kept track of the progress of RA by using four severity states. While these states are not as refined as originally imagined, this makes it possible to adapt all protocols to this format.
- We have decided to make represent the positive and negative effects of medication in the form of transition weights. This means that each type of medication has a certain weight of remaining the current RA severity of a patient or transitioning to a higher or lower severity state.
- We have decided to represent side-effects as a variable that is used to transition to the stopped state. This is because it is more likely for a patient to stop following a treatment protocol when side effects show up.
- We have made the protocol expendable by setting the transition weights by a method. This method currently uses effectiveness values set to each type of medication, this could be changed to a formula or more credible values when this is required in the future.

13.1.2 Treatment protocol model

- Each type of medication is represented as a medication package. Each package has transition weights defined in the patient model. This means that creating a model with known medication is trivial and the workings of a medication package is dependant on the patient and not the protocol.
- Each step in a protocol is defined by variables that are set by a method. The duration of each step, the medication package and the next steps based on the severity state are set by this method.
- The time that passes is represented by clocks. One clock keeps track of the total time that has passed, other clocks are to keep track of the protocol steps. Once a protocol transistion happens, a clock is reset to zero to keep track that the specified amount of weeks are reached.
- We implement protocol transitions by setting an invariant on the state in the protocol model. Once the protocol clock reaches the value set by a protocol step, the method to change the current values is called and the clock is reset.
- A protocol can be personalized by adding if statements to the protocol steps. This personalizing of the protocol has been placed out of the scope for this research, however, the option is left available.

13.1.3 Evaluation treatment protocol models

- The data generated by the simulation is evaluated in the form of queries. The required data is the cost and severity state at each moment in time.
- This data is extracted at specific times by setting a time limit on the simulations and outputting the results at that point in time.
- The data is currently represented by a list of values of doubles.

13.2 Is it possible for medical professionals to create and evaluate models?

This research has shown that it is theoretically possible for anyone to create and edit a protocol with the use of our created tool.

13.2.1 Prototype tool for creating a protocol model

- The information that is required to make a protocol is not as much as originally expected. For each protocol step we have six variables that are required. The medication package and how many weeks this package should be used. The last four variables are the identifiers of the next protocol steps when the patient is in high severity, moderate severity, low severity, or in remission. The final information that is required is the initial protocol step.
- A protocol model in constructed by a simple loop that calls a method and a method that sets the data specified in each protocol step.
- Each model is an xml file that can be loaded into our python program as a xml tree.
- We select the part of the xml file that contains the method and read all the information that is included in it. This information is stored in dictionaries. The values in these dictionaries is edited by removing, adding, and editing the values, after which the entire method is generated. This method overwrites the original method.
- Queries are stored in a separate query file. These queries are loaded and edited in the same way as the method from the protocol file.
- The newly created protocol can be exported by writing the xml element to an xml file. This method is supported out of the box by python.
- The commands that are required are showing all medication packages, editing and adding one, showing all protocol steps, adding removing and editing protocol steps, showing, editing, and adding queries, and a run query command. The run query command has the option to export the newly made xml files.

13.2.2 Linking the tool to UPPAAL

- Connecting python to UPPAAL is done by using the standalone SMC called VERIFYTA. The program sends the a model that combines the patient and protocol model, as well as a query file to VERIFYTA and the results of these queries are returned to the program.

- the length of the simulation is decided by the queries. Each query has the option to set the number of weeks each simulation should run for, as well as the number of runs required to verify each query.

13.2.3 Making the tool user friendly

The main focus of this research was getting the models in credible shape, which meant that the development of the tool had to be pushed back on the timeline of this research. Because a thesis only has a limited time span to perform the research and a proof of concept tool was set as a goal for this research, a prototype tool has been created without user testing. All functionality and methods have been tested by the developers, the usability has not been tested extensively. The tool has been discussed in terms of usability and functionality with H. Vonkeman, but should be discussed and explored more extensively with a large test panel before the tool is further developed..

13.3 GitHub

The final versions of the tool, required models and a manual of how to use our tool can be found on our GitHub [14]. Our tool is a steppingstone that enables the use of models to experiment with new protocols that could potentially be used as a discussion tool, after more research is done and the medicine combination values are improved.

14 Future study

This thesis has shown that it is possible to digitally evaluate the effectiveness of existing treatment protocols for Rheumatoid Arthritis and of new untested protocols. The models created to proof this concept are still in the proof of concept stage and must be extended and tested more extensively before it should be adapted as a discussion piece during the development of new protocols.

14.1 future study

As stated above, this thesis proves that it is a viable option to model a patient and treatment protocol to predict the effectiveness of the protocol. However, this research has shown that it is able to predict the course of the severity for the general patient population. This could and should be improved in further research.

14.1.1 personalization

The patient model created for this research represents the general patient population. The end goal, of which this research is a stepping stone, is to be able to create a model of a specific patient. Multiple protocols can be tested on this model, resulting in a list of reasons why some protocols are a good viable option and why some protocols should be avoided. This requires more research in the following fields:

- The effect of patient characteristics.
  How much does the sex, BMI, or RA factor of a person affect the positive workings of medication and how much does this influence the probability of developing side effects?

14.1.2 Medication

The medication packages represented in the patient model are good first iteration. However, these should be improved before the tool can be used as a discussion piece.

- The exact workings of medication
  The effectiveness of each type of medication used in this research is set to the best of the ability of the researcher. These values are based on available research data and must be further tuned to be more reliable before our tool is used to decide the best protocol for a patient. These values can be improved by using existing medical data or by conducting new clinical studies on the exact workings of a type of medication and transforming these to the required format.

- Adding a better overview for side effects.
  In the current form of the models, side effects are solely modeled in the form of a "stop weight". This could be improved such that the protocol returns the probability of which side effects have a high probability to show up, which could be discussed with the patient to make a well informed decision about which protocol to use.

14.1.3 tool

The tool created for this research has not been tested by a test panel on usability and unexpected behaviour. Because the expectations of end users and the vision of a developer

can differ vastly, this testing is an important step before the tool should be released for practical use.

- Graphical user interface.
  The current version of the tool is written in python and requires exact inputs in order to function. While a textual user interface is usable, it is far from user friendly. This is why the creation of a GUI could be a useful. This could make it possible to visualise the protocol and make editing errors less common.

- Testing.
  Each piece of software should be thoroughly tested before it becomes available to the public. This has not been done for the prototype tool due to time constraints. Possible ways of testing includes user tests, probing and functional testing.

14.1.4 verification

The created models have been tweaked to be in line with data from publicly available medical research. In order to verify the credibility of the models, more research must be sampled and the results from those papers must be reproducible with our models. One example of new research that is not yet available to the public or the medical world, are the result of the PRIMERA trial. We have created a protocol model that allows us to predict the outcomes for the PRIMERA trial, which are at this moment no more than speculation based on earlier trials. Once the trial has been conducted, the credibility of our tool could increase if the predicted outcomes are in line with the real outcomes, or our models can be improved to be in line with the real outcomes. After updating the patient model, the outcomes for a potential new trial could be predicted, after which it can be verified or improved. This is the never ending cycle that the models must go through in order to become and remain credible for the coming years.

References

[1] Kimmo Aho and Markku Heliövaara. Risk factors for rheumatoid arthritis. *Annals of Medicine*, 36(4):242–251, 2004. PMID: 15224650.

[2] Jacqueline Bullock, Syed A A Rizvi, Ayman M Saleh, Sultan S Ahmed, Duc P Do, Rais A Ansari, and Jasmin Ahmed. Rheumatoid arthritis: A brief overview of the treatment. *Medical principles and practice : international journal of the Kuwait University, Health Science Centre*, Sep 2018.

[3] Cynthia S. Crowson, Eric L. Matteson, Elena Myasoedova, Clement J. Michet, Floranne C. Ernste, Kenneth J. Warrington, John M. Davis III, Gene G. Hunder, Terry M. Therneau, and Sherine E. Gabriel. The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. *Arthritis & Rheumatism*, 63(3):633–639, 2011.

[4] Dr. P.H.P. de Jong. PeRsonalIzed MEdicine in Rheumatoid Arthritis (PRIMERA trial): A multicentre, open-label, randomized controlled trial comparing usual care with a tailor-made approach. aug 2022.

[5] Thomas Frisell, Marie Holmqvist, Henrik Källberg, Lars Klareskog, Lars Alfredsson, and Johan Askling. Familial risks and heritability of rheumatoid arthritis: Role of rheumatoid factor/anti–citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis & Rheumatism*, 65(11):2773–2782, 2013.

[6] Vijay Joshua, Katerini Chatzidionisyou, and Anca I. Catrina. Role of the lung in individuals at risk of rheumatoid arthritis. *Best Practice & Research Clinical Rheumatology*, 31(1):31–41, 2017. Individuals at risk of rheumatoid arthritis—the evolving story.

[7] Zorginstituut Nederland. Farmacotherapeutisch kompas. `https://www.farmacotherapeutischkompas.nl/`, Oct 2022.

[8] Elena Nikiphorou, Sam Norton, Lewis Carpenter, Josh Dixey, David Andrew Walsh, Patrick Kiely, Adam Young, and on behalf of the Early Rheumatoid Arthritis Study and the Early Rheumatoid Arthritis Network Cohorts . Secular changes in clinical features at presentation of rheumatoid arthritis: Increase in comorbidity but improved inflammatory states. *Arthritis Care & Research*, 69(1):21–27, 2017.

[9] Marta Novella-Navarro, Chamaida Plasencia-Rodríguez, Laura Nuño, and Alejandro Balsa. Risk factors for developing rheumatoid arthritis in patients with undifferentiated arthritis and inflammatory arthralgia. *Frontiers in Medicine*, 8, 2021.

[10] J. E. Oliver and A. J. Silman. Risk factors for the development of rheumatoid arthritis. *Scandinavian Journal of Rheumatology*, 35(3):169–174, 2006. PMID: 16766362.

[11] Francesca Oliviero, Paolo Spinella, Ugo Fiocco, Roberta Ramonda, Paolo Sfriso, and Leonardo Punzi. How the mediterranean diet and some of its components modulate inflammatory pathways in arthritis. *Swiss Medical Weekly*, 145(4546):w14190, Nov. 2015.

[12] Jeffrey A. Sparks, Chia-Yen Chen, Linda T. Hiraki, Susan Malspeis, Karen H. Costenbader, and Elizabeth W. Karlson. Contributions of familial rheumatoid arthritis or

lupus and environmental factors to risk of rheumatoid arthritis in women: A prospective cohort study. *Arthritis Care & Research*, 66(10):1438–1446, 2014.

[13] L. M. M. Steunebrink, G. A. Versteeg, H. E. Vonkeman, P. M. ten Klooster, H. H. Kuper, T. R. Zijlstra, P. L. C. M. van Riel, and M. A. F. J. van de Laar. Initial combination therapy versus step-up therapy in treatment to the target of remission in daily clinical practice in early rheumatoid arthritis patients: Results from the dream registry - arthritis research & therapy. *BioMed Central*, Mar 2016.

[14] Thomas Stouten. Modeling the effectiveness and side effects of treatment for rheumatoid artritis with uppaal tool and models. `https://github.com/TStouten/Modeling-Rheumatoid-Artritis-with-UPPAAL`, JAN 2024.

[15] National Guideline Centre (UK). Treat-to-target: Rheumatoid arthritis in adults: Diagnosis and management. *National Center for Biotechnology Information*, Jul 2018.

[16] Frederik van Delft, Mirte Muller, Rom Langerak, Hendrik Koffijberg, Valesca Retèl, Daan van den Broek, and Maarten IJzerman. Modeling diagnostic strategies to manage toxic adverse events following cancer immunotherapy. *Medical Decision Making*, 41(6):693–705, 2021. PMID: 33813943.

[17] A van der Heide and J.W.G. Jacobs. The effectiveness of early treatment with "second-line" antirheumatic drugs. a randomized, controlled trial, Apr 1996.

[18] P.L.C.M. van Riel. Das28. `https://www.das-score.nl/en/`, Jan 2023.

[19] P.L.C.M. van Riel. Twenty-eight joints. `https://www.das-score.nl/en/das-and-das28/how-to-meaure-the-das28/twenty-eight-joints`, Jan 2023.

[20] Kimberly Velthuis, Fenna Poppelaars, Peter M. ten Klooster, Harald E. Vonkeman, and Naomi T. Jessurun. Impact of adverse drug reactions on the treatment pathways of early rheumatoid arthritis patients: a prospective observational cohort study. *Expert Opinion on Drug Safety*, 22(8):753–762, 2023. PMID: 36683590.

[21] Jake Weddell and Samantha Hider. Diagnosis and management of rheumatoid arthritis. *Prescriber*, 32(10):15–20, 2021.

[22] Peter Wetselaar, Frank Lobbezoo, Pepijn de Jong, Umarah Choudry, Jasper van Rooijen, and Rom Langerak. A methodology for evaluating tooth wear monitoring using timed automata modelling. *Journal of Oral Rehabilitation*, 47(3):353–360, 2020.

15 Appendices

A variables

Das28 The current DAS28 score multiplied by 1000 to allow scores like 5,234 in UPPAAL.

Generalhealth This is the health of the patient as used in the DAS28 score, a value between 0 and 100.

Sex This is the sex of a patient at birth, male or female.

Weekinterval This is the clock that is used to simulate a week passing.

mtxmult A value that sets the effectiveness multiplier for MTX. The higher this number, the less the effect of the medication is. This variable exists for every type of medication used by the patient.

mtxrand A value specifically generated for MTX to add randomness to the effectiveness of MTX. -2 Meaning that MTX does worsen the DAS28 score, 5 meaning the highest probable effect. This variable exists for every type of medication used by the patient.

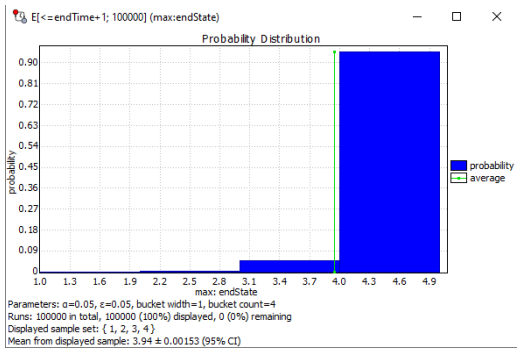
weekMTX This variable keeps count of how many consecutive weeks the medication MTX has been used by the patient. This variable exists for every type of medication used by the patient.

MTXbijw This variable tracks if the patient experiences side effects of MTX during this week. If the variable is 1, the patient experiences some form of side effects, 0 means no side effects. This variable exists for every type of medication used by the patient.

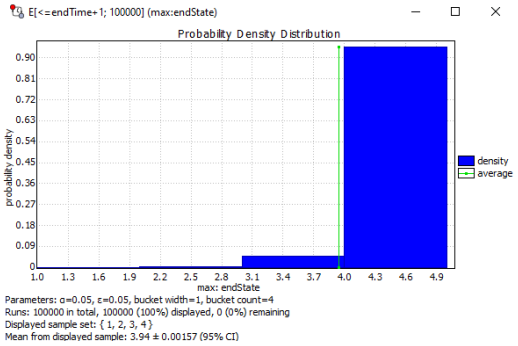
weekmed Keeps track of the week within the protocol. This is not the current week, as this could be reset by the protocol when the patient resumes the protocol after they had reached remission.

weekdeadline This variable indicates when the next step op a protocol should be started.

B Results



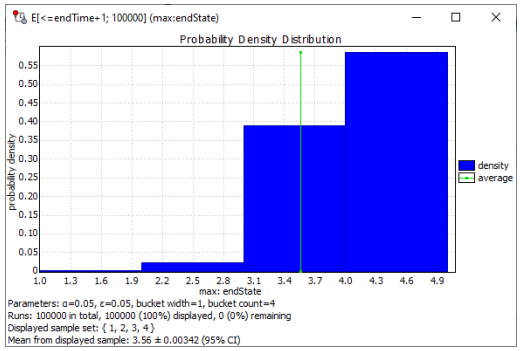
(A) 26 weeks



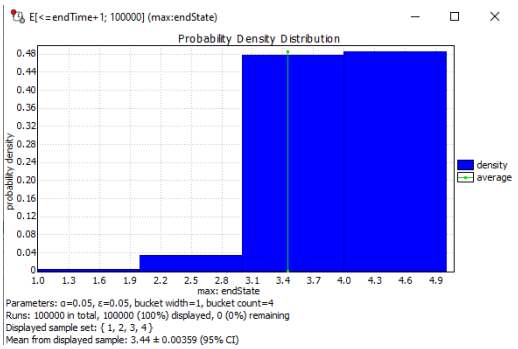
(B) 52 weeks

FIGURE 11: Disease activity state No Medication

These figures show the final severity of the patient after 26 and 52 weeks. State 1 depicts remission, state 2 depicts low disease activity, stage 3 depicts moderate disease activity and stage 4 depicts high disease activity.



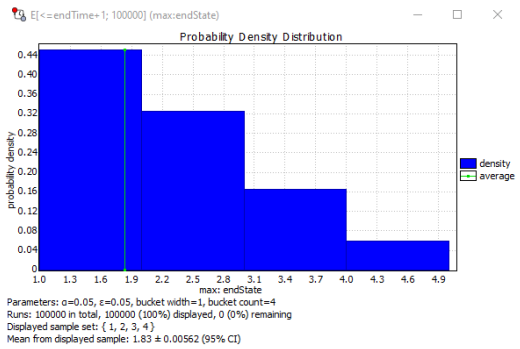
(A) 26 weeks



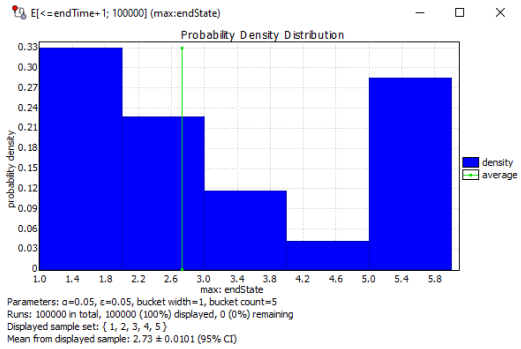
(B) 52 weeks

FIGURE 12: Disease activity state NON-SAARD medication

These figures show the final severity of the patient after 26 and 52 weeks. State 1 depicts remission, state 2 depicts low disease activity, stage 3 depicts moderate disease activity and stage 4 depicts high disease activity.



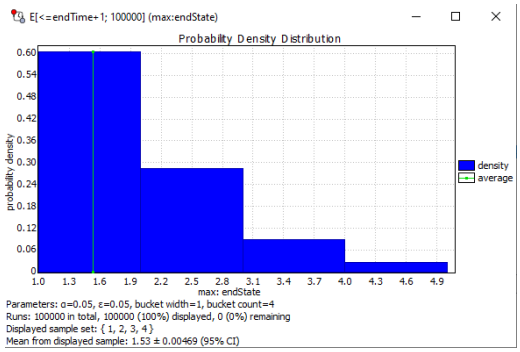
(A) Without Stop criteria



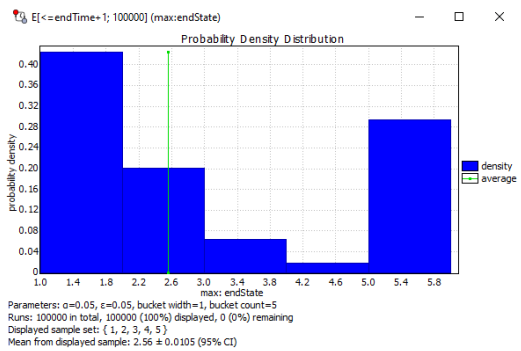
(B) With Stop criteria

FIGURE 13: Disease activity state T2T Step-Up medication 26 weeks

These figures show the final severity of the patient after 26 with a stop criteria disabled and enabled. State 1 depicts remission, state 2 depicts low disease activity, stage 3 depicts moderate disease activity, stage 4 depicts high disease activity, and stage 5 depicts that a patient has stopped using medication for any reason. These models indicate that the severity distribution does not change when patients drop out of the protocol.



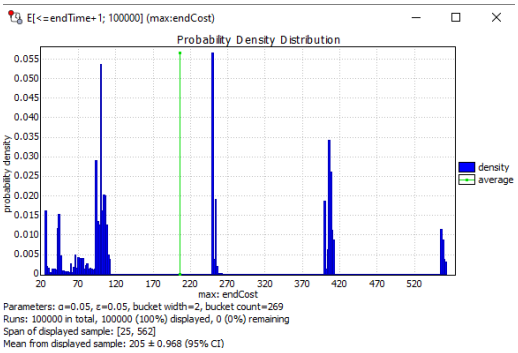
(A) Without Stop criteria



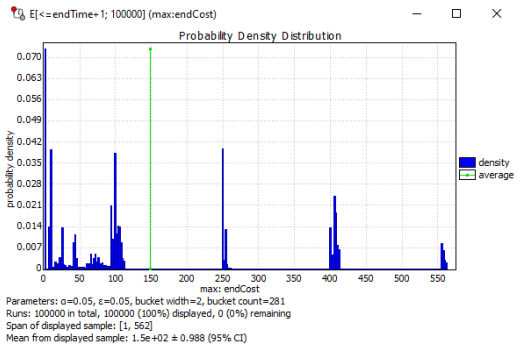
(B) With Stop criteria

FIGURE 14: Disease activity state T2T Step-Up medication 52 weeks

These figures show the final severity of the patient after 52 with a stop criteria disabled and enabled. State 1 depicts remission, state 2 depicts low disease activity, stage 3 depicts moderate disease activity, stage 4 depicts high disease activity, and stage 5 depicts that a patient has stopped using medication for any reason. These models indicate that the severity distribution does not change when patients drop out of the protocol.



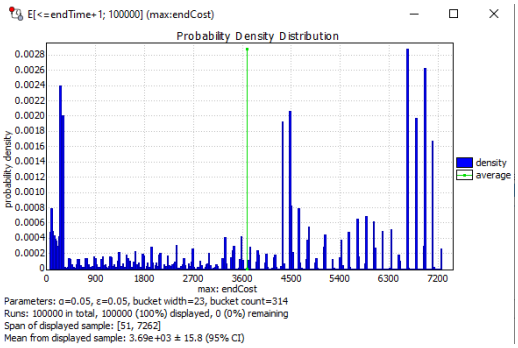
(A) Without Stop criteria



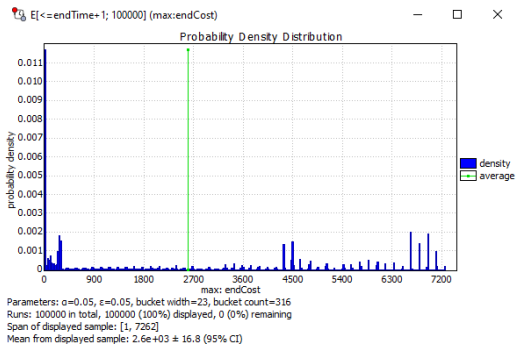
(B) With Stop criteria

FIGURE 15: Costs T2T Step-Up medication 26 weeks

These figures show the costs of the treatment after 26 weeks for each simulation. Patients who stop are most often in the early stages using the type of medication that is not that expensive. This can be seen by the bars between zero and fifty increasing in heights, while higher costs are decreasing in height.



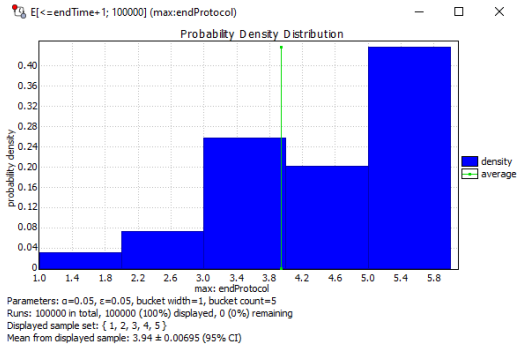
(A) Without Stop criteria



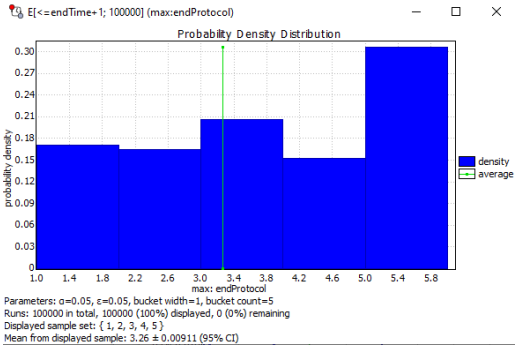
(B) With Stop criteria

FIGURE 16: Costs T2T Step-Up medication 52 weeks

These figures show the costs of the treatment after 52 weeks for each simulation. Patients who stop are most often in the early stages using the type of medication that is not that expensive.



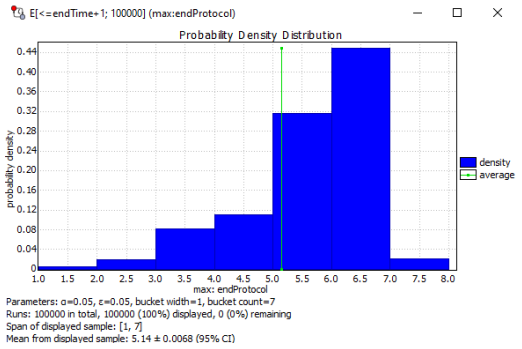
(A) Without Stop criteria



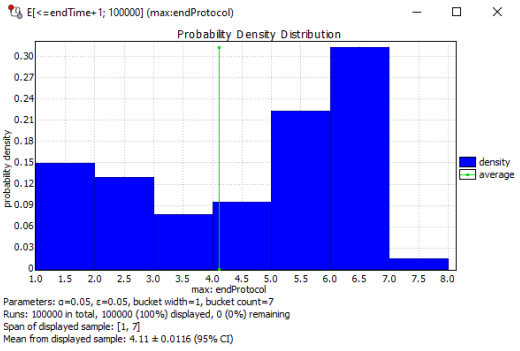
(B) With Stop criteria

FIGURE 17: Final protocol step T2T Step-Up medication 26 weeks

These figures show the final protocol steps of patients after 26 weeks. The numbers of the step correspond with the step number that are used in the protocol. These figures indicate that most patients who stop with the protocol do so in the first few steps.



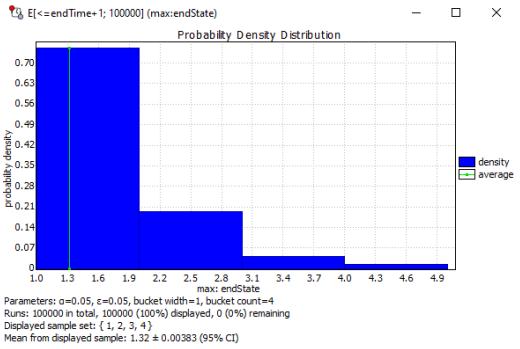
(A) Without Stop criteria



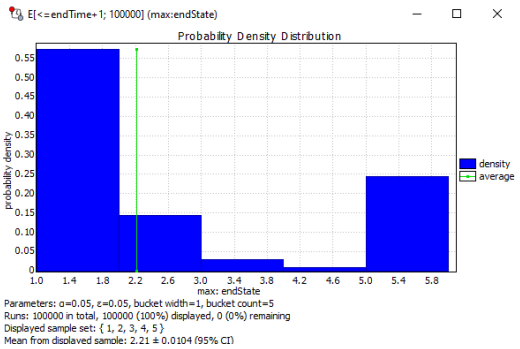
(B) With Stop criteria

FIGURE 18: Final protocol step T2T Step-Up medication 52 weeks

These figures show the final protocol steps of patients after 52 weeks. The numbers of the step correspond with the step number that are used in the protocol. These figures indicate that most patients who stop with the protocol do so in the first few steps.



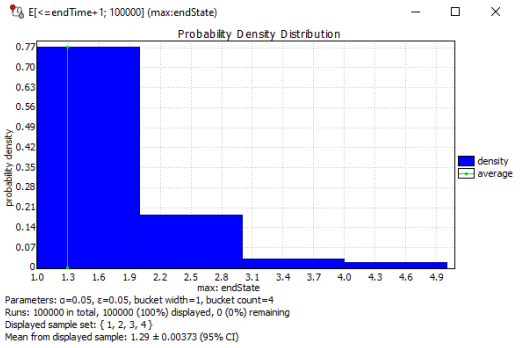
(A) Without Stop criteria



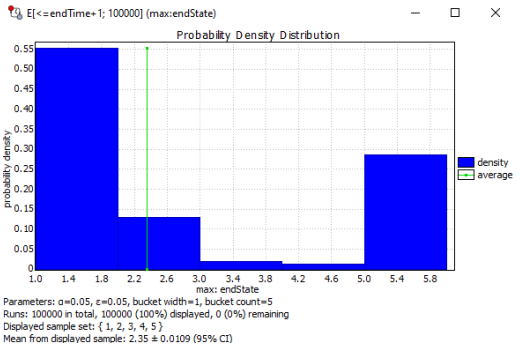
(B) With Stop criteria

FIGURE 19: Disease activity PRIMERA 26 weeks

These figures show the final severity of the patient after 26 with a stop criteria disabled and enabled. State 1 depicts remission, state 2 depicts low disease activity, stage 3 depicts moderate disease activity, stage 4 depicts high disease activity, and stage 5 depicts that a patient has stopped using medication for any reason. These models indicate that the severity distribution does not change when patients drop out of the protocol.



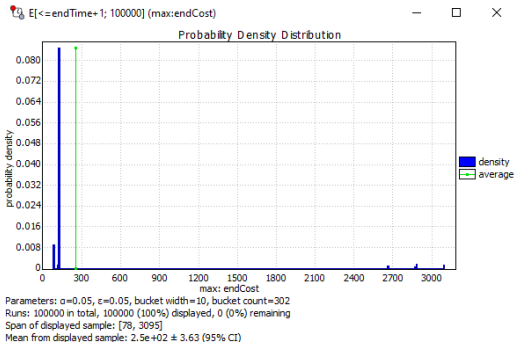
(A) Without Stop criteria



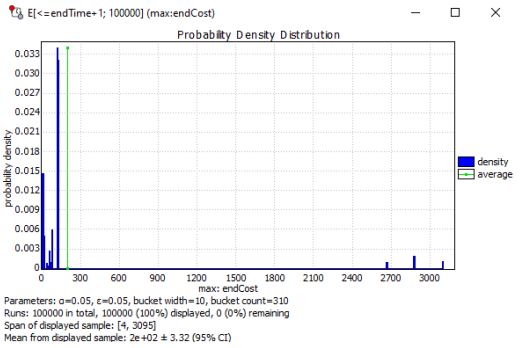
(B) With Stop criteria

FIGURE 20: Disease activity PRIMERA 52 weeks

These figures show the final severity of the patient after 52 with a stop criteria disabled and enabled. State 1 depicts remission, state 2 depicts low disease activity, stage 3 depicts moderate disease activity, stage 4 depicts high disease activity, and stage 5 depicts that a patient has stopped using medication for any reason. These models indicate that the severity distribution does not change when patients drop out of the protocol.



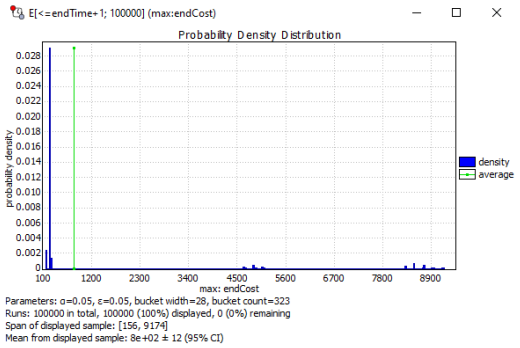
(A) Without Stop criteria



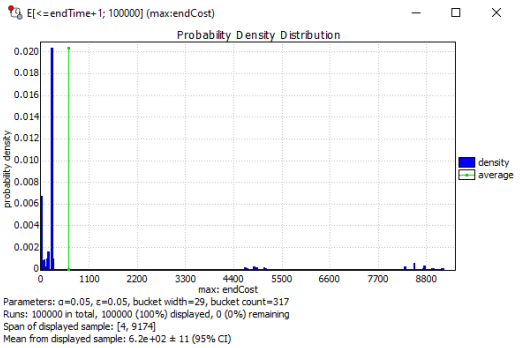
(B) With Stop criteria

FIGURE 21: Costs PRIMERA 26 weeks

These figures show the costs of the treatment after 26 weeks for each simulation. Patients who stop are most often in the early stages using the type of medication that is not that expensive. This can be seen by the bars between zero and fifty increasing in heights, while higher costs are decreasing in height.



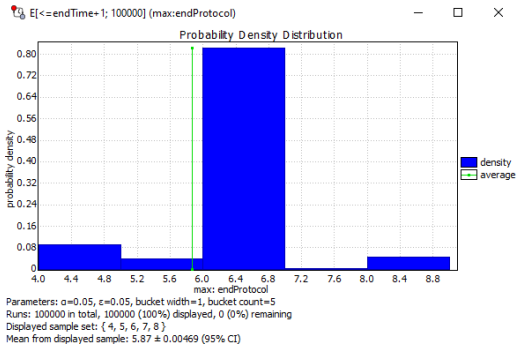
(A) Without Stop criteria



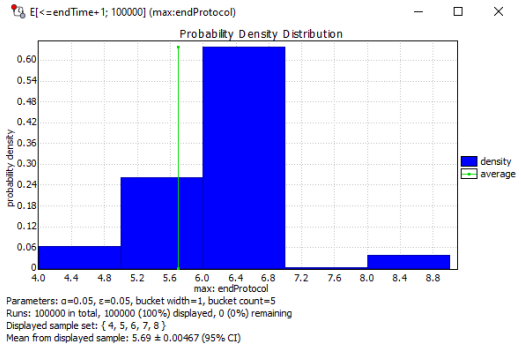
(B) With Stop criteria

FIGURE 22: Costs PRIMERA 26 weeks

These figures show the costs of the treatment after 52 weeks for each simulation. Patients who stop are most often in the early stages using the type of medication that is not that expensive. This can be seen by the bars between zero and fifty increasing in heights, while higher costs are decreasing in height.



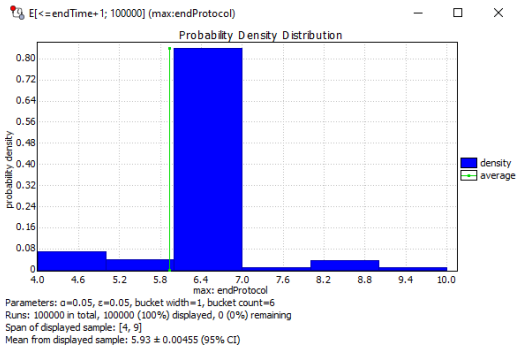
(A) Without Stop criteria



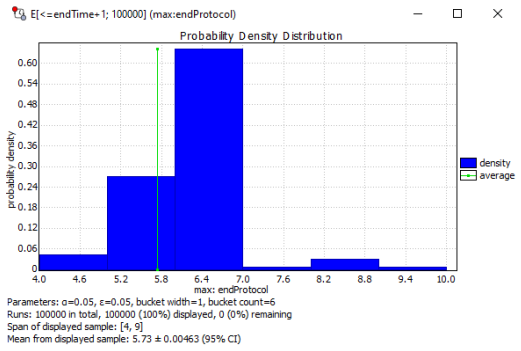
(B) With Stop criteria

FIGURE 23: Final protocol step PRIMERA 26 weeks

These figures show the final protocol steps of patients after 26 weeks. The numbers of the step correspond with the step number that are used in the protocol. These figures indicate that most patients who stop with the protocol do so in the protocol step labeled 5.



(A) Without Stop criteria



(B) With Stop criteria

FIGURE 24: Final protocol step PRIMERA 52 weeks

These figures show the final protocol steps of patients after 26 weeks. The numbers of the step correspond with the step number that are used in the protocol. These figures indicate that most patients who stop with the protocol do so in the protocol step labeled 5.