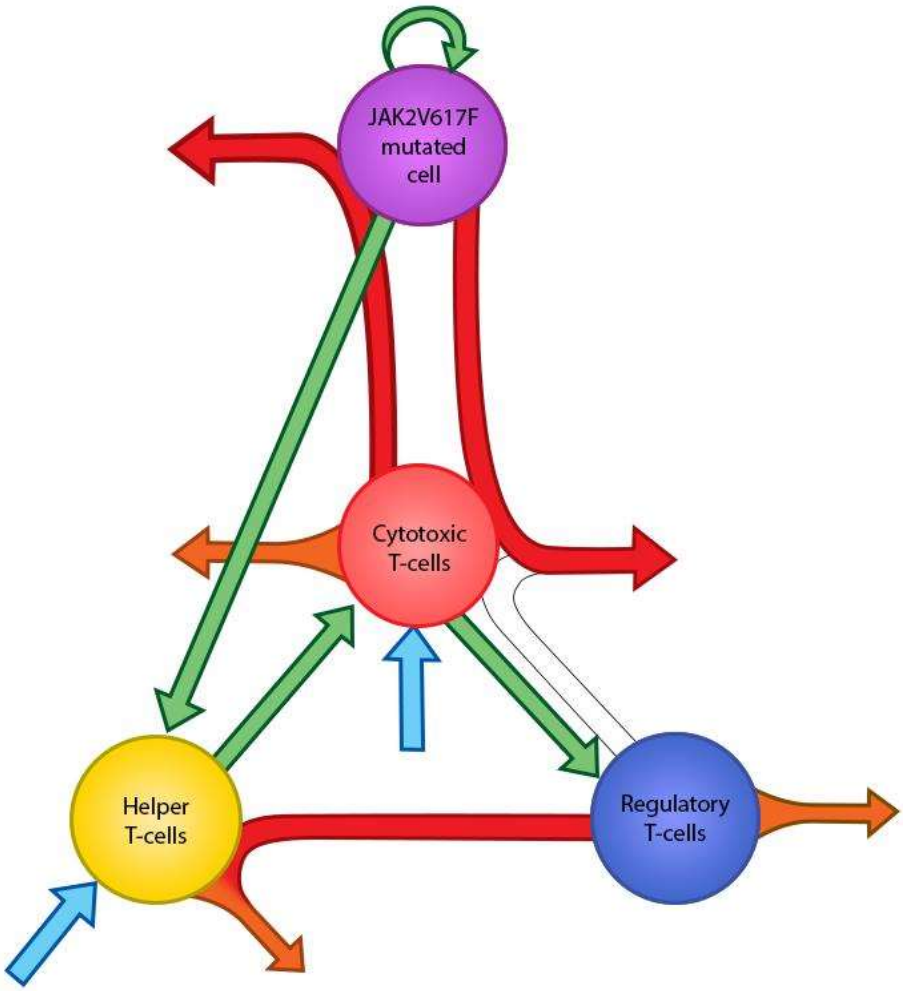
Modeling the interaction between the immune system and leukemia

Authors: Stine G. Fohrmann, Lukas Rasocha, Jūlija Tumanovska, Julien Hirsch and Kata M. Riegels



The conceptional model of our report

Roskilde University

Supervisor: Johnny T. Ottesen

1st semester project NIB 2019, autumn semester

Abstract

This report covers modeling the interaction between the body's immune system and leukemic cells. We suggest a mathematical model with a specific parameter set that models this interaction with possible uses in the medical field in the future. Based on known information about the immune response, we created the model and validated it with patient data showing the cell count of relevant cells. We included Interferon treatment in our model, which has not been done before in the way. By this, we try to improve the general understanding of the immune cells' response to leukemic cells and how the Interferon treatment works.

Table of Contents

Introduction

“Blood is a very special juice” (Johann W. Goethe, Faust Part 1). Any kind of disturbance in its complex system of production and functions can be fatal to the human body and health. Blood cells are separated into three types: red blood cells, white blood cells and platelets. They get produced by hematopoietic stem cells (HSC) in the bone marrow. Over- or underproduction of blood cells can lead to serious complications. (Mader, Windelspecht, and Preston 2011)

The human body defends itself against possible threats with an immune response which targets intruding or malignant cells. The immune reaction is a complex process, which involves many different cells with different tasks. Damaged cell tissue can trigger an inflammatory response as a part of the immune reaction. Inflammation is meant to heal the body, though in some cases it can also harm it if it turns chronic which is defined by the simultaneous damaging and repair of tissue. This can lead to a variety of inflammatory diseases (Chen et al. 2018).

Part of the immune response is the activation of T cells. These types of cells can generally be divided into three categories: Helper T cells, cytotoxic T cells and regulatory T cells (Sharma 2019).

The focus in this report will be a specific kind of blood cancer, Myeloproliferative Neoplasms (MPNs), which is caused by mutations in HSC and believed to trigger and drive inflammatory diseases (Hasselbalch and Bjørn 2015).

MPNs are types of leukemia, which are being more and more acknowledged to have a strong inflammatory component. Although it is commonly recognized as inflammatory disease, it is still a rather novel concept and it is thereby important to find additional proof by mathematical modelling. Mathematical modelling that couples MPN cancer development with the immune response is something that studies only very recently have started to investigate (Andersen et al. 2017).

In this project we will attempt to create a mathematical model that could help treat leukemia. By being able to predict the development of the disease with a mathematical model, decision-making can be simplified, and better treatment schedules can be worked out. The report focuses on finding a specific set of parameters, which are universal for every single patient with this disease and treatment.

This leads us to the research question:

What set of equations could represent a mathematical model, which predicts the development of leukemia and the treatment thereof?

- What set of parameters can fit a specific set of data, and how well would these parameters work for other data?
- How can the interaction between T cells and leukemia cells be modelled?
- How can our model assist in understanding the disease with and without IFN-treatment?
- How can a mathematical model predict the progression of the disease and treatment for it?
- How can our mathematical model be used in the medical field?

To create the model, we received data from Rigshospitalet in Copenhagen with counts of mutated cells and T cells of three patients diagnosed with essential thrombocythemia (a type of MPN). Based on this data we created a model of four differential equations to predict the development of cancer cells and T cells during a treatment period of 10 years and without the treatment.

By building the model we went through a process of research, building, fitting, studying and testing, which will be further explained in chapter 3.1.

Chapter 1 and 2 give an overview over relevant biological and mathematical theory. In the following chapters 3 and 4 we explain the method we are using to create our model and introduce our assumptions and equations. We also present a conceptual version of our model. In chapter 5 “Analysis” we analyze the model from the graph, statistics and sensitive parameters.

In chapter 6 “Discussion” we discuss the results of our analysis as well as considering how to improve the model and other possible approaches to our model. Because of a limited amount of data and a complicated model, we ended up with a parameter set for our model which did not show good prediction. Furthermore, it is important to mention that the model is built on many assumptions, which strongly influence its applicability and accuracy. Through sensitivity analysis of our parameters we get an idea of which parameter has most influence in fighting cancer and can thereby give a guess of what treatment could be most effective. Additionally, we’ll have a look at the usefulness of our model in the medical field.

In chapter 7 “Conclusion” we give a summary of all the work we have done to answer our research question and associated sub-questions, concluding the achievements and abilities of our model.

Terminology

The intention of this page is to have definitions to return to when coming across terms which one may not know.

Antibodies	"A protein produced by certain white blood cells in response to entry into the body of a foreign substance in order to render it harmless." (Singleton and Sainsbury 2006)
Antigens	"Any substance that the body regards as foreign and that therefore elicits an immune response, particularly the formation of specific antibodies capable of binding to it." (Martin and Hine 2015)
Cell lysis	The rupture of cells. (Singleton and Sainsbury 2006)
Cytotoxic	"Causing death or injury to cells" (*OED: Oxford English Dictionary: The Definitive Record of the English Language* 2010)
Downregulation	Opposite of upregulation (Martin and Hine 2015)
Hemoglobin	"One of a group of globular proteins occurring widely in animals as oxygen carriers in blood." (Martin and Hine 2015)
Hematopoiesis	"The formation of blood" (*OED: Oxford English Dictionary: The Definitive Record of the English Language* 2010)
Homeostasis	**"**The maintenance of a dynamically stable state within a system by means of internal regulatory processes that tend to counteract any disturbance of the stability by external forces or influences**"** (*OED: Oxford English Dictionary: The Definitive Record of the English Language* 2010)
Humoral	"Relating to the blood or other body fluids." (Martin and Hine 2015)
Lymphocytes	White blood cells that are responsible for immune reactions. (Martin and Hine 2015)
Immunologic tolerance	"The phenomenon by which the cells of the immune system are constrained from mounting an immune response against 'self' tissues." (Martin and Hine 2015)
Inflammation	"The defence reaction of tissue to injury, infection, or irritation by chemicals or physical agents." (Martin and Hine 2015)
Interleukin	"Any of numerous cytokines that are produced by leucocytes and perform a range of regulatory functions for cells of the immune system,… ." (Martin and Hine 2015)
Microenvironment	"A small-scale, local, or specialized environment, esp. as a distinct part of a larger environment" (*OED: Oxford English Dictionary: The Definitive Record of the English Language* 2010)
Phagocytes	Any kind of cell, which ingests and digests particular matter. (Singleton and Sainsbury 2006)
Proliferation	The instant increase or growth, esp. of or involving cells of the same type (*OED: Oxford English Dictionary: The Definitive Record of the English Language* 2010)
Upregulation	"An increase in the sensitivity of a cell to a chemical substance, such as a hormone, signal molecule, or drug, due to an increase in the density of cell-surface receptors for that molecule." (Martin and Hine 2015)

Chapter 1: Biological Theory

This section provides knowledge that will aid with understanding the model later on, giving insight into how the immune response works, what the functions of the different cells involved are and how they interact with each other. The emphasis here is on lymphocytes - they are the cells to be modeled in this project. In this chapter we also describe the type of leukemia (essential thrombocythemia) that we will model.

1.1 Blood cells

Blood is arguably the most important part of the body. Its many functions include transporting nutrients, waste and hormones, regulating body temperature, and defending against foreign matter within the body. The latter mechanism is called the immune response, a process that we will further examine in the next chapter. (Nagle 2008)

Blood is classified as a liquid connective tissue, although some classify it as an organ. It consists of formed elements (cells and cell fragments) and plasma (Mader, Windelspecht, and Preston 2011).

Red blood cells

Red blood cells are small red biconcave (concave on both sides) shaped disks without a nucleus. Red blood cells (also known as erythrocytes) contain hemoglobin, the substance that combines reversibly with oxygen. The main purpose of erythrocytes is gas exchange - hemoglobin binds with oxygen in the lungs and releases it in the tissues, where it picks up carbon dioxide to transport it back to the lungs (Kuby 1992).

Platelets

Platelets (thrombocytes) are small irregularly shaped cell fragments. These formed elements are involved in the process of blood clotting. When platelets come into contact with air or rough surfaces, they bind to each other, causing blood clot formation (Nagle 2008).

White blood cells

White blood cells (leucocytes) are larger than red blood cells. White blood cells are classified into the granular leucocytes and the agranular leucocytes (Mader, Windelspecht, and Preston 2011). The main focus of this project is the T lymphocytes, which are a type of agranular leucocytes (and will be looked into in later chapters).

Both granular leucocytes and agranular leucocytes have granules, but granules of agranular leucocytes are hard to see even under a microscope. Agranulocytes include the monocytes and the lymphocytes. Monocytes are involved in activation of one of the types of lymphocytes, and after further differentiation they also destroy illness inducing microorganisms, old cells and cellular debris.

Lymphocytes are responsible for specific immunity – specific lymphocytes recognize specific antigens by having the specific membrane receptors for the foreign material. There are three types of lymphocytes – natural killer (NK) cells, B cells and T cells (Mader, Windelspecht, and Preston 2011). Natural killer cells' main function is destruction of virus-infected self-cells and some types of tumor cells (Lydyard, Whelan, and Fanger 2011), B cells are responsible for antibody production (Kuby 1992), while T cells regulate immune responses and provide help for B cell responses (Lydyard, Whelan, and Fanger 2011).

1.2 Immune response

This chapter introduces the T cell immune responses, as well as the three main types of T cells – T helper (T_h) cells, T cytotoxic (T_c) and T regulatory (T_{reg}) cells. T_h cells upregulate the production of T_c cells and provide help to B cells by signaling. T_c cells kill infected and malignant cells. T_{reg} cells suppress other T cells, ensuring that autoimmune processes do not happen (Lydyard, Whelan, and Fanger 2011). Signaling proteins called cytokines are also introduced, with an emphasis on Interleukin-2. This information is later used to come up with assumptions, which are the basis for the equations in this model.

Antigens and immune response activation

Antigens are the substances that induce an immune response. Lymphocytes can recognize proteins, carbohydrates, lipids, nucleic acids and other antigens. Once recognized by lymphocytes, they start proliferating, producing cytokines (explained below) and/or antibodies (depending on the type of lymphocyte) (Lydyard, Whelan, and Fanger 2011).

Cytokines

Cytokines are an array of protein mediators involved in cell growth, inflammation, immunity, differentiation and repair that are produced by leucocytes (Paul 1993). Moreover, they are immune system proteins that are biological response modifiers. They serve as chemical communicators (messengers of the immune system) from one cell to another by combining with surface receptors on target cells, which enables them to coordinate T cell immune system interactions and amplify immune reactivity (Cruse and Lewis 1999).

Cytokines regulate the intensity and duration of the immune response by stimulating or inhibiting the activation, proliferation or differentiation of various cells. Thus, the cytokines secreted by a single lymphocyte can influence the activity of various cells involved in the immune response. For example, cytokines produced by activated T helper cells can influence the activity of cytotoxic T cells (Kuby 1992). One of the two producers of cytokines are T helper cells. A binding of a cytokine to its receptor induces numerous responses including the development of a cellular and humoral immune response (happens outside of cells), induction of the inflammatory response, regulation of blood cell production, control of cellular proliferation and differentiation and induction of wound healing (Kuby 1992).

Their effects may be:

- autocrine - acting on cells that produce them,
- paracrine – acting on neighboring cells,
- endocrine – acting on cells at distant sites.

Interleukin

Interleukin is a group of cytokines that can be divided into different major groups. **Interleukin-1** (also known as lymphocyte activating factor (Klein 1990)) is an initiation molecule that acts on cells that participate in the inflammatory response(Tizard 1992). **Interleukin-2,** which is the T cell growth factor, is synthesized by T helper cells (and a minor fraction of cytotoxic T cells) and has an autocrine effect on it. The amount of Interleukin-2 that T helper cells produce is a principal factor in determining the strength of an immune response (Cruse and Lewis 1999). Generally, Interleukin-2 affects cytotoxic T cells by supporting long-term growth, enhancing their activity and inducing proliferation (Kuby 1992). **Interleukin-3** is created by T helper cells and participates in the immune response by facilitating

proliferation (rapid cell reproduction) of some hematopoietic cells and promoting proliferation and differentiation (cell changes from one to another) of other lymphocytes (Cruse and Lewis 1999).

T cells

T lymphocytes originate in the bone marrow but mature in the thymus. They are responsible for the cell mediated immune response and control of the immune response. This kind of immune response does not involve antibodies, but includes the activation of cells that ingest foreign particles and T cells, and release of various signaling proteins (explained in the section "cytokines") to deal with the threat instead (Sharma 2019). T cells engage in controlling the immune responses by providing specific cells that are capable of helping or suppressing these responses (Cruse and Lewis 1999). T cells need to be activated in order to carry out their function. The T cells can be distinguished from each other by the antigen receptors that are present on the surface of said cells.

Helper T cells (CD4+)

Helper T cells are a type of T lymphocytes which are critical for induction of an immune response to a foreign antigen by creating lymphokines (a type of cytokines) which play a central role in the activation, control and coordination of cytotoxic T cells and variety of other cells that participate in the immune response (Kuby 1992).

They are divided into two subpopulations - Th1 and Th2. They have different growth characteristics, respond to cytokines in different ways and produce different Interleukins. Th1 secretes Interleukin-2 which effects the activity of T_c cells (Tizard 1992).

Activation:

T_h cells are activated when a ligand (cell surface protein) binds to its receptors. The binding is followed by clustering and internalization of the receptor – the receptor moves to the inside of the cell. An exposure to an antigen for 30 – 60 minutes suffices to make the cell leave the first phase and enter the second phase; the T_h cell, however, stops (unless it receives a second signal in the form of Interleukin-1) (Klein 1990).

Function:

Antigen is presented by an antigen presenting cell such as the macrophage. Once activated the T helper cells express Interleukin-2 and produce Interleukin 2 molecules, which act in an autocrine fashion by combining with the Interleukin-2 receptors and stimulating T helper cells to proliferate. Differentiated T helper cells synthesize and secrete lymphokines that affect the function of other cells of the immune system such as T_c cells, B cells and NK cells. (Cruse and Lewis 1999)

Cytotoxic T cells (CD8+)

Cytotoxic T cells are a type of lymphocytes which, in contrast to T_h, do not secrete many cytokines and instead exhibit cell destructing activity. They have a vital function in monitoring the cells of the body and eliminating any that display antigens, such as virus-infected cells, tumor cells and cells of a foreign tissue transplant (Kuby 1992). They recognize antigens through the T cell receptor on cells of the host infected by viruses. Death of the target cell occurs a few hours later after recognition and binding to it. (Cruse and Lewis 1999)

The T_c cell is activated by interaction with an antigen on the surface of an infected cell in the presence of appropriate cytokines, once activated it kills the infected cell. (Kuby 1992)

T_c cell mediated lysis (as seen in Figure 1) is a major mechanism for the destruction of virally infected cells. If activated during the period in which the virus/cancer is present but undetectable, T_c cells may

be capable of eliminating the virus/cancer and curing the host with relatively limited cell destruction. On the other hand, vigorous T_c cell activity after a virus has been widely distributed may lead to substantial tissue damage because of the large number of infected cells. Thus, a large number of cells may potentially be killed by the action of T_c cells (Paul 1993).

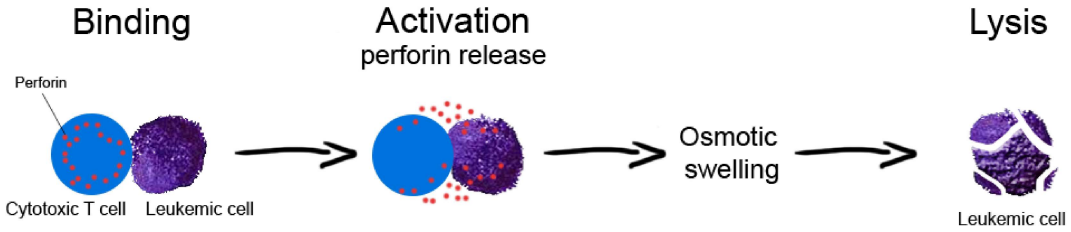


Figure 1. Destruction of the leukemic cell by a cytotoxic T cell.

Activation:

T_c cells do not require Interleukin-1 for their activation. They are, however, fully dependent on Interleukin-2 for their growth. Some T_c cells can produce Interleukin-2 but they are in the minority, so they are dependent on the Interleukin-2 supplied by the T_h cells (which is one of the main functions of T_h cells) (Klein 1990).

When T_c cells get activated by an antigen, they develop into T_c and T_h cells or memory (T_{cm}) cells. T_{cm} cells act as a reservoir for producing specialized T_c and T_h cells, both of which are able to migrate to inflammatory sites (Lydyard, Whelan, and Fanger 2011).

Function:

T_c cells secrete lymphokines that attract other lymphocytes to the area and release perforins (specific proteins) that produce ion channels in the membrane of the target cell leading to cell lysis (destruction of the cell). They play a significant role in tumor immunity (Cruse and Lewis 1999).

Regulatory T cells (suppressor T lymphocytes)

T_{reg} cells are a T lymphocyte subpopulation that downregulates the ability of T lymphocytes to mount a cellular immune response. An overall immune response may be a consequence of the balance between T_h and T_{reg} stimulation. They are significant in the establishment of immunologic tolerance and are particularly active in response to unprocessed antigen, meaning that the response is weaker if an antigen has been interacted with. T_{reg} cells are able to suppress the immune reactivity of other cells and are regulated by the interaction of other T_h and T_{reg} cells (Tizard 1992; Cruse and Lewis 1999).

T_{reg} specialize in the regulation of T_h and T_c cells. The T_{reg} cells are believed to have both CD8 (T_c) and CD4 (T_h) antigen receptors, however this fact makes them almost indistinguishable from the T_c (CD8 antigen) cells and T_h cells (CD4 antigen), except that the former is believed to lack cytotoxic activity and also for the functional difference between them (Klein 1990).

Nonspecific suppression – occurs when T cells stimulated by one antigen inhibit the response of lymphocytes to nearly all antigens, including the one that originally induced the immune response (Klein 1990).

Specific suppression – occurs when T_{reg} cells that have been activated by a particular antigen inhibit the response of other T cells to this but not to other antigens. They are supposed to act by way of soluble specific suppressor factors (Klein 1990).

1.3 Myeloproliferative Neoplasms

Myeloproliferative Neoplasms (MPNs) are types of blood cancer that disturb the production of blood cells in the bone marrow. MPN can be classified as Polycythemia vera (PV), essential thrombocytosis (ET) or myelofibrosis (MF). PV is characterized by an increased count of white and red blood cells and platelets in the blood tests. ET patients show a higher number of platelets in their blood stream. MF is a disturbance in the production of blood stem cells in the bone marrow that ultimately leads to a decreased count of blood cells and scarred bone marrow. (Silberstein and Anastasi 2017)

Connection to JAK2-V617F

Patients diagnosed with these diseases mostly show mutations in hematopoietic cytokine signaling pathways in blood stem cells. One important intracellular communication tool is the Janus Kinase – Signal Transducers and Transcription (JAK-STAT) pathway (Gadina 2013). Especially Janus Kinase 2 (JAK2) – one of the four existing subgroups of Janus Kinases in the human body (McLornan, Percy, and McMullin 2006) and key component of the cell growth (Kundranda, Tibes, and Mesa 2011)– plays an important role in MPN (Gadina 2013). If JAK2 carries the so called V617F mutation, it is considered to be the so far most related factor in leading to and progressing MPN. Studies from 2005 show that all three classic disorders of MPN are linked to mutations in the JAK2 gene. In 81-99% of all patients with PV, 41-72% of the patients diagnosed with ET and 39-57% of MF patients the JAK2-V617F mutation was present (Abdel-Wahab and Levine 2010).

Connection to chronical inflammation

A Swedish epidemiological study has recently shown that the JAK2-V617F mutation is also a factor in chronical inflammatory diseases. Inflammatory diseases such as Crohn's disease (an inflammation of the gastrointestinal tract) were discovered to likely develop in patients with PV, ET and MF. Also, a history of any type of autoimmune disease has been linked to patients being at much higher risk of being diagnosed with MPN (Hasselbalch and Bjørn 2015).

1.4 Interferon therapy

Leukemia can be treated in many ways. The model of this project uses data from patients that were treated with interferon therapy by inducing the immune response.

The interferon protein

Interferons are a subgroup of cytokines. Their main responsibility is to interfere with the replication of viruses within a cell. Interferons (IFN) can be divided into type 1 and type 2. Type 1 includes IFN- α and IFN-β. Type 2 is represented by IFN-γ, which appears naturally in the body. IFN-γ has many properties, but its primary function is to regulate microenvironments. (Williams 2012)

Type 1 interferons have powerful antiviral properties. Besides this, type 1 interferons have shown to work as anti-tumor proteins, in part by improving the expression of antigens in connection to tumors. (Williams 2012)

Interferon therapy in essential thrombocythemia

IFN-α therapy aims to enhance the immune response (Williams 2012) and is used as treatment for several diseases including ET (Kiladjian, Chomienne, and Fenaux 2008).

The mechanism behind the IFN-α therapy is unknown, and the speculations are typically complex. IFN-α can antagonize growth factor obtained by platelets, which may play a part in the treatment (Kiladjian, Chomienne, and Fenaux 2008). Furthermore, all interferon types have demonstrated to

increase the expression of an MHC class 1 (Kuby 1992), which is a set of genes triggering immune response that signals T cell activation (Williams 2012).

Despite absent knowledge of the process behind the therapy, studies have shown IFN-α therapy to be an effective form of treatment with an average response rate of 84% for patients with ET (Kiladjian, Chomienne, and Fenaux 2008).

Chapter 2: Mathematical Theory

This section aims to provide the background knowledge on mathematical modeling and basic mathematics necessary to understand the model's equations.

2.1 Introduction to modelling

Mathematical modelling is a way to translate hypotheses about how systems in the real world behave into mathematics. However, most systems in the world are far too complicated to be able to include every aspect of them. Therefore, compromises must be made in order to simplify the system we would like to model. This is achieved by focusing on the most important parts of the system and purposefully excluding the rest, which naturally bears the risk of oversimplification.

Since the underlying assumptions and ideas are the basis for a mathematical model, modelling makes us formulate and question our existing beliefs about the system to be modelled. Another beneficial aspect of using mathematical models is the fact that they allow us to run calculations on computers to determine the effect of changes in the system. Mathematical models can thereby improve scientific understanding and help in decision making. (Marion and Lawson 2008)

2.2 Types of mathematical models

There are different types of mathematical models. Depending on the nature of the outcome a model predicts, it can be either deterministic or statistical. In a deterministic model, the outcome will be the same if all initial conditions are the same, whereas a statistical model includes random variation and therefore results in a distribution of different possible outcomes. (Marion and Lawson 2008)

In addition to this division by the type of outcome, we can differentiate between mechanistic models and empirical models. The basis of a mechanistic model is the theoretical information about the mechanisms behind the processes that are being modeled, while an empirical model does not consider these mechanisms and instead only focuses on the quantitative aspect (Marion and Lawson 2008). An example for such an empirical model could be any calculation of a regression equation, since this shows the functional correlation between the parameters, but ignores the specific underlying mechanisms and therefore does not show the reason for the correlation.

2.3 Modelling in biology and the medical field

Generally, mathematical models are utilized to generate and validate hypotheses, but in biological and medical research especially, they are also used to make predictions or perform further experiments. Mechanistic models are preferable in biology and medicine, since they aim to understand and simulate the biological processes occurring in the body, which is required to accurately understand a disease and predict its progression for example. Another advantage of mechanistic models in biology is the fact that it is possible to ensure that natural laws such as conservation laws are not violated, since they can also be included in the model.

To evaluate a model's validity, experimental data is usually applied to fit parameters. However, there are cases where a model acts as a "thought experiment" and no experimental data is used (Hernandez-Vargas and Sanchez 2019).

2.4 Differential equations

A differential equation is an equation that contains derivatives of a function with respect to one or more variables. Since they show the change of a function with respect to the function, differential equations are used to describe various deterministic systems in science and economics (Hernandez-

Vargas and Sanchez 2019). Using differential equations, it is possible to model universal laws or general processes, while being able to make specifications by applying different initial conditions. Since they show the change of a function with respect to the function, differential equations are used to describe various deterministic systems in science and economics (Hernandez-Vargas and Sanchez 2019)(Hernandez-Vargas, Esteban A.; Sanchez 2019). Using differential equations, it is possible to model universal laws or general processes, while being able to make specifications by applying different initial conditions. If the initial conditions are known, differential equations can then be used to make predictions for deterministic systems (Adams 2018).

Differential equations can contain derivatives with respect to any number of variables. If they only involve derivatives of **one** variable, they are classified as **ordinary differential equations** (ODEs), whereas **partial differential equations** (PDEs) contain derivatives with respect to **several** variables. Another characteristic to differentiate between differential equations is their order; it is determined by the order of the highest order derivative in the equation (Adams 2018).

For modeling dynamical systems in biology, first order ODEs of the following general form are commonly used (Hernandez-Vargas and Sanchez 2019):For modeling dynamical systems in biology, first order ODEs of the following general form are commonly used (Hernandez-Vargas, Esteban A.; Sanchez 2019):

$$\frac{dx}{dt} = f(x,t) \qquad (1)$$

Here, $\frac{dx}{dt}$ corresponds to the change in variable $x(t)$ over time t. When dynamical systems represented by ODEs reach a point for which the partial derivative with respect to time is 0, the system will not change and will therefore remain at this value $x = x^*$ forever. These points are called equilibrium points, because the system will remain in a steady state, once x^* is reached (Adams 2018). Equilibrium points are important for studying and analyzing a model, since they can provide new insights about the system's behavior.

2.5 The modelling cycle/ process

The process of modelling according to (Marion et al. 2008) consists of four main stages. However, should any flaws be found in our model during the later phases, earlier stages must be returned to in order to correct these weaknesses. By continuously following this method of repeated iteration, modelling becomes a circular process, as visualized in Figure 2. The four main stages are: building, studying, testing and use.

1. **Building**

It is prerequisite that the goal of the model is clear at the beginning to be able to make reasonable **assumptions** and **delimitations**, which depend on the purpose of the model. Based on these, a **conceptual model** is usually created, which translates the system's parts and their interactions to a visual representation like a flow diagram, for example. Then, suitable **mathematical equations** describing the interactions

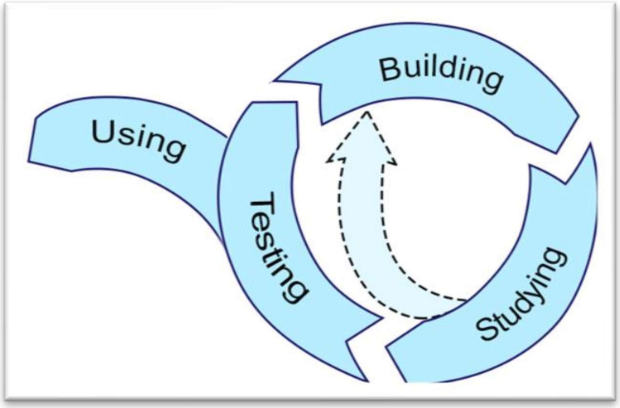


Figure 2. The modelling cycle

shown in the conceptual model are formulated and solved, either analytically or numerically (Marion and Lawson 2008).

2. Studying

In this phase, the model is analyzed both qualitatively and quantitatively. For a qualitative description of the model, equations are often re-written in terms of dimensionless quantities, which decreases the number of parameters. This may reveal that that the dynamics are governed by clusters of fewer parameters and it also facilitates comparing different model types directly. Another qualitative feature of models to be analyzed is long-term behavior. The modelled system might approach an asymptote, which corresponds to an equilibrium, since there is no change in value once the asymptote is reached. Alternatively, the system might show oscillating behavior, characterized by recurrence at regular intervals, or, besides these possibilities, the system may behave completely irregularly.

Studying the model quantitively involves altering the parameters to examine how the outcome of the model changes accordingly. This method is known as sensitivity analysis and can aid in detecting soft spots of the model. Additionally, the model output is analyzed quantitatively using computers for repeated evaluation or by performing experiments, whose results can be used to estimate outcomes for other conditions (Marion and Lawson 2008).

3. Testing

The next step is to assess the model's validity by testing the underlying assumptions, assessing the parameters, and making predictions. To clarify, a model is considered valid if it has not been falsified, since a perfect validation of a model is seldom possible.

To verify the assumptions of the model that we might be skeptical about, they should be relaxed individually to be able to investigate if the model is still representative of reality. This also includes assumptions about the structure of the model, since those could be inaccurate, as well.

It should be noted that there is always a discrepancy between model predictions and observed data due to natural variability and measurement error, but errors might also arise from the exclusion of important factors, invalid assumptions, and errors in parameter estimates. Testing the model predictions should be done using a data set that was not used for parameter estimation, since not doing so could result in perceiving a false accuracy of predictions. It is also important that the data used for validation covers the whole range of cases which the model is supposed to represent. Summary statistics are useful to show the disparity between the predictions and the observed data, and thereby serve as a measurement of model performance.

Most commonly, the parameters are evaluated by minimizing the mean square error to find the best set of parameters. Another possibility is to compute their maximum likelihood estimates, which is the certain parameter set with the greatest probability of predicting the observed data.

Additionally, different models are compared to decide which is the best and should be used for application. To guide this decision, the aspects of the testing of a single model are used and compared: the assumptions, the effect of exclusions, model predictions and their accuracy, as well as the simplicity. In case of similar predictions, the simpler model might be better, while combining the predictions in a model averaging process might be beneficial if the predictions differ.(Marion and Lawson 2008)

4. Use

Finally, the model is applied to solve problems in the real world. However, it is important that, instead of simply using the predictions of the model, their precision is also taken into consideration. Estimates for the predictions' precision and accuracy were ideally already acquired during the studying or testing phase (Marion and Lawson 2008). Furthermore, any limitations of the model need to be discussed and should be specified clearly to prevent others from misusing the model, since they highly depend on the assumptions and delimitations established during the building phase.

Chapter 3: Method and data

This section aims to explain the process of our project and how we proceeded to be able to answer our research question, as well as to present the data used in the project. We illustrate how we adapted the modelling cycle described in the previous chapter to adjust it to our specific project. Additionally, a visual representation and short description of the data is given in this chapter.

3.1 Building process

Our process of creating this model is fairly similar to the process described by (Marion et al. 2008) however with small alterations.

Research

One stage we believe is worth adding is an initial *research.* In the research phase we collected the needed information regarding the interaction between the leukemia cells and T cells, which is a necessity to create a mechanistic model, as this attempts to describe the actual process behind the modelled system.

Building

Based on the acquired knowledge and stated assumption we started setting up the equations and one by one adding terms each consisting of one or more parameters representing numerical value(s) describing the "strength" of the variables' interactions.

Fitting

Besides research we split a part of the "Building" stage described in the Mathematical theory sections into a fitting stage. This stage consists of manually adjusting our parameters one at a time to make the graph of our model fit better to the data, which is described in the section below. First, we adjusted parameters to approximately fit the first data points for each kind of cell reflecting disease and immune system before the treatment. Our primary focus was on the number of cancer cells. When this was done, we tried to change one or two parameters at a time to decide which parameters must be tailored to represent the effect of the treatment. Since the actual effect of the treatment is unknown, we regularly reconsidered which parameters should change as well as how much.

Occasionally while altering the parameters, we experienced that the model would not steer towards our desired curve, which caused us to go back to the building phase and reconsider the structure of the model.

Studying

When we felt sufficiently confident with our model and parameters, we continued to study and analyze it. We did a statistical analysis where we compare our model to one of the simplest models:

the mean model, which is a model predicting the next outcome with a constant value determined by the history of data. We also did a sensitivity analysis, allowing us to discuss what parameter an ideal treatment could affect.

Testing
We tested our model on data from two other subjects. We are using summary statistics to evaluate how representative our model is for other cases. Just like in the studying phase, we analyzed how well the model fits with the data from two other patients.

Use
This is not a stage we have covered, but this would be the ideal result. How and how well the model could be used in the medical field is discussed later in this report.

Figure 3 visualizes our modelling process. It shows how we start by researching, then start building our model and afterwards fit parameters, which would sometimes lead a return to the building phase shown by the gradient arrow appearing from the "Fitting" arrow. When confident in the model, we would continue to *studying* and *analyzing*.

The *bright green dotted arrow* shows how studying could lead back to building to achieve a more precise model, which was not within our timeframe.

The *bright blue* dotted line shows that the model would ideally be used in the real world, but has not been done in our report, and we are discussing its ability to do this.

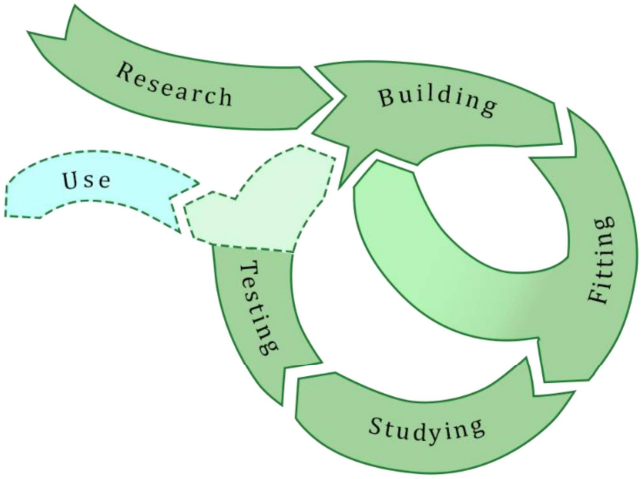


Figure 3. Our process in creating the model.

3.2 Data description

To validate our model and to adjust our parameters we used medical data from three different treated patients and one untreated patient, which helped us to understand the rise and decline of T_c and T_h cells during the same time period as cancer cells, so we could clearly see how the specific types of T cells react to the cancer or how the cells that partake in the immune response depend on each other. The data was collected at Rigshospitalet and a lot of different measurements were done, but we decided to analyze and display only the most relevant ones for our model. They start the measurements as they begin the treatment and then took new measurements every 3 months for 12 months. They measured the number of leukocytes which they then used to calculate the JAK2 percentage of all leukocytes, number of lymphocytes, T_h cells, T_c cells and cancer cells.

3.3 Data in tables

Patient 1 [months]	0	3	6	9	12
T_h cells (CD4+)					
cell count [10^9/L]	0.35	0.17	0.18	x	0.29
T_c cells (CD8+)					
cell count [10^9/L]	0.43	0.11	0.12	x	0.13
Cancer cells (N)					
cell count [10^9/L]	1.63	0.52	0.17	x	0.23

Table 1. Blood test results for patient 1 containing Th, Tc and cancer cell count in billion cells per liter.

Patient 2 [months]	0	3	6	9	12
T_h cells (CD4+)					
cell count [10^9/L]	0.68	0.68	0.54	x	0.76
T_c cells (CD8+)					
cell count [10^9/L]	0.16	0.12	0.14	x	0.10
Cancer cells (N)					
cell count [10^9/L]	1.56	1.63	0.56	0.56	1.06

Table 2. Blood test results for patient 2 containing Th, Tc and cancer cell count in billion cells per liter.

Patient 3 [months]	0	3	6	9	12
T_h cells (CD4+)					
cell count [10^9/L]	0.30	0.16	0.27	x	0.19
T_c cells (CD8+)					
cell count [10^9/L]	0.51	0.31	0.28	x	0.25
Cancer cells (N)					
cell count [10^9/L]	4.87	3.78	4.37	3.88	3.71

Table 3. Blood test results for patient 3 containing Th, Tc and cancer cell count in billion cells per liter.

In Table 1 we have our data from the first patient, which we used to set our parameter values. Table 2 contains the data from the second patient and Table 3 has the third patient's data. These tables have all the different types of cells that we used in our model and they are counted in billions per liter. The months are based on time since treatment start.

Visual representation of the data

In the Figure 4 and Figure 5 we visualized the given data by plotting JAK2 percentage against time. Here JAK2 percentage is given by the ratio of leukemic cells to all leukocytes multiplied by 100:

$$JAK2 = \frac{leukemic\ cells}{all\ leukocytes} \cdot 100.$$

Untreated patient

Plotting the JAK2 percentage of all leukocytes over time (shown in Table 4), we found the best fitting exponential regression curve (blue curve) shown in Figure 4, therefore we can conclude that in an untreated patient the mutated cells grow almost exponentially, which corresponds well to a standard population growth.

Years	JAK2[%]	Years	JAK2[%]
0	0.7	3.89	12.1
0.54	4	4.55	22
0.61	0.9	4.89	19
0.81	1.3	4.99	30
1.17	2.2	5.18	31
1.51	6	5.24	12.1
2.72	2.5	5.32	32
2.95	7.9	5.37	31
3.1	7.9	5.74	21.8
3.35	8.7	5.79	30
3.45	6.9	5.85	31
3.62	10	X	X

Table 4. JAK2 percentage of all leukocytes in the untreated patient.

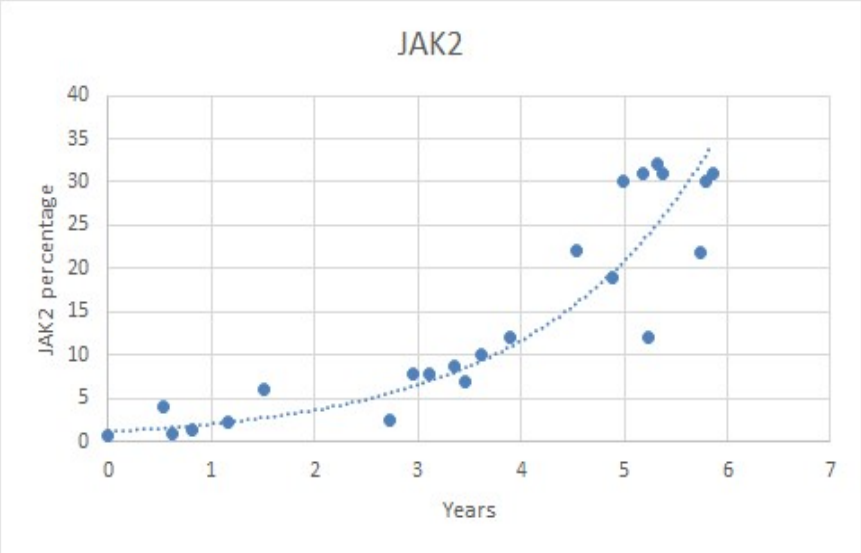


Figure 4. Plotting JAK2 percentage of all leukocytes of untreated patient over time in years. Blue line represents the best fitting regression line.

Treated Patient

In Figure 5 mutated cells rapidly drop only in a few months once we start the treatment.

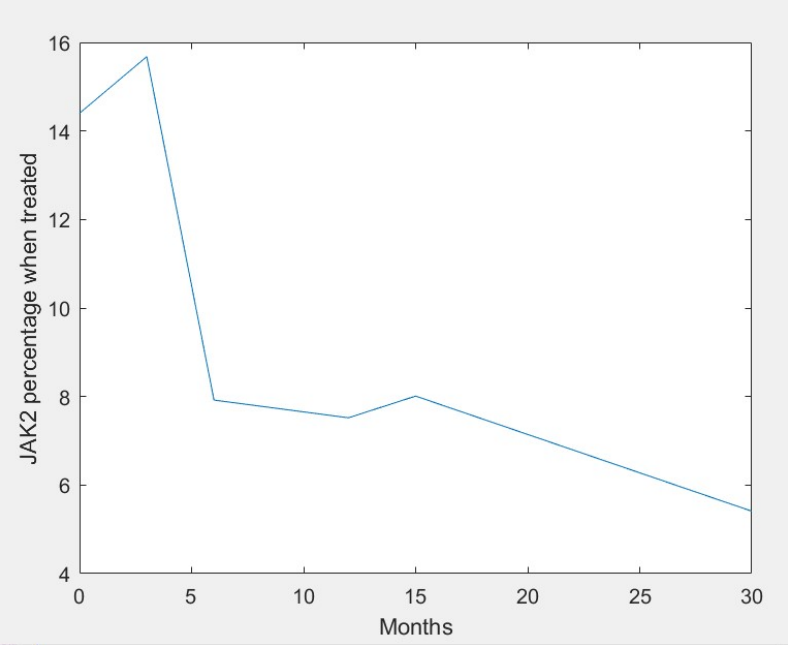


Figure 5. JAK2 cell percentage of all leukocytes over months during treatment.

Tendencies

In Figure 6, Figure 7 and Figure 8 we tried to find the tendencies which the data from all three patients follow.

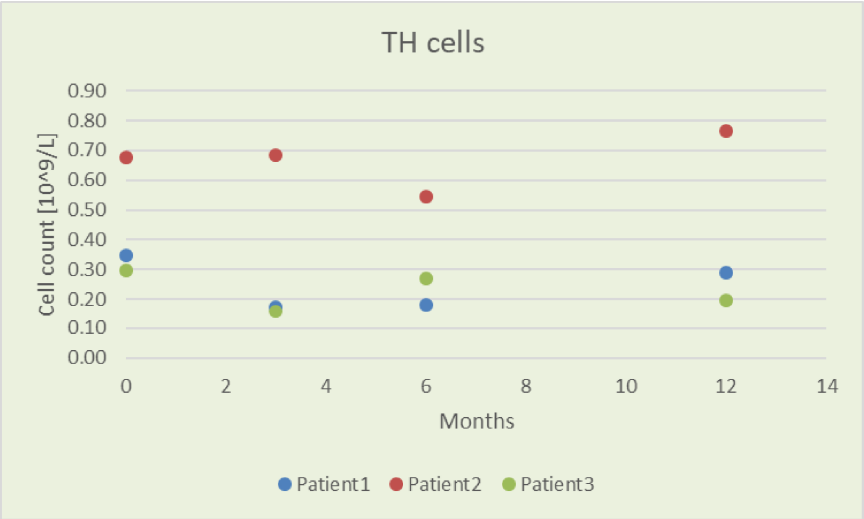


Figure 6. Count of Th cells in billion cells per liter in three different patients over 12 months.

In Figure 6, Patient 1 and Patient 3's data points are almost the same and they follow the same pattern – the T_h cell count declines. Patient 2 datapoints don't seem to have any similarities with the other patient's data. We lack data from the ninth month.

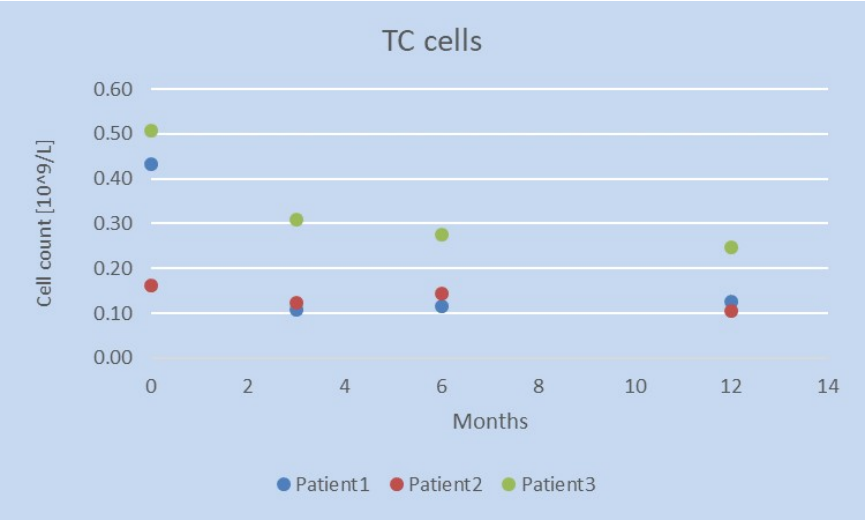


Figure 7. Count of Tc cells in billion cells per liter in three different patients over 12 months.

In Figure 7 the cell count of T_c cells declines in every patient, but Patient 1 and Patient 3 data follow the same pattern with a significant elbow at the third month. We lack data from the ninth month.

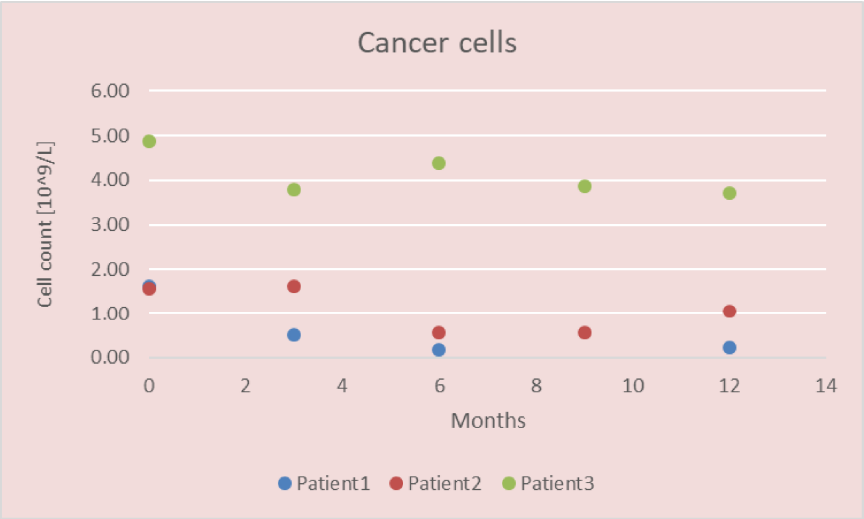


Figure 8. Count of cancer cells in billion cells per liter in three different patients over 12 months.

In Figure 8 the points of all patients tend to decline, however not dramatically, that would be the ideal state. Data from the Patient 1 is lacking in the ninth month.

Chapter 4: The model

We are aiming to develop a deterministic mechanistic mathematical model.
Our model should consider the biological processes occurring in the body, which makes it a mechanistic model, as opposed to empirical models. The latter can be useful to show correlation, but they do not provide information about causation (Marion and Lawson 2008). We want to suggest a mechanistic model instead, because we would like to contribute to enhancing existing knowledge about the interaction of T cells and leukemia cells, and how Interferon treatment affects this.

4.1 Assumptions and delimitations

The underlying ideas and assumptions are the basis for a mathematical model and depend on the model's purpose. They need to be included to be able to simplify the system that we are trying to model, since the whole immune system is far too complex to be modeled in its entirety (Hernandez-Vargas, Esteban A.; Sanchez 2019)("Mathematical Modeling Principles" 2019)("Mathematical Modeling Principles" 2019)("Mathematical Modeling Principles" 2019)("Mathematical Modeling Principles" 2019)(Hernandez-Vargas and Sanchez 2019).

Due to its precise nature, mathematics requires these assumptions to be clearly formulated, which in turn requires us to review and reconsider said assumptions before implementing them (Marion and Lawson 2008). However, some assumptions made at the beginning may not be valid. Therefore, we need to continue reexamining our assumptions if the model is inaccurate or not representative of the patient data (Hernandez-Vargas, Esteban A.; Sanchez 2019)("Mathematical Modeling Principles" 2019)("Mathematical Modeling Principles" 2019)("Mathematical Modeling Principles" 2019)("Mathematical Modeling Principles" 2019)(Hernandez-Vargas and Sanchez 2019).

These assumptions were made:

1. **General assumptions**
   To align our model with its purpose as illustrated by the research questions some general assumptions and delimitations are necessary.
   a. *The body's reaction to leukemia can be reduced to the interaction of T helper cells, cytotoxic T cells, and regulatory T cells with the cancer cells.*
      We are purposefully excluding any other cells that partake in the immune response to simplify the model and only focus on T cells' interaction with cancerous cells.
   b. *The dynamic change in number of cells can be described using differential equations as deterministic, mechanistic sub-models.*
      Ordinary differential equations can describe presumably deterministic and dynamic systems (Hernandez-Vargas, Esteban A.; Sanchez 2019). In our case, first order ordinary differential equations are suitable, because we assume that the number of cells affect each other's rate of change over time. This corresponds to taking the first derivative with respect to time.
   c. *The cell types interact and affect each other, so their corresponding differential equations together build a simplified system model of the immune response to leukemic cells.*
      Based on our current biological knowledge the different T cell types interact with each other in different ways, creating the body's immune system. By incorporating cancer cells into the model, their interaction with leukemic cells is included.
2. **Assumptions about cancer cells**
   a. *The cancer cells grow logistically.*

The data from the untreated patient suggests exponential cancer growth. We adjusted this to logistic growth because this is the conventional way of describing population growth more realistically since it also accounts for the crowding effect which arises when a population exists in a limited space (Marion and Lawson 2008).

b. *Cancer cells are killed by cytotoxic T cells.*
Since the function of cytotoxic T cells is to monitor the body's cells and eliminate any cells that display foreign antigens, they kill tumor cells due to their cells destructing activity (Kuby 1992).

c. *Treatment enhances killing of cancer cells by cytotoxic T cells.*
We assume that Interferon treatment increases the T_c cells' effectiveness, which means they are more likely to be successful at killing a tumor cell they meet.

3. Assumptions about T cells

a. *T helper cells and cytotoxic T cells are also produced when there are no cancer cells in the body.*
The immune system is not only responsible for fighting cancer, but any kind of intruder (Lydyard, Whelan, and Fanger 2011), therefore its cells need to have a baseline production, even when the person does not have cancer.

b. *T helper cells are activated by cancer cells.*
T_h cells are activated when a ligand (cell surface protein) binds to its receptors (Klein 1990). These antigens are present on leukemic cells.

c. *T helper cells die naturally.*
Cells do not live an endless amount of time, but die after some time period.

d. *T helper cells stimulate the production of cytotoxic T cells*.
Interleukin-2 is vital for the long-term growth of T_c cells (Kuby 1992) and is mainly produced by T_h cells, so T_c cells depend on the T_h cells' supply of Interleukin-2 (Klein 1990).

e. *Cytotoxic cells are killed by cancer cells.*
We assume that there is a certain chance that the cytotoxic T cells die when they meet a cancer cell, instead of killing the leukemic cell.

f. *Regulatory T cells decrease the number of T Helper cells without treatment.*
The natural behavior of regulatory T cells is to suppress the immune reactivity (Tizard 1992) by regulating T_h and T_c cells (Klein 1990).

g. *Regulatory T cells decrease the number of cytotoxic T cells when treatment begins.*
We are assuming that Interferon treatment affects the regulatory T cells in such a way that they shift their attention from decreasing T_h cell count to decreasing T_c cell count in order to prevent the cytotoxic T cells from killing more cells than necessary. This aims to minimize the danger of severe tissue damage through vigorous T_c cell activity (Paul 1993).

h. *Regulatory T cells are stimulated by cytotoxic T cells and die naturally.*
Depending on the number of T_c cells the regulatory T cells are produced to prevent the cytotoxic T cells from killing too many cells, thereby regulating the immune reactivity (Tizard 1992). Just like the T helper cells, regulatory T cells die after some time period.

i. *Interferon treatment increases the effectiveness of cytotoxic T cells.*
As explained above, we assume that Interferon treatment increases the rate of leukemic cells being killed by cytotoxic T cells by raising T_c cells' effectiveness. This corresponds to assumption 2.c.

j. Interferon treatment increases the baseline production of T helper cells and cytotoxic T cells.

Raising the production of T_h and T_c cells by introducing Interferon treatment corresponds to an enhanced immune response, so the cancer can be defeated.

4.2 Conceptual model

The conceptual models in Figure 9 and Figure 10 show the four cell types that we are focusing on (cancerous cells, cytotoxic cells, regulatory cells and helper cells) and the interaction between them. Upregulation of cells is denoted by parameters *k*, while parameters *d* describe downregulation of cells.

Without treatment

The mutated cells constantly reproduce, and their production increases the creation of new helper cells. They also kill cytotoxic T cells. Death of cancerous cells is accounted for by using logistic growth.

Helper cells do reproduce on their own, but this process is greatly accelerated by the presence of cancer cells. Their main function is stimulation of cytotoxic cell production. Helper cells die naturally, but in a case without treatment regulatory T cells also greatly induce their death.

New cytotoxic cells constantly develop, and this process is induced by the helper cells. Cytotoxic cells die naturally, and some of them are also killed by the malignant cells. Cytotoxic cells are the only contributor to the death of cancerous cells (in our model). The number of cytotoxic cells determines the extent of the production of regulatory T cells.

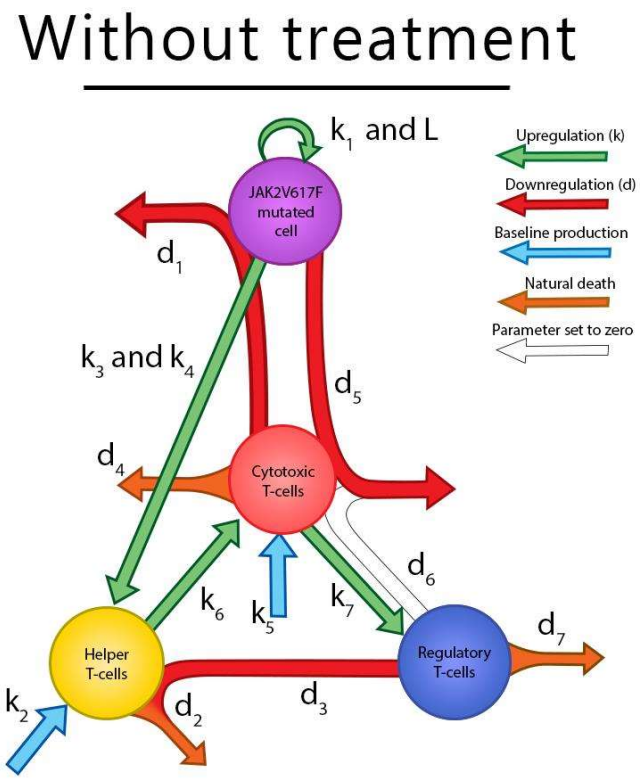


Figure 9. The conceptual model of the immune system interaction with cancer for a patient without treatment.

Regulatory cells suppress the immune response by inducing T helper cell death. The production of regulatory cells is mostly actuated by the presence of cytotoxic T cells. Regulatory cells die naturally.

This whole interaction can be explained by our conceptual model depicted in Figure 9. The figure shows green arrows representing the upregulation by one cell type to another. Red arrows represent downregulation by one cell type to the cell they touch as it continues to point away from the downregulated cell, which resembles the death of this cell. Blue arrows represent a baseline production, and the orange arrows shows the natural death. The figure also shows the white arrow from the regulatory cells downregulating the cytotoxic cells. This represents a parameter which is set to zero for the model before treatment, and the effect of the parameter is shown in the conceptual model for when treatment is started, which is shown in Figure 10.

With treatment

There are some key differences in the model with treatment.

The introduction of treatment makes regulatory T cells switch from killing helper T cells to killing cytotoxic cells.

This is shown in the conceptual model Figure 10 as the arrow between the regulatory cells and helper cells changing to white, and the one between regulatory cells and cytotoxic cell turning red.

Three parameters get increased with the introduction of treatment, as we assume that treatment enhances the production of cytotoxic and helper T cells, and also increases the destructive effect of cytotoxic cells on malignant cells.

The effects of these changes are strong enough for the body to start battling cancer intensely enough to defeat it. This is shown with a thicker outline of the arrows in the conceptual model.

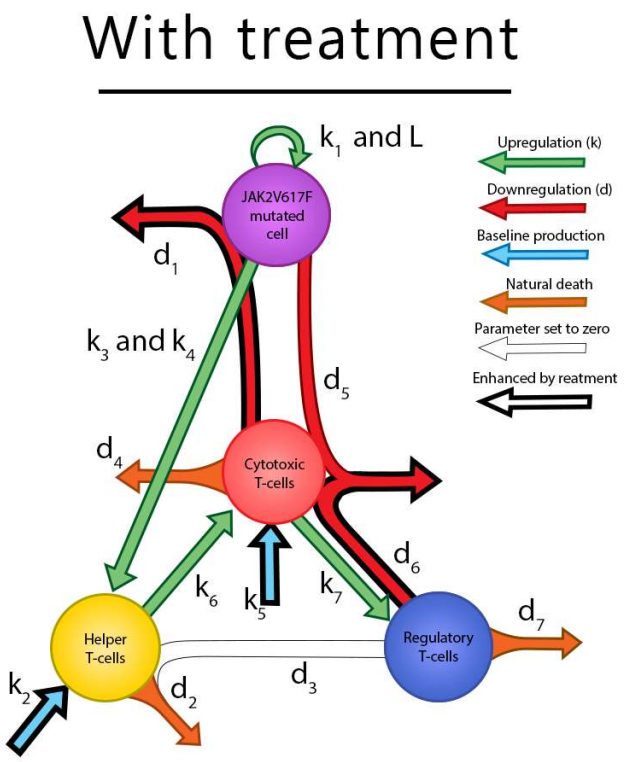


Figure 10. The conceptual model of the immune system interaction with cancer for a patient with treatment.

4.3 Parameters

Our differential equations include 14 parameters, all included in Table 5, that represent numbers which express growth and decrease rates of and certain relationships between cell populations. The letter 'k' stands for a positive change in cell count whereas 'd' stands for a negative change in cell count.

Five out of those parameters change their value after treatment starts. These parameters include d_1, k_2, d_3, k_5 and d_6.
d_1 gets increased due to the immune enhancement by the IFN-α treatment. T_C cells kill mutated cells more efficiently.
k_2 gets increased because the rate of T_H cell production gets enhanced by the treatment.
d_3 is set to zero after treatment start because we assume that the rate of T_{reg} cells killing T_H cells is approaching zero because of the enhancement of the immune response by the IFN-α treatment.
k_5 gets increased because the rate of T_C cell production gets enhanced by the treatment.
d_6 is set to zero until treatment starts because, unless IFN-α treatment is involved, we assume that the T_C cell production doesn't get suppressed by T_{reg} cells.

In order to find a set of parameters that would make our model fit to the data we went through a process of 'trial and error'. This means that we changed the parameter values in our MATLAB-code over and over again one by one until we found a set that would satisfy our expectations.

We learned that little changes can have big influence in the final outcome and vice versa. Chapter 5.3 focuses on this aspect by performing a sensitivity analysis of our parameters.

	Parameter	Function	Value		Unit
			Without treatment	With treatment	
Cancer	**k_1**	cancer growth rate	0.0015		$days^{-1}$
	d_1	rate of T_c cells killing cancer cells	$1.44 * 10^{-4}$	0.0437	$(10^9 * cells)^{-1}/liter^{-1} * days^{-1}$
	L	limiting capacity	7		$10^9 * cells/liter$
Helper	**k_2**	baseline production of T_h cells	0.0014	0.0028	$10^9 * cells/liter * days^{-1}$
	k_3	rate of cancer cells stimulating T_h cells	0.0063		$days^{-1}$
	k_4	threshold value	0.7		$10^9 * cells/liter$
	d_2	natural death rate of T_h cells	$3.0 * 10^{-4}$		$days^{-1}$
	d_3	rate of T_{reg} cells killing T_h cells	0.0096	0	$(10^9 * cells)^{-1}/liter^{-1} * days^{-1}$

Cytotoxic	**k_5**	baseline production of T_c cells	$2.0 * 10^{-6}$	$4.4 * 10^{-6}$	$10^9 * cells/liter * days^{-1}$
	k_6	rate of T_h cells stimulating T_c cells	0.0019		$(10^9 * cells)^{-1}/liter^{-1} * days^{-1}$
	d_4	natural death rate of T_c cells	$2.0 * 10^{-4}$		$days^{-1}$
	d_5	rate of cancer cells killing T_c cells	$1.0 * 10^{-6}$		$(10^9 * cells)^{-1}/liter^{-1} * days^{-1}$
	d_6	rate of T_{reg} cells killing T_c cells	0	$2.4 * 10^{-4}$	$(10^9 * cells)^{-1}/liter^{-1} * days^{-1}$
Regulatory	**k_7**	rate of T_c cells stimulating T_{reg} cells	0.005		$days^{-1}$
	d_7	natural death rate of regulatory T cells	$1.5 * 10^{-4}$		$days^{-1}$

Table 5. All the parameters used in the model. Any parameters k have an increasing effect on cell count, while any parameters d have a decreasing effect on cell count.

4.4 Equations

We assume that the immune response is a deterministic system, which implies that it can be predicted if all necessary information is known (Adams and Essex 2018). As a presumably deterministic and dynamic system the immune response can be described by using ordinary differential equations (Hernandez-Vargas and Sanchez 2019).
The following *first order* ordinary differential equations describe the changes in cell count over time for each of the cell types.

Cancer cell equation

$$\frac{dN}{dt} = \overbrace{k_1 N \left(1 - \frac{N}{L}\right)}^{\text{Cancer production}} - \underbrace{d_1 N\, T_c}_{\text{Killed by } T_C} \qquad (2)$$

The number of cancer cells is represented by N (in billion cells per liter).
Generally, the change in cancer cells over time, $\frac{dN}{dt}$, consists of a term for cancer production and one for cancer cell death.
Based on assumption 2.a, the term for cancer production follows the general form of logistic growth, written as a differential equation (Marion and Lawson 2008):

$$\frac{dy}{dt} = ry(a - y) \qquad (3)$$

Where, in our case, $y = N, r = \frac{k_1}{L}$, and $a = L$. Here, L represents the capacity in billion cells per liter and k_1 denotes the cancer growth rate in days^{-1}. Once this limit L is reached, the cancer stops growing:

$$\lim_{N \to L} \left(1 - \frac{N}{L}\right) = 0 \quad \Rightarrow \quad \lim_{N \to L} k_1 N \left(1 - \frac{N}{L}\right) = 0 \qquad (4)$$

Since we assume that cancer cells are killed by cytotoxic T cells (assumption 2.b), the death of cancer cells is proportional to the number of cancer cells N and the number of cytotoxic T cells T_c (in billion cells per liter). d_1 describes how effective the T_c cells are at killing cancer; this parameter is increased by interferon treatment (assumption 2.c), making the cytotoxic T cells more effective (assumption 3.i).

Helper T cell equation

$$\frac{dT_h}{dt} = \overbrace{k_2}^{\text{Baseline production}} + \underbrace{k_3 \frac{N}{k_4 + N}\, T_h}_{\substack{\text{Stimulation} \\ \text{from cancer}}} \overbrace{- d_2 T_h}^{\text{Natural death}} \underbrace{- d_3 T_{reg} T_h}_{\text{Death from } T_{reg}} \qquad (5)$$

The equation for the change in T_h cells over time, $\frac{dT_h}{dt}$, includes a baseline production k_2 (in billion cells per liter per time) and a term for the stimulation of T_h cells by cancer cells, as well as two terms describing the death of T_h cells.

Parameter k_2 represents the number of (billion) cells per liter that we assume are produced per time when there is no cancer present in the body, corresponding to assumption 3.a. With interferon

treatment, this parameter increases (assumption 3.j), because that results in an increased immune response.

Based on assumption 3.b, the term for the T_h cell stimulation term includes the number of cancer cells N. Furthermore, it is proportional to the number of T helper cells T_h (in billion cells per liter per time) and the stimulation rate k_3 (in time^{-1}). By including parameter k_4 the effect of cancer cells is limited:

$$\frac{N}{k_4+N} = 1 \; for \; k_4 \ll N \; and \; \frac{N}{k_4+N} = 0 \; for \; k_4 \gg N \qquad (6)$$

This means that the T_h cell stimulation is proportional to the number of T helper cells for a high number of cancer cells compared to k_4. For a very low number of cancer cells the stimulation approaches 0. Due to its limiting effect, k_4 is also referred to as threshold value (in billion cells per liter).

T helper cells die naturally (assumption 3.c) or get killed by regulatory T cells (assumption 3.f). Their natural death rate is represented by d_2 (in days^{-1}), while the death of T_h cells induced by regulatory T cells is proportional to the number of T_h and T_{reg} cells (both in billion cells per liter) and the killing rate d_3 ((billion cells)$^{-1}$ per liters^{-1} per days). Parameter d_3 is equal to 0 once treatment is introduced, since we assume interferon treatment affects the regulatory T cells in such a way that they stop killing T helper cells (assumption 3.f).

Cytotoxic T cell equation

$$\frac{dT_c}{dt} = \overbrace{k_5}^{\text{Baseline production}} + \underbrace{k_6 T_h T_c}_{\text{Stimulation from } T_H} \overbrace{- d_4 T_c}^{\text{Natural death}} \underbrace{- d_5 N \, T_c}_{\text{Killed by cancer}} \overbrace{- d_6 T_{reg} T_c}^{\text{Death from } T_{reg}} \qquad (7)$$

The equation describing rate of change in cytotoxic T cell count, $\frac{dT_c}{dt}$, consists of a baseline production, a stimulation term, and three cell death terms, each representing a different way of how T_c cells die. Parameter k_5 is the baseline production of cytotoxic T cells in billion cells per liter per time (assumption 3.a). We assume, it is increased by Interferon treatment (assumption 3.j).

Furthermore, we assume that T helper cells stimulate the production of additional cytotoxic T cells (assumption 3.d), therefore its associated term is proportional to the number of T_h and T_c cells. k_6 describes this influence as it corresponds to the stimulation rate in (billion cells)$^{-1}$ per liters^{-1} per time. The natural death rate (in time^{-1}) of cytotoxic cells is denoted by d_4.

We also assume that cancer cells kill the cytotoxic cells (assumption 3.e); that number is determined by the killing rate d_5 (in (billion cells)$^{-1}$ per liters^{-1} per time) and proportional to the number of T_c and cancer cells.

When Interferon treatment is introduced, the regulatory T cells start killing the cytotoxic T cells (assumption 3.g). This T_{reg} cell induced death is therefore proportional to the number of T_{reg} and T_c cells during treatment, with killing rate d_6 (in (billion cells)$^{-1}$ per liters^{-1} per time). Without treatment this parameter d_6 is equal to 0, because the regulatory T cells do not affect the cytotoxic cells in that case.

Regulatory T cell equation

$$\frac{dT_{reg}}{dt} = \overbrace{k_7 T_c}^{\text{Stimulation from } T_C} - \underbrace{d_7 T_{reg}}_{\text{Natural death}} \qquad (8)$$

The equation above corresponds to the change in T_{reg} cell count over time, $\frac{dT_{reg}}{dt}$, one term each for the stimulation and natural death of the regulatory T cells is included.
The number of regulatory T cells is directly determined by the number of T_c cells (assumption 3.h), so the parameter for production of T_{reg} cells, k_7 (in time^{-1}), is multiplied by the number of T_c cells. We also suppose that they die naturally (assumption 3.h), so the death rate d_7 (in time^{-1}) is multiplied by the number of regulatory T cells.

Chapter 5: Analysis

When building a model, it is important to consider the model's abilities and whether it is useful or not. We do this by analyzing its graphs and using statistics to evaluate how well it fits our data and its ability to predict development of cell count in the blood. Afterwards we do a sensitivity analysis, which determines the most sensitive parameters; this allows a later discussion of what parameter an ideal treatment could affect.

5.1 Model outcome as graphs

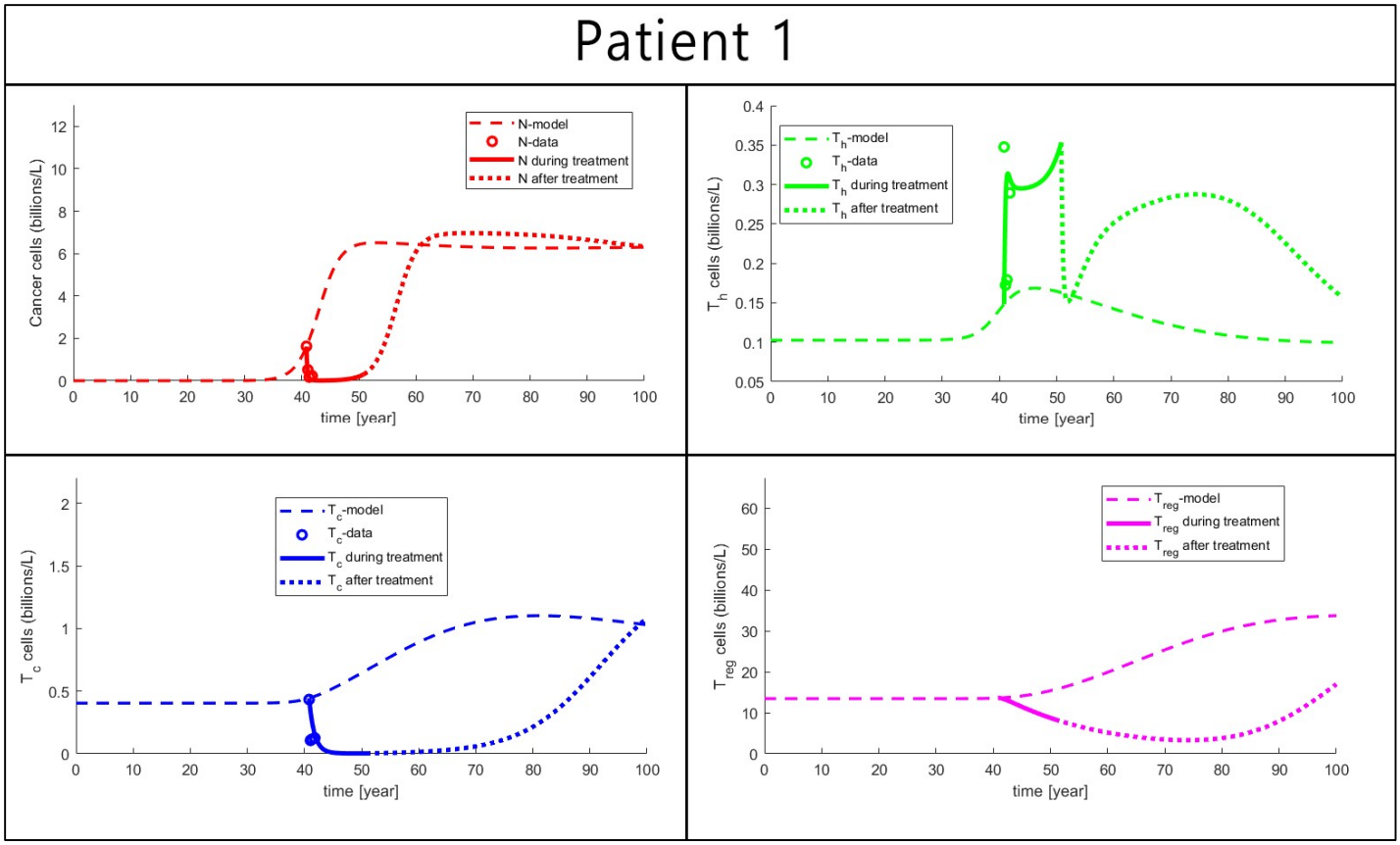


Figure 11. Model prediction and data of the first patient's cell count over 100 years from the first cancer cell appearance. This patient's data was used in parameter estimation.

Figure 11 shows a simulation based on our model, created using MATLAB. It shows graphs for the number of Cancer cells *N* (red), T_h cells (green), T_c cells (blue), and T_{reg} cells (magenta), each in billion cells per liter, over a time period of 100 years from when the first cancer cell appeared, as well as the data points of the patient that we used for parameter estimation. The dashed line represents the model's prediction of cell growth without treatment, whereas the line that is drawn through shows the number of cells during the treatment phase, and the dotted line corresponds to cell growth after treatment has been stopped. The data points are depicted as circles in the respective color. The dashed line in Figure 11 for the cancer cells estimates how many cancer cells per liter a person would have after x years from the first cancer cell appearance. According to our model, it would take around 40 years for 1 cancer cell to replicate into $1.5 * 10^9$ cancer cells.

$N(x)$ is the solution of our differential equation for the cancer cells $\left(\frac{dN}{dt}\right)$. We calculate the time, x_a, of the first observation by solving the equation $N(x_a) = N_{data}$, where N_{data1} is the observed

cancer cell count per liter at the *first* observation at treatment start: This means that for Patient 1, our model estimates that their first cancer cell appears 40.8 years before treatment starts and the data at treatment start $N_{data} = 1.63 * 10^9$ is measured, see Figure 12.

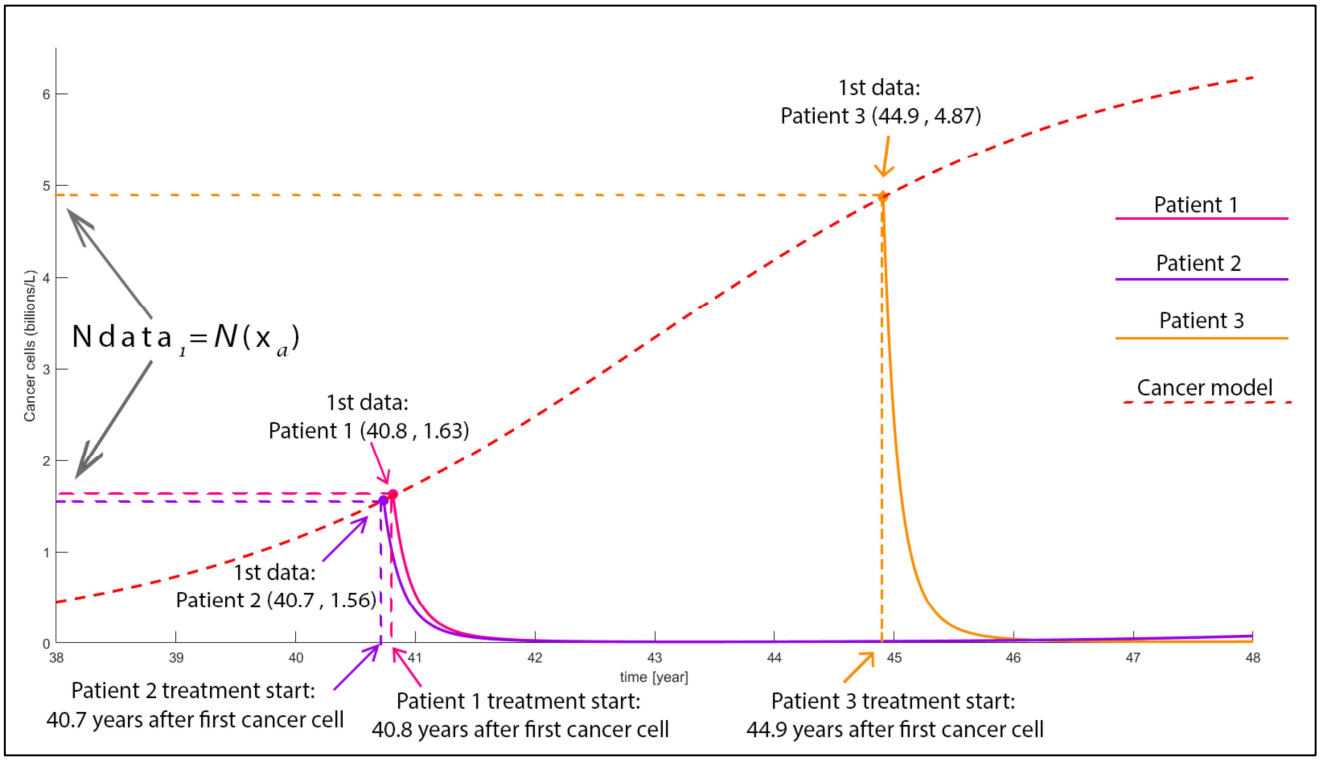


Figure 12. Showing year of treatment start depending on our model's estimation

The cancer cell graph in Figure 11 starts seemingly at zero, but there is one single cancer cell at the start. From there on the cancer grows exponentially, just like the data from the untreated patient suggests. The cancer cell count seems to stay constant in Figure 11 due to its scale, but it is growing logistically. Because of the low number of cancer cells during the first 35 years the number of T_h cells stays close to constant, since their production is barely stimulated compared to the baseline production of T_h cells. Due to this, the amounts of cytotoxic and regulatory T cells are also seemingly constant during this period.

Between year 35 and 40 the number of cancer cells increases rapidly and, based on our model, it would increase further without treatment. This initiates the body's immune response and the production of T_h cells gets stimulated, having an increasing effect on the T_c cells, which then start killing the cancer cells. Without treatment the amount of cancer cells would approach 6.5 billion cells per liter at about year 50 and would stay relatively constant.

After the diagnosis around year 40, the patient is treated with IFN-α treatment, causing several changes (see Figure 13 for graphs zoomed in around treatment phase). Based on our assumptions about Interferon treatment (assumptions 2.c, 3.g, 3.f, and 3.i), regulatory T cells shift from killing T_h cells to killing T_c cells instead (assumptions 3.f and 3.g). Additionally, Interferon elevates the baseline production of T_h and T_c cells (assumption 3.j) and enhances the cytotoxic T cells' effectiveness by raising the rate of T_c cells killing cancer cells (assumptions 2.c and 3.i). Due to these changes the of T_h cells increases rapidly at the beginning of the treatment phase, while the numbers of cytotoxic and regulatory T cells decrease during this phase. The amount of cancer cells in the body still drops considerably because of the T_c cells' enhanced effectiveness. After this significant decrease in cancer cells, they stay constantly low for the rest of the treatment phase, which causes the T_h cell count to

decrease a little during this period. However, a slight cancer growth can be perceived before the end of the treatment phase.

The treatment is stopped after ten years. At this point the number of helper T cells drops rapidly as the effects of IFN-α treatment are missing. There are still only a few cytotoxic T cells in the blood, since the treatment lowered their amount. Due to this the cancer can grow very fast again; this time even to a slightly higher number of cancer cells than our model predicts for disease progression without treatment. Here, the body reacts to the cancer again as the immune response is activated again. The rising number of T_h cells causes the number of cytotoxic and regulatory T cells to increase as well, resulting in a barely significant killing of the cancer cells.

Our simulations show that, without treatment, the number of cancer cells would keep increasing until just before year 50 and stay rather constant after that despite the immune system. Using treatment, the cancer cells are rapidly reduced, but not eliminated entirely, resulting in a relapse just after the end of the treatment phase, with the cancer being more aggressive the second time.

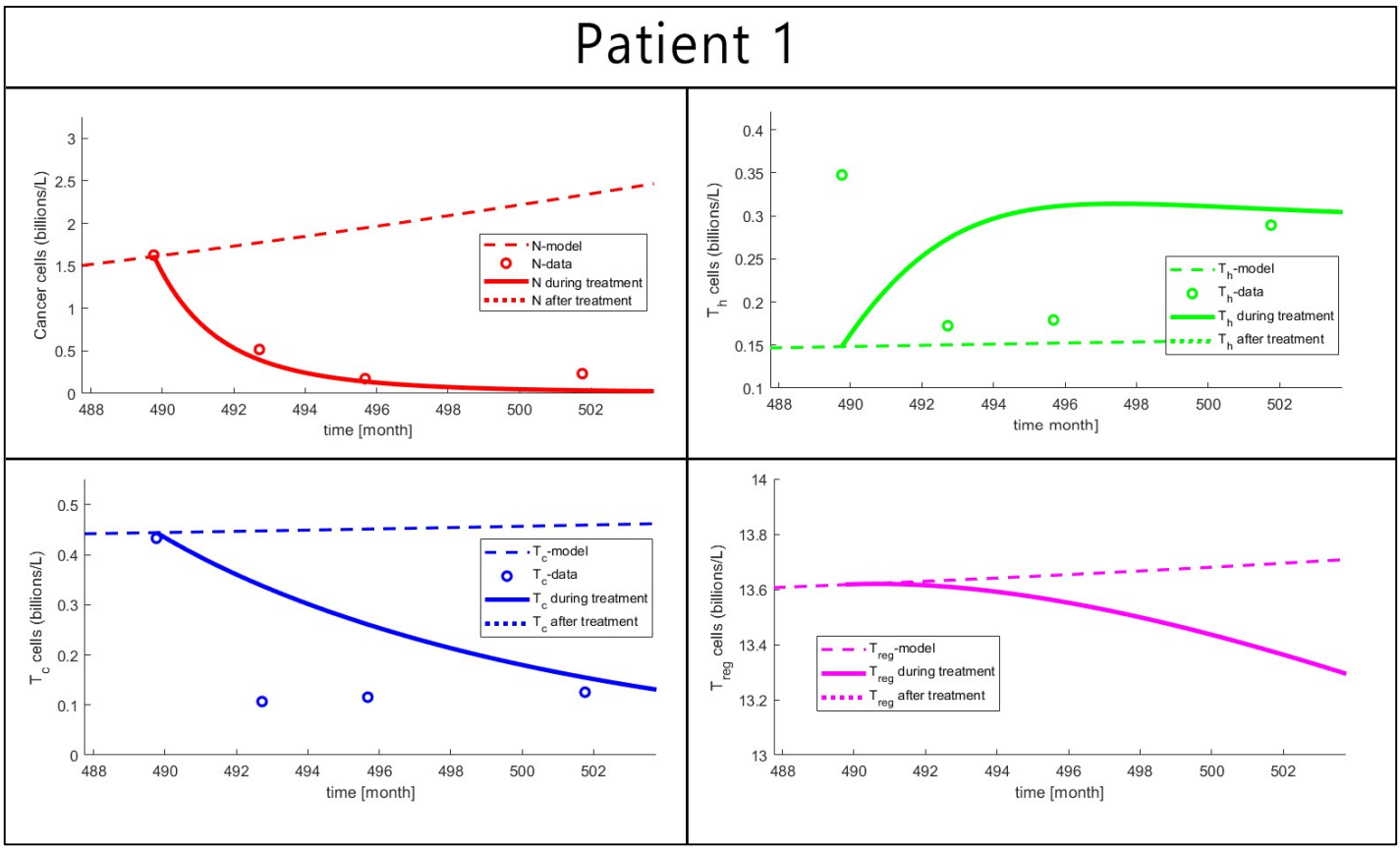


Figure 13. Model prediction and data of the first patient's cell count, zoomed in around data.

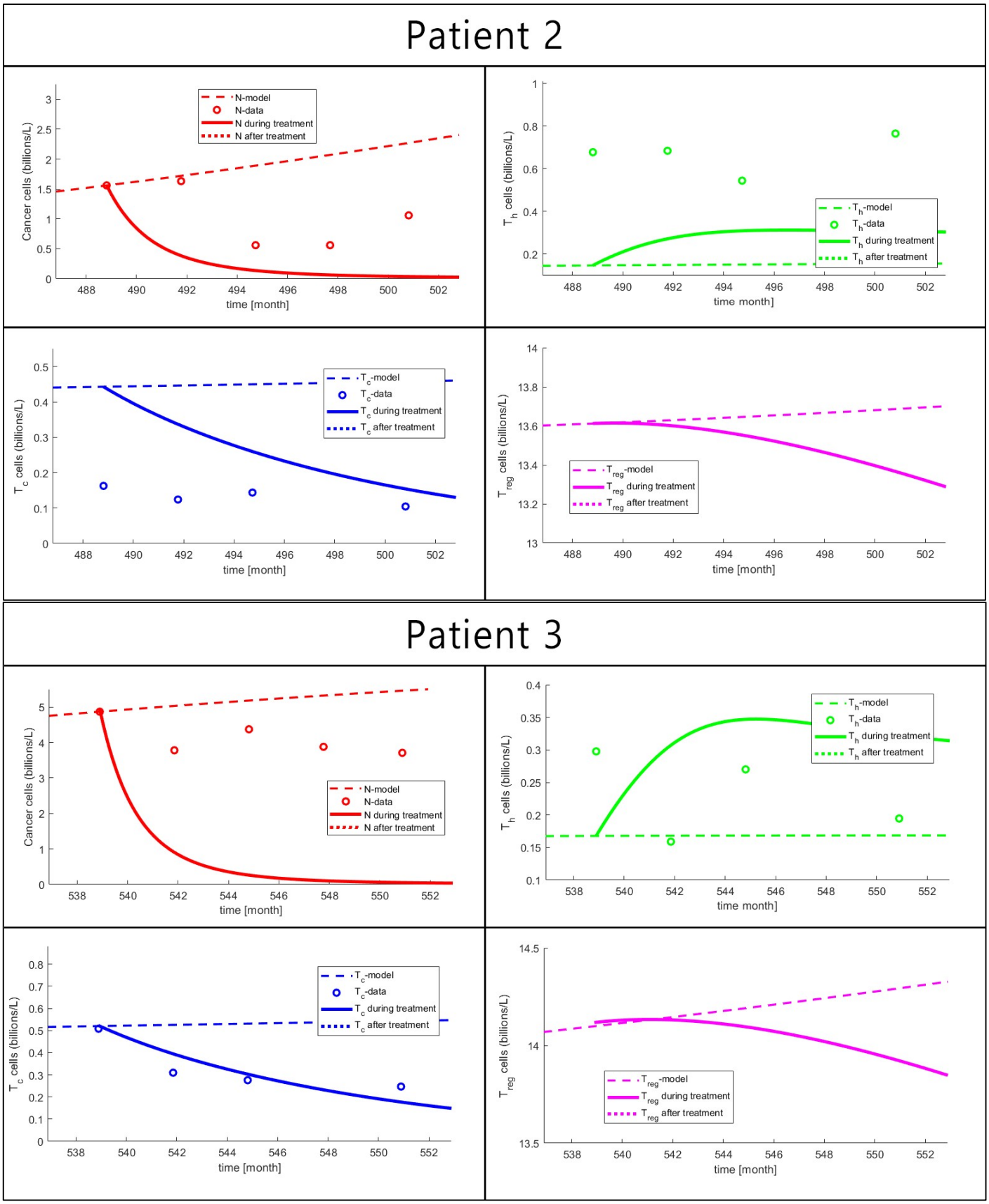


Figure 14. Model prediction and data of the second and third patient's cell count, zoomed in around data.

Description for the other patients

When looking at the graphs for the other two patients shown in Figure 14 several observations can be made. The number of cytotoxic cells seems to be similar for all three patients and is predicted well by the model. Patient 2 has similar cancer cell data to the first patient, but their T_h cell count differs significantly. Since the number of T helper cells is so high in the second patient, the data points do not fit the model.

For the third patient the number of T_h cells resembles the first patient's data. In this case, however, the cancer cell count is higher, meaning that the disease was further progressed at the time of diagnosis. Patient 3 was diagnosed at a later point than the first two patients, but when zoomed in as seen in Figure 14, it is clear that the cancer does not drop as rapidly in reality as the model predicts.

This suggests that our model with the specific parameter set (see Table 7 for parameter values) is only valid for the first patient; this will be discussed in chapter 6.

5.2 Deviations between model and data

Note that to understand the reasoning behind the statistics of this section, one must have basic knowledge of statistics.

To determine whether our model with the decided set of parameters is good, we analyze how well it fits the patient's data with simple statistics. We need something to compare to if we want to say anything about its functionality. We do this by comparing to a very simple model, which we refer to as mean model.

The mean model takes the mean of data and uses this to estimate the next value. It is individual since it predicts depending only on the values of the specific observations and the predictions only work for that specific individual case. It does not consider what mechanics may affect the next outcome and the predicted value will be constant.

We want to investigate how well our model can predict the data. To measure how good the prediction is, we can consider the 95% prediction interval defined as the interval within which 95% of the observations should be. If this interval is narrow, then we are good at predicting; if it is wide, then we are bad at predicting. We will investigate the prediction ability for each cell type and for the cells overall.

If we can assume that the data are normally distributed around the predicted value based on our model with variance σ^2. Then the residual standard deviation (σ) determines how wide the prediction interval is. Therefore, we want to have estimates of the σ for each cell type and for the cells overall. This variance σ^2 covers both measurement errors and the fluctuations, within a patient as well as across patients, in the true number of cells for a random patient.

Sum of Squares of Residuals (SSR)

Sum of Squares of Residuals is a measure of deviations from a model to the data. This is done by calculating the sum of the residuals squared:

$$SSR = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 \qquad (9)$$

where y_i is the i^{th} value of the data corresponding to the value x_i. $f(x)$ is the function of the model, where $f(x_i)$ is the predicted value corresponding to y_i and *n* being the number of observations.

The SSR can therefore be used as an indicator of how far the model is from the observations. However, this method only takes the sum, meaning it does not consider the number of observations. When comparing the SSR of different data, the data must contain the same amount of observations. To take the number of observations into account, we can use the mean square of deviations described below.

The model estimating the population growth of the cancer cells suggests how long it takes for one cancer cell to replicate into a certain number of cells. The first data points from each patient are fitted to this model, suggesting that the first cancer cell appears around 40 years before the first measurement, which differs slightly for each patient. This means the cancer cell number from the first data points for each patient is not included in the SSR, as the time of the first data point has been fitted to the cancer cell value corresponding to the graph or our model, as described in chapter 5.1 Figure 12.

Mean square of Residuals (MSR)
The mean square of residuals is the SSR divided by the number of observations:

$$MSR = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 \quad (10)$$

As described above, this allows for a comparison of errors between data with different numbers of observations. This is necessary, as the calculated SSR is done with data not containing the same number of data points.

Root-mean-square residuals (RMSR)
The root-mean-square deviation is the square root of the mean square of residuals.

$$RMSR = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2} \quad (11)$$

RMSR is useful because it is in the same scale as our observation, and it allows for a comparison with the usual standard deviation.

Let Y_i be the random variable that generates the observed y_i. If for example, we look at the number of cancer cells at the start of treatment for Patient 1, then x_i is time of start of treatment for Patient 1, Y_i is the true but unknown number of cancer cells at time x_i for Patient 1, and y_i is the number of cancer cells observed for Patient 1 at time x_i.

If Y_i is normally distributed with the expected value of Y_i being equal to $f(x_i)$ and the variance of Y_i being equal to σ^2, then RMSR is an estimate of σ and 95% of the y_i should be within $\pm 1.96 \cdot \sigma$ from $f(x_i)$, so the width of the prediction interval is proportional to σ. (Bro 1996)

With Patient 1 however, the RMSR does not estimate prediction, since our set of parameters were chosen while trying to fit Patient 1 data, so the RMSR here *only* states how well the model fits the data.

Sample Standard Deviation (SD)

In the simple mean model, we assume that a constant mean for the cells of a specific type from an individual patient can predict future measurements. This can be given by the standard deviation:

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (12)$$

This can be compared to the RMSR to evaluate how well we can expect our model to fit relative to what a simple mean model would. (Bro 1996)

5.3 Analysis results

	Cancer cells		T helper cells		T cytotoxic cells		Total RMSR
Patient	RMSR	$RMSR/SD$	RMSR	$RMSR/SD$	RMSR	$RMSR/SD$	
1	0.1352	0.2000	0.1300	*1.5152*	0.1374	0.8662	0.1341
2	0.8706	*1.7087*	0.4228	*4.6324*	0.1868	*7.4491*	0.5691
3	3.6320	*12.1154*	0.1235	*1.9135*	0.0572	0.4834	2.0984

Table 6. The RMSR and RMSR/SD for each patient for each cell and all-cell (total) RMSR for each patient. Green underlined numbers show more precise predictions, red/italic numbers represent the opposite.

Table 6 shows the **RMSR** of each cell type for each patient. As mentioned above, multiplying this number with ±1.96 gives us a prediction interval, where 95% of data should lie, under the assumption that they are normally distributed.

The table also shows the **RMSR/SD** which is an estimate of how well our model could predict that certain cell type for that patient compared to the simple mean model.

The green underlined numbers show when our model would give more precise predictions/fits the data better than the mean model, whereas the red and italic numbers represents the opposite.

We see the model fitting equally well in terms of number of cells for Patient 1. However, by looking at the first data points of the first patient, we see that the amount of cancer cells is 1.6272/0.3475≈*16.3* times bigger than the number of T helper cells. This means that a prediction interval of around ±1.96·0.13 ≈ 0.26 billion cells per liter are remarkable for the T helper cells but have a much smaller impact on the cancer cells.

While the prediction interval is almost the same for every cell for Patient 1, it shows to be worse in predicting compared to a simple mean model for the *T helper cells*. Our model estimates a 52% larger σ than the simple mean model, meaning predicting the next number should give a better result when simply calculating with the mean.

For Patient 2, we see that our model in every scenario shows remarkably worse predictions than a simple mean model.

In terms of absolute number of cells per liter, our model overall shows best fit for Patient 3's cytotoxic T cells. The RMSR is half as big or smaller than any other RMSR. Looking at the total RMSR for each patient, we see that Patient 3 has an overall prediction interval 2.0984/0.1341 = 15.6 times wider than for Patient 1.

From the red/italic numbers in the T helper cell – RMSR/SD column, we see that for every patient, the simple mean model would have been better. This suggests that a mean model may be more appropriate in any case.

If our model were to be used in the medical field, the most important type of cell to predict would be the cancer cells. The model for patient 1 seems fine, however for patient 2 and 3 we see that it would be notably worse than the simple mean model. In fact, the estimate of σ is 12 times bigger than the mean model standard deviation. The RMSR for patient 3 shows a prediction interval of ±1.96·3.632 = ±7.229 billion cancer cells per liter. With patient 3 having between 3.6-4.9 billion cancer cells per liter during the tests the prediction interval is so big, our model could predict large negative numbers.

Negative cell counts are not possible, and to further analyze this number, one could take the logarithm to the cell numbers when measuring the fit of the model. This means that prediction interval would be on a logarithmic scale requiring transforming it back before it can be applied to absolute cell numbers.

5.4 Sensitivity analysis

Sensitivity analysis is the study of how the uncertainty in the output of a model can be divided to different sources of uncertainty in the model input (Saltelli et al. 2004). We used this method as one of the final steps to analyze our mathematical model to see what parameters are most responsible for the output change and are therefore most sensitive and also to increase our overall understanding between the input and output variables. We could look for redundant parts of the structure and simplify our model by seeing what parts do not have any effect on the output of the model. This statistical method has many different approaches, for our case we used the most common one – the OAT (one at a time). We took each variable one at a time, changed it and then analyzed the output. We could use this method, since our model does not consist of many equations, it was possible to change the variables and run the model how many times we wanted.

Original parameters

par.k1	par.L	par.d1	par.k2	par.k3	par.k4	par.d3	par.d2	par.k5	par.k6	par.d5	par.d4	par.d6	par.k7	par.d7
0.0015	7.0000	1.44E-04	0.0014	0.0063	0.7000	0.0096	3.00E-04	2.00E-06	0.0019	1.00E-06	2.00E-04	0	0.0050	1.50E-04

Table 7. Original parameter values.

Increasing each of the parameters one at a time by 15 % and comparing the output value for each single one of them with the original output value we could measure the sensitivity of each input parameter. We got the sensitivity values by dividing the absolute value of percentage change in output by the percentage change in input.

$$S = \frac{|output\ cha\quad|}{input\ change} \qquad (13)$$

Change [%]	TC	TH	Treg	N	SUM_TC	SUM_TH	SUM_TREG	SUM_CANCER
par.k1	21.36	-6.75	13.85	2.12	1.92	0.08	2.35	5.52
par.L	0.64	0.73	0.19	14.52	0.52	0.64	0.38	14.11
par.d1	-1.01	0.18	-0.52	-1.27	-0.10	-0.03	-0.10	-1.53
par.k2	32.04	-6.63	42.08	-3.61	32.77	2.17	38.23	0.85
par.k3	12.44	8.81	3.70	-0.66	15.03	4.43	11.34	2.10
par.k4	-1.99	-0.55	-0.82	0.12	-1.62	-0.73	-1.32	0.02
par.d3	-30.63	0.00	-31.10	2.61	-25.82	0.43	-27.70	0.53
par.d2	-11.76	-0.73	-11.78	1.02	-12.69	-0.32	-12.49	1.22
par.k5	0.60	-0.55	0.99	-0.07	0.32	-0.35	0.55	-0.12
par.k6	32.04	-18.78	42.08	-3.61	32.77	-11.16	38.23	0.85
par.d5	-0.25	0.00	-0.06	0.01	-0.75	-0.06	-0.58	-0.26
par.d4	-25.17	24.32	-35.56	2.56	-23.99	11.63	-29.23	-2.29
par.k7	-11.76	-0.73	1.45	1.02	-12.69	-0.32	0.64	1.22
par.d7	12.84	-0.06	-0.03	-1.17	9.71	-1.21	-2.48	-3.35
MAX	32.04	24.32	42.08	14.52	32.77	11.63	38.23	14.11

Table 8. Output changes in percentages for every parameter.

Since our model does not have only one output variable, we needed to look at every output and find the percentage change from the base value. We chose to calculate the change at 50 years from the cancer beginning, as that is the point where the graph changes most drastically and then we calculated the change from the sum of all the cell counts during 100 years. When looking at Table 8 we can conclude what the sensitive values are; these are then put into Table 9.

Cells	Sensitive parameter	Sensitive value
T_C	k_2, k_6	2.14
T_H	d_4	1.62
T_{reg}	k_2, k_6	2.81
N	L	0.97
SUM_T_C	k_2, k_6	2.19
SUM_T_H	d_4	0.78
SUM_T_{REG}	k_2, k_6	2.55
SUM_CANCER	L	0.94

Table 9. Sensitive parameters and values for every type of cell.

By looking at Table 9 we can also conclude that the parameter k_2 and k_6 are affecting T_c cells and T_{reg} cells in the same way, which can easily be explained when we look at those parameters in Figure 9 in chapter 4.2.

Chapter 6: Discussion

In this section we focus on the problems with the data and assumptions, as well as the general validity of our model. We address the fitting analysis and also illustrate how our model could be improved.

6.1 Problems with the data

To assess the accuracy and applicability of our model, it is important to evaluate potential deficits of any data used in the process of building and testing the model. In our case, this refers mainly to the number of patients, amount of data points for each patient, and general shortcomings of data collection.

Working with data from a small number of patients leads to very specific results that might not be applicable for every individual patient, especially since the data from different patients varies considerably, as described in chapter 5.1 and visualized in Figure 6 and Figure 7. Due to these discrepancies in the data from different patients, some data points are not represented well by the model. This could be improved by including data from more patients, in order to be able to compute average values and recognize outliers more easily.

The amount of available data points for each patient is also rather small, limiting our ability to optimize the model's accuracy. In addition to some missing data points for time t = 9 months after the beginning of the treatment phase, the data was collected over the period of one year in intervals of three months, providing us with only 4 to 5 data points per patient, as shown in Table 1, Table 2, and Table 3. A shortening of the data collection intervals and/or a longer collection period would have resulted in more data points, giving us a more precise picture of the disease and its progression. However, we are aware that more data collection is time consuming and expensive, and might be inconvenient for patients, if they have to be tested more frequently. The consequences of this is further discussed later.

Whenever measurements are performed, the general measurement uncertainty must be considered, as well. Since the data we used was collected by medical professionals at Rigshospitalet instead of scientists, it should be noted that the measurements were not taken for scientific research or mathematical modelling, but with the purpose of medical application, and could therefore have been collected under different standards. This could also provide an answer for the small number of data points for each patient, as well as the fact that the number of T_{reg} cells was not measured at all. Additionally, the number of cancer cells, N, was not measured directly, but instead the amount of DNA carrying a mutation of the JAK2 gene was measured. This results in the risk of miscalculation when determining the number of cancer cells, since one cell could carry several mutations.

6.2 Problems about the assumptions and structure

In order to assess the validity of our model the underlying assumptions and delimitations must be examined with regard to their validity, since they are the basis for the model.

As described in chapter 4.1, delimitations have to be made when modeling a complex biological process to simplify the model. Our model only considers the interaction between the cancer cells and the T cells (assumption 1.a). However, the specific immune response also contains two other types of lymphocytes: natural killer (NK) cells and B cells (Mader, Windelspecht, and Preston 2011). These could affect the cancer cells, as well, especially since NK cells can destroy some types of tumor cells

(Lydyard, Whelan, and Fanger 2011), and by purposefully excluding these factors we risk oversimplifying the whole process of the immune response.

The most crucial uncertainty about our model is the fact that it is still unknown how the IFN-α treatment works exactly. We assume that the treatment enhances the downregulation of cancer cells by T_c cells (assumptions 2.c) by increasing the T_c cells' effectiveness in killing cancer (assumption 3.i), which corresponds to increasing parameter d_1. Additionally, the treatment prevents the T_{reg} cells from killing T_h cells and makes them kill T_c cells instead (assumptions 3.f and 3.g), so parameter d_3 is set to 0, while d_6 is not 0 anymore. IFN-α treatment also increases the baseline production of T_h and T_c cells (assumption 3.j), corresponding to an increase in parameters k_2 and k_5, respectively. However, it is only known that IFN-α treatment enhances the immune response by increasing a specific molecule responsible for signaling T cell activation in general (Williams 2012). IFN-α could possibly affect the production of regulatory T cells in a way that is dependent on the number of cancer cells. This case is not considered in our model; it is therefore possible that our assumptions about IFN-α treatment are invalid.

Due to a limited amount of time available we might not have gotten the best possible parameter set. Additionally, the parameters were mostly determined by guessing through a process of trial and error, so it is possible that our set of parameters is not the only option and another set might fit the equations just as well.

6.3 Discussion of fitting analysis

Sufficiency of the model

As described in the Data fitting analysis chapter 5.2 we decided to compare our differential equation model to the one of the simplest possible models: the simple mean model.

We see a model that fits the number of cancer cells during treatment for Patient 1. There is a clear difference in the prediction ability when using our model instead of a simple mean model. This, however, says nothing about the sufficiency of the model being used on other patients, since the parameters only were changed to fit the data of Patient 1, and with more weight on making the cancer cell model fit.

This means that even though we got a fairly well fitted model here for cancer cell data of Patient 1 (calculated in section 5.2), it says more about our ability to fit parameters, than it says about the general validity for using the model.

Since our aim is to find a model that can be used in the medical field, especially for the change of cancer cells over time, this is what we consider to be the most important model to fit. For Patient 2 and 3, the model shows a very large, and therefore bad, prediction interval. Based on these numbers, our model would be worse than using a simple mean model.

Whether the prediction interval (RMSR) in absolute cell number has a biological impact must be assessed by someone in the cell biology field.

Ability of our model and the mean model

Even though the simple mean model seems to allow better predictions than our model, a simple mean model would require more data. One data point is not enough, since the model is built from the mean of historical data, and this means you need to measure every type of cell several times to be able to make predictions. To achieve better results with the mean model, one would have to continuously

take samples and update the model, since the model simply gives a constant, which only allows for short term predictions.

Our model is quite complicated and as a dynamical model, so it should do much better predictions than a model built on a mean, which is not necessarily impossible for our model, but at least not with our set of parameters.

The set of parameters intend to be universal for every patient with the ET disease being treated with IFN-α and with the chosen parameter set, our model would, if sufficient, only need the cells per liter count of JAK2 mutated cells at the time of treatment start to allow prediction of cancer cell count. This suggests our model having more practical use than a mean model.

A way to better this issue would be to find the best fitting parameters for each patient and create a way to convert an observation into a set of parameters, which would fit for an individual patient based on specific character traits such as age. To do so a program or model that would calculate individual parameters based on knowledge of the patient could be developed. A large Danish epidemiological study concluded that factors like age, gender and smoking behavior, for example, are associated with JAK2-V617 positive MPN. Increasing age increases the chance of the presence of the JAK2-V617 mutation. So does male sex and increasing current smoking behavior. (Nielsen et al. 2013).

Deciding what parameters correlates with characteristics requires a lot of data and a lot of correlation statistics analyses. If doctors could simply insert patient traits in a program like that, they would receive the correct parameters to insert in the model and thereby be able to apply the model on individual patients.

Fitting parameters

Trying to fit parameters by changing one at a time and seeing if it fits well with a graph is not the ideal way to go.

A better way to fit parameters would be to use the least squares method: a kind of regression analysis which minimizes the SSR. This, however, is technically very complicated, since our model of the mean is not a linear model, but a differential equation model. Besides this, we would need more observations and fewer parameters to identify a unique, optimal set of parameters.

What could make our model better

Converting our model to be dimensionless could help detecting less important parameters. Since a big part of the fitting issues stem from a complicated model with many parameters, removing unnecessary parameters could allow better fit (Marion and Lawson 2008). One can also determine what parameters could be cut is a sensitivity analysis. Even though we have done a sensitivity analysis, we did not use it to remove parameters.

Another way of lowering our number of parameters could have been to leave out the regulatory T cells, since the effect of these is not certain and we do not have any data to suggest a tendency of these. With the missing knowledge of the role of regulatory T cells, there is less certainty of how to apply this cell in a mechanism model.

As discussed in the problems with assumptions, the effects of treatment that we assume to be true may have been wrong. Knowing the mechanism behind the treatment's effect, we could have adjusted the model/parameters to match this outcome.

Other approaches

Our model seems quite impractical to predict cell count per liter over time. Something else our model may have been of use for, is to simply show the interaction between these cells. Instead of being used in the medical field as an attempt to predict anything, it could serve the purpose of contributing to understanding the mechanism behind the biological interaction.

As described in chapter 5.1, cancer cell growth in the model starts slow, staying relatively low for many years. This may correspond well to real life situations, as cancer can be asymptomatic in its early stages and is more likely to be diagnosed at later stages. This can be confirmed by the fact that, for patient data that is used, patients are diagnosed after the number of cancer cells starts rapidly increasing (which may correspond to the rapid worsening of their condition), but further research is needed.

After growing rapidly at some point the number of cancerous cells starts staying relatively constant. At that point in real life situations the patient might already have passed away due to the inability of the blood to fulfill its functions.

When treatment starts, the number of cytotoxic cells drops similarly to the cancer cells. This may seem counterintuitive but might be true due to how interferon treatment works (further research is needed). After cancer the patients may exhibit lower immunity, but further research is needed.

Number of T_h cells grows rapidly after starting the treatment, following with a rapid drop and increasing again, albeit slower this time. It is hard to judge if the model outcome is precise in this instance as there is no patient data from this extended time period.

After treatment discontinuation the cancer cells start growing rapidly again, because there are less cytotoxic cells that can deal with cancer cells and their effect on cancer cells is also less significant. This may be true due to the fact that cancers are usually more aggressive upon relapsing, but further research is needed.

Our model however still shows insufficient fit for the data, so also this approach would be insufficient.

6.4 Best treatment

This section will discuss the best possible treatment with the help of sensitivity analysis. It should be noted that this assumes our model to represent reality.

First, we will try to find the ideal parameters that we will consider changing by seeing which parameter has the biggest effect on cancer cells and ideally not drastic effect on the immune cells. After that we will change those parameters and monitor the changes it has on the patient when we start with the treatment. The goal is to find a parameter or a set of parameters that will affect the treatment in such a way that it makes the cancer decline faster or completely kills it without affecting the rest of the system too much.

When looking at parameters that decrease the cancer the most (Table 9), we could conclude that the best possible treatment would change either parameter k_2 or k_6. However, these two also increase the T_c cells and T_{reg} cells count the most which could be a problem because the best possible treatment should not affect the rest of the immune system, since more T_c cells can cause trouble in the body by killing other healthy cells. However, by increasing k_2 and k_6 we also increase T_{reg} cells. These have a downregulating effect on T_c cells, so it is possible that this would prevent the increase of T_c cells.

Looking at Table 9 we can see that the parameter that is most sensitive in affecting the cancer is the L parameter. Since this is the limiting capacity, we cannot change it as it would mean decreasing or

increasing the cell amount in the blood which is impossible. This fact leaves us with the second most sensitive parameter, k_1, which is the cancer growth rate if we decreased this parameter, we would get big decrease in the cancer count and small change in the immune cells count. So ideally, the best possible treatment needs to change the parameter k_1. When monitoring the cancer after the treatment starts and after the treatment ends both before and after the decrease of k_1, as we can see in Figure 15 it makes the cancer decrease.

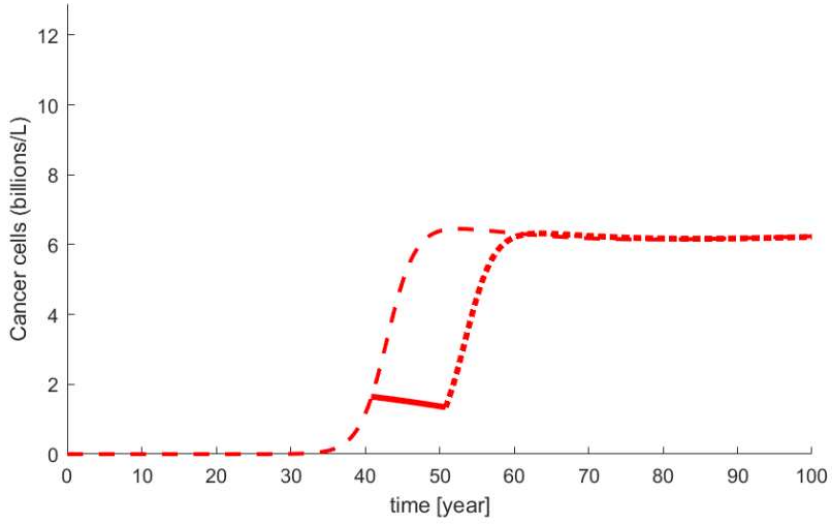


Figure 15. Cancer when we start with the treatment (full line) and after we stop with the treatment (dotted line).

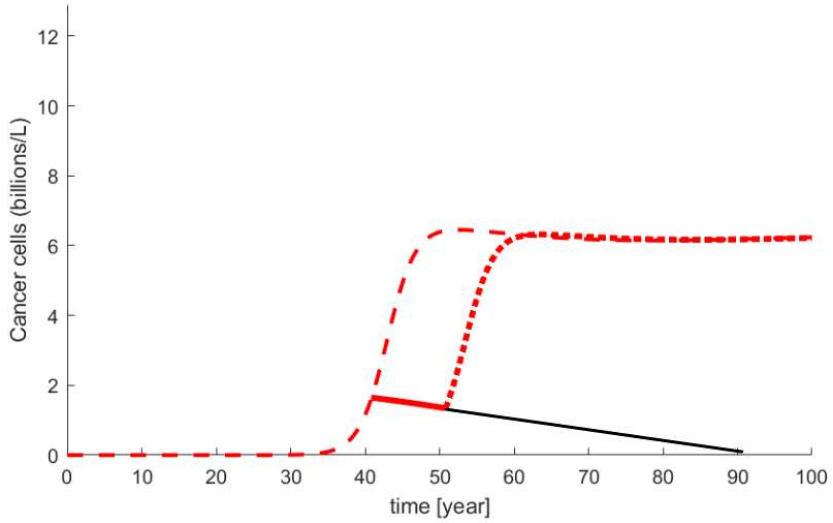


Figure 16. Cancer when we start with the treatment (full line) and after we stop with the treatment (dotted line) and the trend (black line) that the treatment follows.

If we had not stopped with the treatment and the cancer count declined in the same rate for the whole time, we can see that the cancer count would reach 0 around 50 years after the treatment started. This is not the best ideal treatment as we would have to treat our patient for around 50 years. Therefore, we need to look for another parameter, that would eliminate the cancer faster. Looking at Table 8 we will now choose the parameter d_1, which is the rate of T_c cells killing cancer cells. We picked this parameter, since it changes the cancer count by a lot compared to the immune cells.

Decreasing k_1 by factor 0.02 and increasing d_1 by factor 10 we get a result shown below.

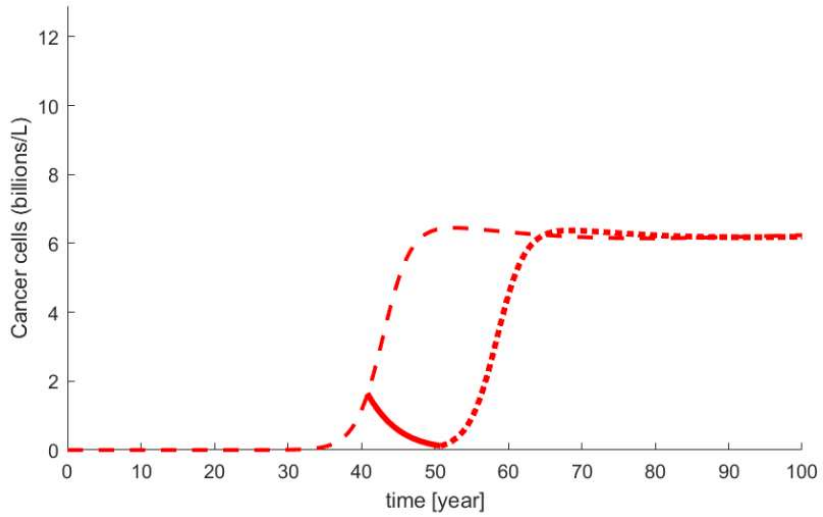


Figure 17. Cancer when we start with the treatment (full line) and after we stop with the treatment (dotted line) when decreasing k_1 and increasing d_1.

This combination of parameters shows promising result as we were able to decrease cancer in a shorter period of time compared to only changing k_1 and still, we were able to keep immune cells in a state that is not dangerous to the body.

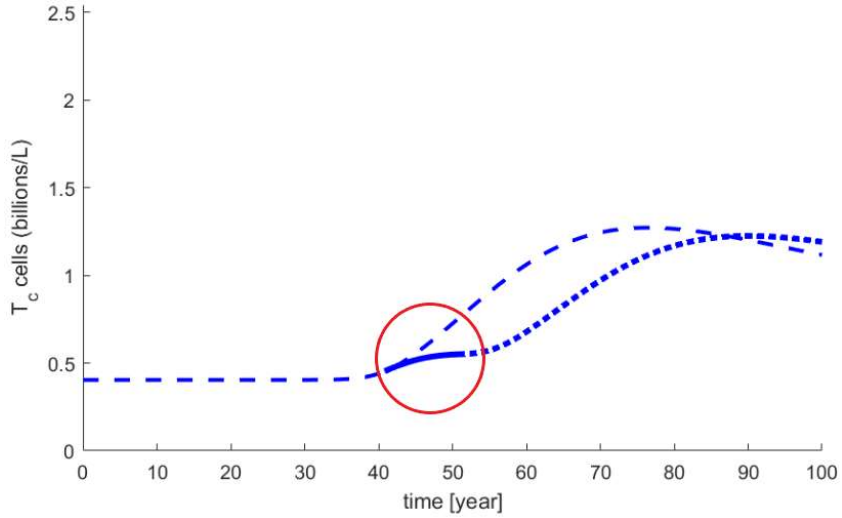


Figure 18. T_c cells count after we started with the treatment (full line) and after we stopped with the treatment (dotted line) when decreasing parameter k_1 and increasing d_1.

In Figure 18 we see that T_c cells count slows down and therefore is closer to the initial state of T_c cells count (dash line).

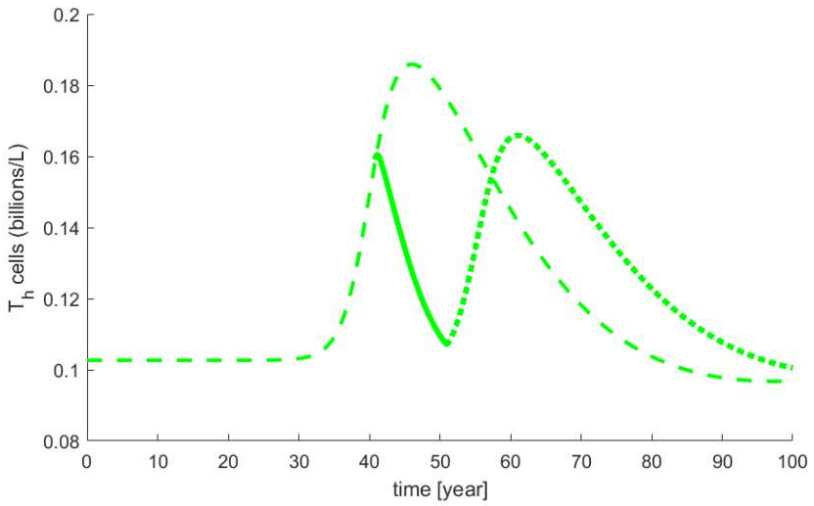


Figure 19. T_h cells count after we started with the treatment (full line) and after we stopped with the treatment (dotted line) when decreasing parameter k1 and increasing d1.

In Figure 19 we see that T_h cells count approaches the initial healthy state when we start with the treatment (full line).

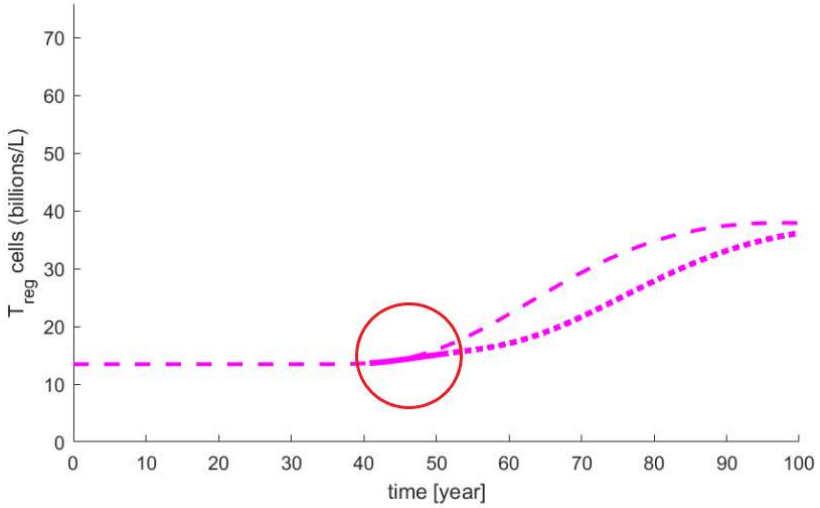


Figure 20. T_{reg} cells count after we started with the treatment (full line) and after we stopped the treatment (dotted line) when decreasing parameter k_1 and increasing d_1.

In Figure 20 we can see that T_{reg} cells count slows down and therefore is closer to the initial healthy state.

When changing parameters k_1 and d_1 we get the promising result where cancer decreases as shown in Figure 17 while all the immune cells stay relatively close to the normal state as shown in Figure 18, Figure 19 and Figure 20. Even though this might be what we were looking for, we still cannot say that this is the best ideal treatment, since there might exist an even better set of parameters that could be changed. An example for that could be parameters that change the cells in same way, but additionally affect the cancer after we stop with the treatment in a way that it grows back slower or not at all.

It makes sense that k_1 and d_1 are the parameters to affect the cancer cell count the most while not affecting the rest of the cell types as much, since they are the parameters affecting the cancer directly.

Naturally, it would be ideal if a treatment only affected the disease and nothing else to avoid a possible chain of unexpected consequences.

Our chosen parameter set also increases d_1 with around a factor 300, suggesting that, with the effect we assume the IFN-α treatment has, the IFN-α treatment shows similarities to the "best treatment". Whether it is possible for a treatment to affect the growth rate of the cancer cells (k_1) requires more knowledge of the immunology field and will not be assessed further in this report.

If we were to use decrease k_1 by factor 0.02, but increasing d_1 by factor 303.75 instead of ten, as done in our model, we would see that the cancer would not relapse, and the other cell types would return close to their initial values as shown in Figure 21. The figure provides an overview of the difference in the "best treatment" and the IFN-treatment, and it is clear that the left side is much more ideal.

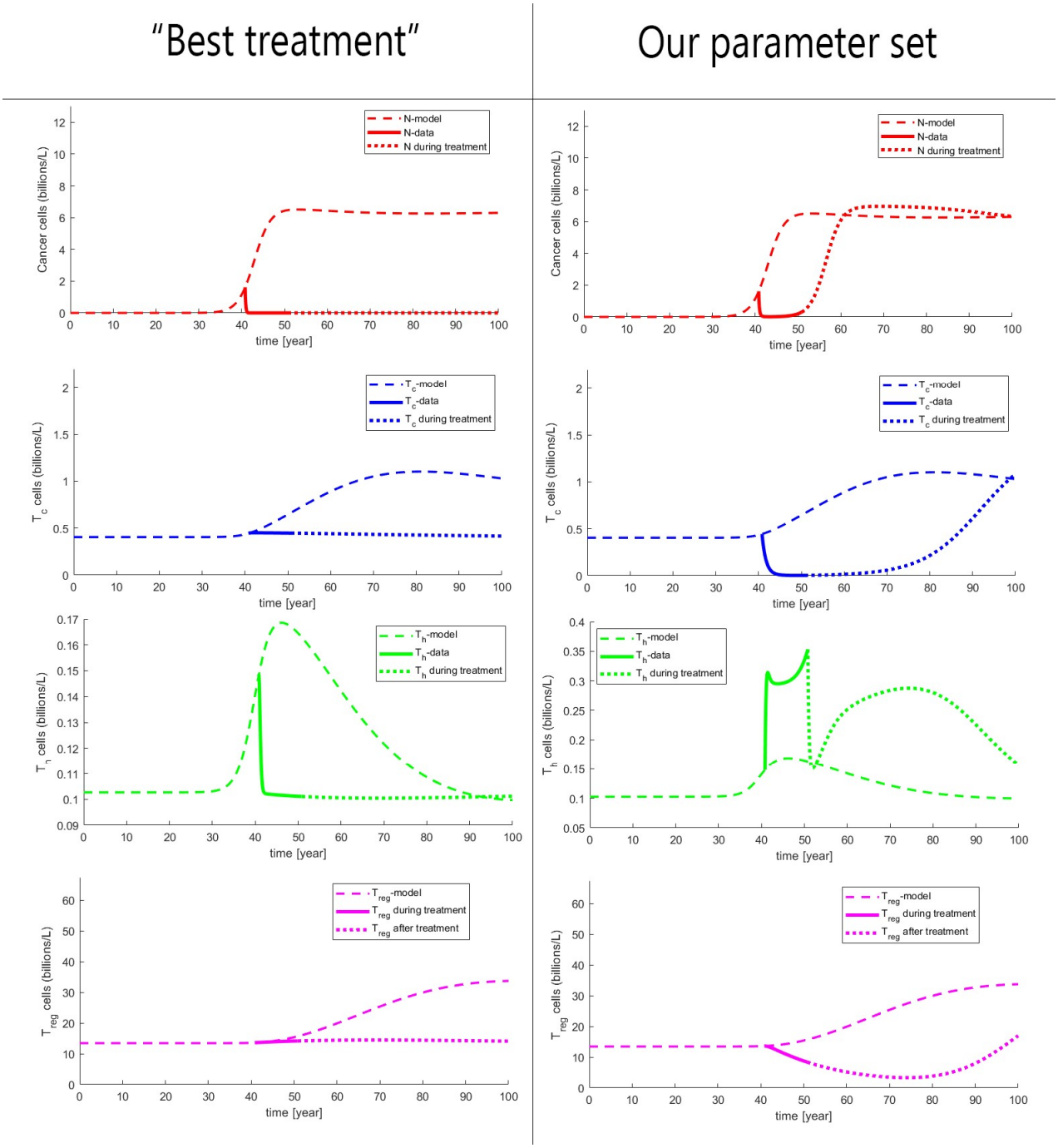


Figure 21. An overview of the different outcomes of the "best treatment" and the IFN treatment (based on our assumptions).

This is of course an imaginary scenario, and especially with the uncertainty in our model, this is purely a thought experiment.

6.5 Use of the model in the medical field

The model presented in this report aims to predict the development of cancer and three essential types of T cells, representing the human immune system, after cancer has started, after treatment has started and after treatment has stopped. With the information given by a model like that, not only can the overall understanding of the pathogenesis of MPN be deepened; it also becomes possible to predict how good the treatment will work depending on at which stage of cancer development it was started and how long it will take for the cancer to return after treatment has been stopped. Knowledge

like this gives crucial information about when, how long and with what dosage treatment should occur.

Besides these practical advantages in the aid to cure MPN, this model tries to describe the development of JAK2-V617F mutated cells in relation to T cells during MPN blood cancer progression. It thereby contributes as an argument for the novel concept that MPN's support and trigger inflammatory diseases and are thereby highly linked to each other. In order to properly treat MPN it must be clear what should be targeted - the mutated cells or the inflammation. A model like the presented one could help making decisions about which treatment would have the biggest effect.

Due to the fact that this model is strongly oversimplified, it is by far not precise enough to being of use in the real world. It is a first attempt though and with more time and knowledge we believe it could be possible to create a model like ours with the ability to contribute in the medical field. As discussed earlier, collecting more data, simplifying the model and finding a way to choose individual parameters based on someone's age, for example, or whether they are smokers. With the prediction confidence we see in our model, it would not be practical nor precise enough to be used in something as serious as the medical field, but it may hold the ability to be improved enough to be of use.

Chapter 7: Conclusion

The interaction between T cells and leukemia cells can be modelled by using the first order ordinary differential equations (2), (5), (7) and (8). Since they are based on assumptions which we derived from current biological knowledge about leukemia and the immune system, the equations represent our beliefs about the interaction of those cells. With that, our model can assist in understanding the disease both with and without treatment by understanding the mechanism behind the disease.

Considering the fact that the exact mechanism and effects of IFN-α treatment are still unclear, the only way to test the accuracy of our assumptions and the change of parameters thereof is to see the change in the outcome of the model. By seeing the change in output and comparing it to the actual patient data, we can evaluate how precise our assumptions about treatment are. Our model also provides visual representation of cancel cell growth, helping in understanding how the disease progresses in both of the scenarios.

This report focuses on a specific set of parameters with the intention of allowing prediction for every single patient with this disease. With these values (shown in Table 5) the suggested model is successful in showing the number of cancer cells over time without treatment, as well as with treatment and after the treatment phase, but only in a very limited sense. The model's graphs also agree with the first patient's data points for cytotoxic T cell count, mainly because we used these data points for parameter estimation.

However, it was difficult to accurately represent the T helper cell data, which suggests that our model still needs improvement to be more accurate. It is also possible that another parameter set fits equally well as ours. Additionally, the specific parameter values that we estimated do not work for the other patients, so the values need to be adjusted to for every individual patient.

Having a universal set of parameters enables the possibility of only needing cancer cell count per liter at the start of the treatment phase to predict the treatment period. As we have discovered in chapter 5.2, this set of parameters is not representative for every patient. This means that patients with the ET disease, that are being treated with IFN-α, need individual parameter sets. If they could be decided from character traits that showed similar parameter sets, the only information required to predict would be the cancer cell count and the characteristics. Finding parameters that fit a character trait would need much more research, testing, statistical analyses and data. The possibility of finding potential clusters of patients was discussed in chapter 6.3. Based on these it might be possible to develop a program which calculates patient specific parameter values from their characteristics.

Since the parameters need to be adjusted for every individual patient, applying the model in medicine does not seem practical, because it would be very time-consuming. However, our model might provide a useful basis for application together with a program that determines the parameters based on patient specific criteria.

The best possible outcome of this project was to create a mathematical model sufficient enough to be practically used in the medical field. Even though we were not able to achieve this result, we could still use our model to find the best possible treatment. The knowledge of what parameters need to be changed and by how much, together with the professional knowledge of the medical field, could be used to find the treatment that affects the parameters in this way and therefore treat patient in the best way possible.

References

Abdel-Wahab, Omar, and Ross L. Levine. 2010. “Genetics of the Myeloproliferative Neoplasms.” In *Myeloproliferative Neoplasms*. Humana Press, Totowa, NJ.

Adams, Robert A. 2018. *Calculus: A Complete Course*. 9th ed. Pearson Canada Inc.

Adams, Robert A., and Christopher Essex. 2018. *Calculus: A Complete Course*. Ninth Edit. Pearson Canada Inc.

Andersen, Morten, Zamra Sajid, Rasmus K. Pedersen, Johanne Gudmand-Hoeyer, Christina Ellervik, Vibe Skov, Lasse Kjær, et al. 2017. “Mathematical Modelling as a Proof of Concept for MPNs as a Human Inflammation Model for Cancer Development.” *PLoS ONE* 12 (8): 1–18. https://doi.org/10.1371/journal.pone.0183620.

Cruse, Julius Major, and Robert Edwin Lewis. 1999. *Atlas of Immunology*. Boca Raton, Fla. ; London : CRC Press.

Gadina, Massimo. 2013. “Janus Kinases: An Ideal Target for the Treatment of Autoimmune Diseases.” *The Journal of Investigative Dermatology. Symposium Proceedings / the Society for Investigative Dermatology, Inc. [and] European Society for Dermatological Research* 16 (1): S70–72. https://doi.org/10.1038/jidsymp.2013.29.

Hasselbalch, Hans Carl, and Mads Emil Bjørn. 2015. “MPNs as Inflammatory Diseases: The Evidence, Consequences, and Perspectives.” *Mediators of Inflammation* 2015. https://doi.org/10.1155/2015/102476.

Hernandez-Vargas, Esteban A.; Sanchez, Edgar N. 2019. “Mathematical Modeling Principles.” In *Modeling and Control of Infectious Diseases in the Host*, 19–33. https://doi.org/10.1016/B978-0-12-813052-0.00012-9.

Kiladjian, J. J., C. Chomienne, and P. Fenaux. 2008. “Interferon-α Therapy in Bcr-Abl-Negative Myeloproliferative Neoplasms.” *Leukemia* 22 (11): 1990–98. https://doi.org/10.1038/leu.2008.280.

Klein, Jan. 1990. *Immunology*. Blackwell Scientific Publications, Inc.

Kuby, Janis. 1992. *Immunology*. New York : W.H. Freeman.

Kundranda, Madappa N., Raoul Tibes, and Ruben A. Mesa. 2011. “Blastic Transformation of BCR-ABL-Negative Myeloproliferative Neoplasms.” In *Myeloproliferative Neoplasms*. Springer, Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-24989-1_12.

Lydyard, Peter, Alex Whelan, and Michael Fanger. 2011. *BIOS Instant Notes in Immunology*. 3rd ed. Routledge.

Mader, Sylvia S, Michael Windelspecht, and Lynn Preston. 2011. *Human Biology*. 12th ed. McGraw-Hill.

Marion, Glenn, and Daniel Lawson. 2008. *An Introduction to Mathematical Modelling*.

McLornan, Donal, Melanie Percy, and Mary Frances McMullin. 2006. “JAK2 V617F: A Single Mutation in the Myeloproliferative Group of Disorders.” *Ulster Medical Journal* 75 (2): 112–19.

Nagle, Erika. 2008. “Sarežģītas Dzīvības Norises - Fizioloģiskie Procesi.” In *Bioloģija Vidusskolai. Mācību Grāmata*. Lielvārds.

Paul, William E. 1993. *Fundamental Immunology*. 3rd ed. New York: Raven Press.

Sabath, D.E. 2013. “Philadelphia Chromosome.” In *Brenner’s Encyclopedia of Genetics*. https://doi.org/10.1016/B978-0-12-374984-0.01155-4.

Saltelli, Andrea, Stefano Tarantola, Francesa Campolongo, and Marco Ratto. 2004. *Sensitivity Analysis in Practice*. John Wiley & Sons, Ltd.

Sharma, Anil K., ed. 2019. *Immunology: An Introductory Textbook*. 1st ed. Jenny Stanford Publishing.

Tizard, Ian R. 1992. *Immunology: An Introduction*. Fort Worth, TX : Saunders College Pub.

Williams, Andrew E. 2012. *Immunology. Mucosal and Body Surface Defenses*. 1st ed. John Wiley & Sons, Ltd.

Appendix

Initial conditions

In order to find the initial conditions for T cells before the cancer growth starts, we assume that the homeostasis of the person is in steady state at the time of the first malignant mutation.

After all the differential equations are set to zero and N=0 there are three equations with three unknown variables, which can be solved as followed.

Setting N to 0, we get these equations:

$$\frac{dT_H}{dt} = k_2 - d_2 T_H - d_3 T_{reg} T_H \quad (1)$$

$$\frac{dT_C}{dt} = k_5 + k_6 T_H T_C - d_6 T_{reg} T_C - d_4 T_C \quad (2)$$

$$\frac{dT_{reg}}{dt} = k_7 T_C - d_7 T_{reg} \quad (3)$$

The right-hand side of the differential equations can be set equals zero since in the case of no cancer cells present in the body, the number of each of the T cells are expected not to change.

$$k_2 - d_2 T_H^* - d_3 T_{reg}^* T_H^* = 0 \quad (4)$$

$$k_5 + k_6 T_H^* T_C^* - d_6 T_{reg}^* T_C^* - d_4 T_C^* = 0 \quad (5)$$

$$k_7 T_C^* - d_7 T_{reg}^* = 0 \quad (6)$$

Equations (4) and (6) can be solved for T_H^* and T_{reg}^* and expressed with respect to T_C^*.

$$T_H^* = \frac{k_2}{d_2 + d_3 (\frac{k_7}{d_7} T_C^*)} \quad (7)$$

$$T_{reg}^* = \frac{k_7}{d_7} T_C^* \quad (8)$$

Equations (7) and (8) can be substituted into equation (5)

$$k_5 + k_6 \frac{k_2}{d_2 + d_3 \frac{k_7}{d_7} T_C^*} T_C^* - \frac{d_6 k_7}{d_7} (T_C^*)^2 - d_4 T_C^* = 0 \quad (9)$$

Moving all negative terms to the right-hand side of the equation, we get two curves on each side.

$$k_5 + k_6 \frac{k_2}{d_2 + d_3 \frac{k_7}{d_7} T_C^*} T_C^* = \frac{d_6 k_7}{d_7} (T_C^*)^2 + d_4 T_C^* \quad (10)$$

The right-hand side describes an increasing convex curve of a function of T_C^* going from 0 to infinity in the positive half plane.

The left-hand side describes a concave curve as a function of T_C^* starting from k5 and approaching $\frac{d_7 k_2 k_6}{d_3 k_7}$.

Those two curves meet only once in the positive half plane. Thus there is only one positive solution for T_C^*.

Equation (10) can be expressed as a third-degree polynomial.

$$-d_3 d_6 (\frac{k_7}{d_7})^2 (T_C^*)^3 - (d_2 d_6 + d_3 d_4) \frac{k_7}{d_7} (T_C^*)^2 + \left(\frac{d_3 k_7 k_5}{d_7} + k_2 k_6 - d_2 d_4\right) T_C^* + d_2 k_5 = 0 \quad (11)$$

The positive solution to this polynomial is the solution for T_C^*.

This third-degree polynomial can be solved with MATLAB's root finder. The solution T_C^* is the steady state of T_C-cells and can be substituted in equations (7) and (8) to get T_H^* and T_{reg}^*.

MATLAB Code

```
clear all;
close all;
```

Choose patient number

--- Choose patient number here----

```
patient=1
```

```
patient = 1
```

Parameters

```
% parameters for dN/dt
    par.k1 = 0.0015;        %(baseline) production of cancer cells
    par.L = 7;              %limit of cancer cells
    par.d1 = 1.4400e-04;    %killing of cancer cells by Tc cells

% parameters for dTh/dt
    par.k2 = 0.0014;        %baseline production of Th cells
    par.k3 = 0.0063;        %increased production of Th cells as immune
response to cancer
    par.k4 = 0.7;           %threshold value
    par.d2 = 0.0096;        %natural dying rate
    par.d3 = 3.0000e-04;    %killing of Th cells by Treg cells (in case of
cancer)

% parameters for dTc/dt
    par.k5 = 2.0000e-06;    %baseline production of Tc cells
    par.k6 = 0.0019;        %stimulation of Tc cell production by Th cells
    par.d4 = 2.0000e-04;    %natural dying rate
    par.d5 = 1.0000e-06;    %killing of Tc cells by cancer cells
    par.d6 = 0;             %killing of Tc cells by Treg cells (in case of
cancer)

% parameters for dTreg/dt
    par.k7 = 0.005;         %production of Treg cells, stimulated by Tc
cells
    par.d7 = 1.5000e-04;    %natural dying rate of Treg cells
```

Initial conditions

```
a = par.d3*par.d6*(par.k7/par.d7)^2;
b = (par.d2*par.d6+par.d3*par.d4)*par.k7/par.d7;
c = par.d2*par.d4-par.d3*par.k7*par.k5/(par.d7)-par.k2*par.k6;
d = -(par.d2*par.k5);

rootfinder=[a b c d];
q=roots(rootfinder);

Tc0   = max(q);
Treg0 = par.k7/par.d7*Tc0;
Th0   = par.k2/(par.d2+par.d3*par.k7/par.d7*Tc0);
N0    = 1e-9            ;

init = [N0, Th0, Tc0, Treg0];
```

Data (1st patient)

```
 if patient==1
     diagnosetime1 = 40.8135 * 365;                                    %time
of diagnose in days (~40 years)
     tdata = [0 90 180 270 365] + diagnosetime1;                       %time
in days
     Thdata = [0.3475472 0.17235855 0.17906184 NaN
0.28923843];         %number od Th cells in 10^9 cells per L
     Tcdata = [0.4326608 0.106743 0.11546064 NaN
0.12544194];           %number of Tc cells in 10^9 cells per L
     Ndata  = [1.6272 0.51744 0.17424 NaN
0.23312];                    %number of cancer cells in 10^9 cells per L
 end

 if patient==2
     diagnosetime2 = 40.7345 * 365;                                    %time
of diagnose in days (~40 years)
     tdata = [0 90 180 270 365] + diagnosetime2;                       %time
in days
     Thdata = [0.677527 0.684147 0.544403 NaN
0.764778];                %number od Th cells in 10^9 cells per L
     Tcdata = [0.162867 0.1243 0.143974 NaN
0.104684];                  %number of Tc cells in 10^9 cells per L
     Ndata = [1.56046 1.63134 0.56088 0.56105
1.05938];                 %number of cancer cells in 10^9 cells per L
 end

 if patient == 3
      diagnosetime3 = 44.9076 * 365;
```

```
     tdata = [0 90 180 270 365] +
diagnosetime3;                          %time in days
     Thdata = [0.297728 0.158948 0.270041 NaN
0.194534];                 %number od Th cells in 10^9 cells per L
     Tcdata = [0.508365 0.309593 0.275869 NaN
0.24687];                  %number of Tc cells in 10^9 cells per L
     Ndata = [4.86744 3.77982 4.37175 3.87729
3.70816];                  %number of cancer cells in 10^9 cells per L
 end

 tdatamax = tdata(end);
 tend =
100*365;                                                    %time ends
at 100 years in days
```

Solution for no treatment

```
 T0 = 0;
 Tfinal  = tend;

 solNT = Battle( T0, Tfinal, par, init);

 time = solNT.x;
 N    = solNT.y(1,:).*(solNT.y(1,:)>0.5e-9);
 Th   = solNT.y(2,:);
 Tc   = solNT.y(3,:);
 Treg = solNT.y(4,:);
```

Solution for treatment phase

```
 YearOfTreatmentStart = tdata(1);
 tstart2  = YearOfTreatmentStart;
 Treatmentperiod = 10*365;
 tfinal2   = YearOfTreatmentStart + Treatmentperiod;

 %effects on parameters of Interferon treatment
 par2         = par;
 par2.d1      = par.d1 * 303.75;                          %killing of
cancer cells by Tc cells increases (factor 100)
 par2.d6      = par.d6 + 2.4000e-04;                      %Treg start
killing Tc
 par2.d3      = par.d3 * 0;                               %Treg stop
killing Th
 par2.k2       =par.k2 * 2;                               %baseline
production of Th cells
 par2.k5       =par.k5 * 2.2;                             %baseline
production of Tc cells
```

```
idx          = find(time>=tstart2,1);                          %time index for
treatment starts

init2 = [solNT.y(1,idx), solNT.y(2,idx), solNT.y(3,idx), solNT.y(4,idx)];
sol2 = Battle(tstart2, tfinal2, par2, init2);

time2 = sol2.x;
N2    = sol2.y(1,:).*(sol2.y(1,:)>1e-9);
Th2   = sol2.y(2,:);
Tc2   = sol2.y(3,:);
Treg2 = sol2.y(4,:);

idx3  = find(time2>=tfinal2,1);
init3 = [sol2.y(1,idx3), sol2.y(2,idx3), sol2.y(3,idx3), sol2.y(4,idx3)];
```

Solution for discontinued treatment/ treatment pause

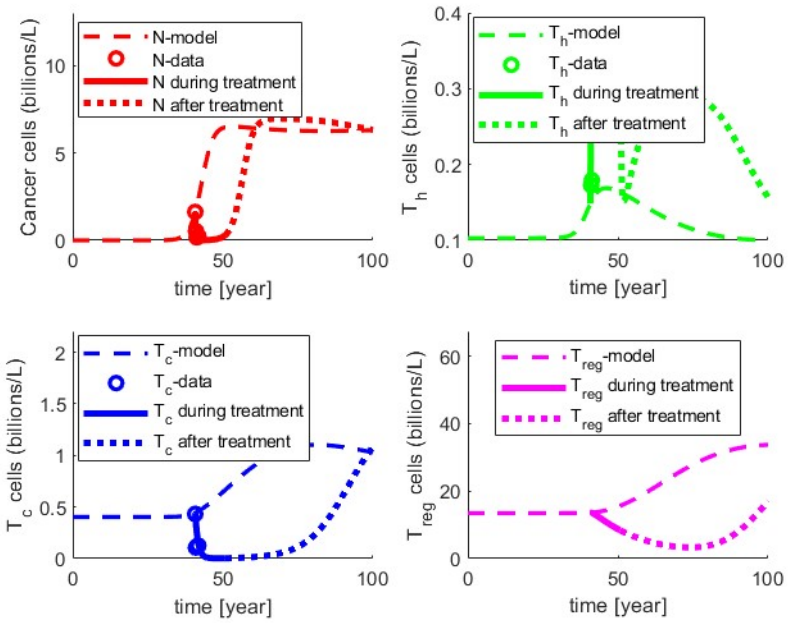
```
tstart3 = tfinal2;
tfinal3 = tend;


solTP = Battle( tstart3, tfinal3, par, init3);

% Transforming into nice variable names
time3 = solTP.x;
N3    = solTP.y(1,:).*(solTP.y(1,:)>0.5e-9);
Th3   = solTP.y(2,:);
Tc3   = solTP.y(3,:);
Treg3 = solTP.y(4,:);

par.d1
```

```
ans = 1.4400e-04
```

Plots

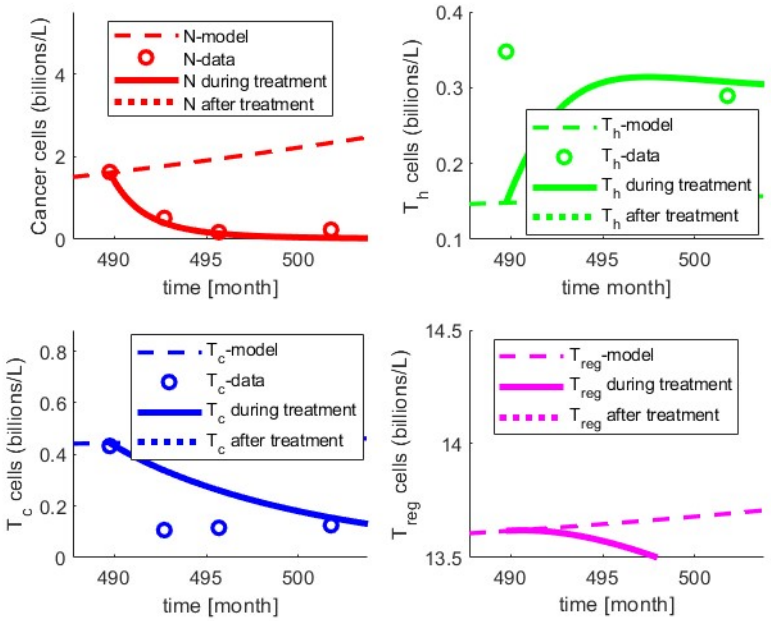
```
%-------------- Figure 1 ----------------------------------------------------

figure(1)   %shows cell count for full 100 years

%graph for cancer cells
subplot(2,2,1)
hold on
plot(time/365,N,'--r', tdata/365,Ndata,'or', 'linewidth',2)
plot(time2/365,N2,'-r', 'linewidth',3)
plot(time3/365,N3,':r', 'linewidth',3)
```

```
 xlabel('time [year]');
 ylabel('Cancer cells (billions/L)');
 legend('N-model','N-data','N during treatment','N after
treatment','location','best');
 ylim([0 2*max(N)]);

 %graph for Th cell count
 subplot(2,2,2)
 hold on
 plot(time/365,Th,'--g', tdata/365,Thdata,'og', 'linewidth',2)
 plot(time2/365,Th2,'-g', 'linewidth',3)
 plot(time3/365,Th3,':g', 'linewidth',3)
 xlabel('time [year]');
 ylabel('T_h cells (billions/L)');
 legend('T_h-model','T_h-data','T_h during treatment','T_h after
treatment','location','best');

 %graph for Tc cell count
 subplot(2,2,3)
 hold on
 plot(time/365,Tc,'--b', tdata/365,Tcdata,'ob', 'linewidth',2)
 plot(time2/365,Tc2,'-b', 'linewidth',3)
 plot(time3/365,Tc3,':b', 'linewidth',3)
 xlabel('time [year]');
 ylabel('T_c cells (billions/L)');
 legend('T_c-model','T_c-data','T_c during treatment','T_c after
treatment','location','best');
 ylim([0 2*max(Tc)]);

 %graph for Treg cell count
 subplot(2,2,4)
 hold on
 plot(time/365,Treg,'--m', 'linewidth',2)
 plot(time2/365,Treg2,'-m', 'linewidth',3)
 plot(time3/365,Treg3,':m', 'linewidth',3)
 xlabel('time [year]');
 ylabel('T_{reg} cells (billions/L)');
 legend('T_{reg}-model','T_{reg} during treatment','T_{reg} after
treatment','location','best');
 ylim([0 2*max(Treg)]);
```




```
 %------ Figure 2 ------------------------------------------------------------
        %--------Zoomed in on data/treatment area.-------------
  figure(2)

  subplot(2,2,1)
  hold on
  plot(time*12/365,N,'--r', tdata*12/365,Ndata,'or', 'linewidth',2)
  plot(time2*12/365,N2,'-r', 'linewidth',3)
  plot(time3*12/365,N3,':r', 'linewidth',3)
  hold off
  xlabel('time [month]');
  ylabel('Cancer cells (billions/L)');
  legend('N-model','N-data','N during treatment','N after
treatment','location','best');
  xlim([tstart2*12/365-2 tdatamax*12/365+2]);
  ylim([0 5.5]);


  subplot(2,2,2)
  hold on

  plot(time*12/365,Th,'--g', tdata*12/365,Thdata,'og', 'linewidth',2)
  plot(time2*12/365,Th2,'-g', 'linewidth',3)
  plot(time3*12/365,Th3,':g', 'linewidth',3)
  hold off
  xlabel('time month]');
  ylabel('T_h cells (billions/L)');
```

```
  legend('T_h-model','T_h-data','T_h during treatment','T_h after
treatment','location','best');
  xlim([tstart2*12/365-2 tdatamax*12/365+2]);
  ylim([0.1 0.4]);

  subplot(2,2,3)
  hold on
  plot(time*12/365,Tc,'--b', tdata*12/365,Tcdata,'ob', 'linewidth',2)
  plot(time2*12/365,Tc2,'-b', 'linewidth',3)
  plot(time3*12/365,Tc3,':b', 'linewidth',3)
  hold off
  xlabel('time [month]');
  ylabel('T_c cells (billions/L)');
  legend('T_c-model','T_c-data','T_c during treatment','T_c after
treatment','location','best');
  xlim([tstart2*12/365-2 tdatamax*12/365+2]);
  ylim([0 0.8*max(Tc)]);

  subplot(2,2,4)
  hold on
  plot(time*12/365,Treg,'--m', 'linewidth',2)
  plot(time2*12/365,Treg2,'-m', 'linewidth',3)
  plot(time3*12/365,Treg3,':m', 'linewidth',3)
  hold off
  xlabel('time [month]');
  ylabel('T_{reg} cells (billions/L)');
  legend('T_{reg}-model','T_{reg} during treatment','T_{reg} after
treatment','location','best');
  xlim([tstart2*12/365-2 tdatamax*12/365+2]);
  ylim([13.5 14.5]);
```



Function for solving the equations

```
 function SOL = Battle(T0, Tfinal, par, init)

      opt = odeset('RelTol',1e-8,'AbsTol',1e-10);
      SOL = ode45(@RHS,[T0 Tfinal],init,opt,par);    %solves ordinary
differential equations

 end
```

Differential equations

```
 function  dz = RHS(~,z,par) %RHS=right hand side, z: array containing N, Th,
Tc, and Treg

      N    = z(1)*(z(1)>0.5e-9);
      Th   = z(2);
      Tc   = z(3);
      Treg = z(4);

      dN    = par.k1 * N * (1 - (N/par.L)) - par.d1 * N * Tc;
      dTh   = par.k2 + par.k3*N/(par.k4+N)*Th - par.d2 * Th - par.d3 * Treg *
Th;
      dTc   = par.k5 + par.k6*Th*Tc - par.d4 * Tc - par.d5 * N * Tc  - par.d6
* Treg * Tc;
      dTreg = par.k7*Tc - par.d7*Treg;

      dz = [dN; dTh; dTc; dTreg];

 end
```