Aus dem Institut für Medizinische Informatik und Biometrie
Direktor: Herr Prof. Dr. Ingo Röder

Application of mathematical modelling to describe and predict treatment dynamics in patients with *NPM1*-mutated Acute Myeloid Leukaemia (AML)

D i s s e r t a t i o n s s c h r i f t

zur Erlangung des akademischen Grades
Doctor rerum medicinalium
(Dr. rer. medic.)
vorgelegt
der Medizinischen Fakultät Carl Gustav Carus
der Technischen Universität Dresden

von

M.Sc. Helene Hoffmann

aus Halle(Saale)

Dresden 2022

2. Blatt (2. Seite)

1. Gutachter: Prof. Dr. Ingo Röder
2. Gutachter: Prof. Dr. Martin Bornhäuser

Tag der mündlichen Prüfung: 25. Mai 2023

gez.: Prof. Dr. Henning Morawietz
Vorsitzender der Promotionskommission

Contents

Summary

Background: Acute myeloid leukaemia (AML) is a severe form of blood cancer, which in many cases can not be cured. Although chemotherapeutic treatment is effective in most cases, often the disease relapses. To monitor the course of disease, as well as to early identify a relapse, the leukaemic cell burden in the bone marrow is measured. In the genome of these cells certain mutations can be found, which lead to the occurrence of leukaemia. One of those mutations is in the neucleophosmin 1 (*NPM1*) gene. This mutation is found in about one third of all AML patients. The burden of leukaemic cells can be derived from the proportion of *NPM1* transcripts carrying this mutation in a bone marrow sample. These values are measured routinely at specific time points during treatment and are then used to categorise the patients into defined risk groups. In the studies, the data for this work originates from, the *NPM1* burden was measured beyond the treatment period. That leads to a more comprehensive picture of the molecular course of disease of the patients.

Hypothesis: My hypothesis is that the risk group categorisation can be improved by taking into account the dynamic time course information of the patients. Another hypothesis of this work is that with the help of statistical methods and computer models the time course data can be used to describe the course of disease of AML patients and assess whether they will experience a relapse or not.

Materials and Methods: For these investigations I was provided with a dataset consisting of quantitative *NPM1* time course measurements of 340 AML patients (with a median of 6 measurements per patient). To analyse this data I used statistical methods, such as correlation, logistic regression and survival time analysis. For a better understanding of the course of disease I developed a mechanistic model describing the dynamics of the cell numbers in the bone marrow of an AML patient. This model can be fitted to the measurements of a patient

by adjusting two parameters, which represent the individual severity of disease. To predict a possible relapse within 2 years after beginning of treatment, I used data that was generated using the mechanistic model (synthetic data). For the prediction three different methods were compared: the mechanistic model, a recurrent neural network (RNN) and a generalised linear model (GLM). Both, the RNN and the GLM were trained and tuned on part of the synthetic data. Afterwards all three methods were tested using the so far unseen part of the data set (test data).

Results: Following the analysis of the data I found that the decreasing slope of *NPM1* burden during primary treatment as well as the absolute burden after the treatment harbour information about the further course of disease. Specifically I found that a faster decrease of NPM1 burden and a lower final burden lead to a better prognosis. Further I could show that the developed simple mechanistic model is able to describe the course of disease of most patients. When I divided the patients into two different risk groups using the fitted parameters from the model I could show that the patients in those groups show distinct relapse-free survival times. The categorisation using the parameters lead to a better distinction of groups than using current categorisation by the WHO. Further, I tried to predict a 2-year relapse using synthetic data and three different prediction methods. I could show that it had nearly no impact at all which method I used. Much more important, however, was the quality of data. Especially the sparseness of data, which we find in the time courses of AML patients, has a considerable negative effect on the predictability of relapse. Using a synthetic data set with measurement times oriented on the times of chemotherapy I could show that a sophisticated measurement scheme could improve the relapse predictability.

Conclusions: In conclusion, I suggest to include the dynamic molecular course of the *NPM1* burden of AML patients in clinical routine, as this harbours additional information about the course of disease. The involvement of a mechanistic model to asses the risk of AML patients can help to make more accurate predictions about their general prognosis. An accurate prediction of the time of relapse is not possible. All three used methods (mechanistic model, statistical model and neural network) are in general suitable to predict relapse of AML patients. For reliable predictions, however, the quality of the data needs to be drastically improved.

Zusammenfassung

Hintergrund: Akute myeloische Leukämie (AML) ist eine schwere Form von Blutkrebs, die auch heute noch nicht gut behandelbar ist. Obwohl die chemotherapeutische Behandlung in vielen Fällen anschlägt, kommt es oft zu einem Rückfall. Um den Verlauf der Krankheit zu beobachten und auch einen Rückfall frühzeitig zu erkennen, wird die Belastung an leukämischen Zellen im Knochenmark bestimmt. In dem Genom dieser Zellen findet man bestimmte Mutationen, die dazu geführt haben, dass eine Leukämie entsteht. Eine solche Mutation, die bei etwa einem Fünftel der Patienten auftritt, ist in dem Nucleophosmin-Gen (*NPM1*). Der Anteil an *NPM1*-Transkripten mit dieser Mutation in einer Knochenmarksprobe gibt Aufschluss darüber, wie hoch die Belastung mit leukämischen Zellen ist. Diese Werte werden standardmäßig zu bestimmten Zeitpunkten im Behandlungsverlauf gemessen und fließen dann in eine Einschätzung der Risikogruppe der Patienten ein. Im Rahmen von den Studien aus denen die in der Arbeit verwendeten Daten stammen, wurde die *NPM1*-Belastung über die Zeit der Behandlung hinaus bestimmt. Damit ergibt sich ein Bild des molekularen Krankheitsverlaufs der Patienten.

Hypothese: Meine Hypothese ist nun, dass die Einschätzung der Risikogruppe durch eine Betrachtung des dynamischen Zeitverlaufs solcher Messungen verbessert werden kann. Eine weitere Hypothese dieser Arbeit ist, dass mit Hilfe von statistischen Methoden und Computermodellen, diese Zeitverlaufsdaten genutzt werden können, um den molekularen Krankheitsverlauf von AML-Patienten zu beschreiben und daraus abzuleiten, ob ein Patient einen Rückfall erleiden wird oder nicht.

Material und Methoden: Für diese Untersuchungen stand mir ein Datensatz von 340 AML-Patienten mit quantitativen Messungen einer Belastung mit *NPM1*-mutierten leukämischen Zellen im Zeitverlauf zur Verfügung (mit im Median 6 Messungen pro Patient). Diese ha-

be ich mit Hilfe von statistischen Methoden, wie Korrelation, logistischer Regression oder Überlebenszeit-Analyse untersucht. Für das tiefere Verständnis des Krankheitsverlaufs wurde ein mechanistisches mathematisches Modell entwickelt, welches die Dynamiken der Zellzahlen im Knochenmark eines AML-Patienten beschreibt. Dieses Modell lässt sich durch die Veränderung von zwei Parametern, die die individuelle Schwere des Verlaufs eines Patienten widerspiegeln, an die experimentellen Daten für jeden Patienten anpassen. Um einen möglichen Rückfall innerhalb der ersten 2 Jahre nach Behandlungsbeginn vorherzusagen, wurden Daten verwendet, die mit dem mechanistischen Modell generiert wurden. Dafür wurden 3 verschiedene Methoden für Vorhersagen verglichen: das mechanistische Modell, ein rekurrentes neuronales Netz (RNN) und ein generalisiertes lineares Modell (GLM). Sowohl das RNN, als auch das GLM wurden mittels eines Teils der Daten trainiert und angepasst. Anschließend wurden alle drei Methoden auf dem verbliebenen Teil der Daten (Testdaten) auf ihre Vorhersagegenauigkeit getestet.

Ergebnisse: Die Datenanalyse führte zu dem Ergebnis, dass sowohl der Abfall der *NPM1*-Belastung während der initialen Behandlung, als auch die absolute *NPM1*-Belastung nach Abschluss der Behandlung Informationen über den weiteren Verlauf der Krankheit enthalten. Genauer bedeutet das, dass ein schnellerer Abfall, sowie eine niedrigere finale *NPM1*-Belastung zu einer besseren Prognose führten.
Weiter konnte ich zeigen, dass ein einfaches mechanistisches Modell, wie ich es entwickelt habe, in der Lage ist den Krankheitsverlauf von den meisten Patienten zu beschreiben. Anhand der angepassten Modellparameter lassen sich die Patienten in Risikogruppen einteilen. Ich konnte zeigen, dass diese Gruppen unterschiedliche rückfallfreie Überlebenszeiten aufweisen. Diese Einteilung der Gruppen führte zu einer besseren Unterscheidung der weiteren Krankheitsprognose als die aktuelle Risikogruppeneinteilung der WHO. Im weiteren versuchte ich mit Hilfe von drei verschiedenen Vorhersagemethoden und synthetischen Daten vorherzusagen, ob ein AML-Patient innerhalb von 2 Jahren einen Rückfall erleiden wird. Dabei zeigte sich, dass die verwendete Methode kaum einen Einfluss darauf hat, wie gut eine Vorhersage funktioniert. Viel entscheidender hingegen war die Qualität der Daten. Besonders die geringe Dichte an Datenpunkten, die wir bei den Zeitverläufen der AML Patienten finden, hat einen erheblichen negativen Effekt auf die Vorhersagbarkeit eines Rückfalls. Anhand eines simulierten Datensatzes mit Messpunkten zu Zeiten, die an den Chemotherapiezeiten orientiert sind, konnte ich zeigen, dass so ein durchdachtes Messschema die Vorhersagbarkeit eines Rückfalls verbessern kann.

Schlussfolgerungen: Die Ergebnisse führen zu den Schlussfolgerungen, dass es durchaus ratsam ist, den dynamischen molekularen Verlauf der *NPM1*-Belastung bei AML-Patienten zu berücksichtigen, da dieser wichtige Hinweise auf den weiteren Krankheitsverlauf liefern kann. Das Einbeziehen eines mechanistischen Modells in die Risikoabschätzung der AML-Patienten kann einen Beitrag dazu leisten, genauere Vorhersagen über die allgemeine Prognose zu treffen. Eine genaue Abschätzung des Rückfallzeitpunktes ist damit jedoch nicht möglich. Alle drei verwendeten Methoden (mechanistisches Modell, statistisches Modell und neuronales Netz) sind geeignet um Vorhersagen zum Rückfall von AML-Patienten zu machen. Für eine zuverlässige Vorhersage muss die Qualität der Daten allerdings deutlich verbessert werden.

List of Abbreviations

AML acute myeloid leukemia

AraC D-arabinosyl cytosine

CIF cumulative incidence function

CML chronic myeloid leukemia

KM Kaplan-Meier

NPM1 *nucleophosmin 1*

ODE ordinary differential equation

qPCR quantitative polymerase chain reaction

Foreword

Dear reader,
I appreciate that you decided to read this dissertation. Before you start reading, I want to give you a small introduction to the purpose of this work and the style used to achieve this. This dissertation is a cumulative work. That means that the three scientific publications contained here, have already been published and are available to the entire scientific community. Furthermore, to understand this work entirely expert knowledge from multiple disciplines, such as medicine, mathematics and computational science, are required. Therefore, the purpose of this work, is not to summarize my achieving again for multidisciplinary experts, but instead, I would like to make my results available to a broader audience. This is done here, by using easy language, explaining technical terms and further trying to make the read as enjoyable and easy as possible. I choose this to be the additional challenge of my PhD, as it is one thing to be able to communicate your findings within an informed community, but a totally different thing to make it understandable for everyone. From my perspective this ability is important for every scientist, as not only a small group of chosen people should have access to the whole knowledge, but everyone should get the chance of understanding what is going on in the scientific world. Therefore, I will give a detailed and easily readable introduction into the main topic of this work, as well as understandable introductions to the background of each of the three manuscripts followed by a summary of the main findings. In the discussion at the end I will explain what the limits of the findings are and what can be implied from them and what not. I hope this will help to give everyone the chance of understanding the work I have done in the last years and what it means for leukaemia patients and other scientists.
I want to take you on a journey, a journey through my years as a PhD candidate, through obstacles and achievements, frustration and motivation. Let me inspire you with my fascination for clinical time course data and how they can be predictive about the further course of the disease. Let me take you away with my excitement to see how a mechanistic mathematical

model can be used to gain insights into the disease dynamics of individual patients and let me infect you with my passion for modern machine learning methods and how they can be used to predict the relapse for leukaemia patients.

Chapter 1

Introduction

In this work, I will show you what interesting insights can be gained, when studying the dynamic behaviour over time of the proportion of malignant cells (tumour burden) in acute myeloid leukemia (AML) patients throughout the course of the disease, based on a unique data set. Together we will explore how this dynamic information can enlighten us about the further prospects of a patient. We will further see how a simply structured mechanistic model of the bone marrow, where leukaemic cells reside, can not only reproduce the dynamics of the tumour burden throughout the disease, but also provide unexpected deep insights into AML disease dynamics in general and in the disease characteristics of each individual patient. And the last part of this work will confront us with the astonishing peculiarities of three different machine learning methods: mechanistic models, statistical models and deep neural networks and how they prove themselves in predicting the relapse of AML patients. The overall aim of this study was to gain insights of how dynamic time course information of leukaemia patients can be used for prognostic purposes. This study takes place at the medical faculty of TU Dresden, where I am provided with a set of AML time course data and the task to "see what's in there".

1.1 Clinical background of AML

1.1.1 Symptoms

But what exactly is this AML I already repeatedly talked about? AML is a disease, which is also known as a form of "blood cancer", even a really harsh form. The first thing persons with

AML notice about this disease is probably that their skin becomes paler, they are often tired, it becomes difficult to breathe, they might get a fever and start to bleed frequently in the mouth. All in all they will feel miserable and therefore, will visit the doctor.

1.1.2 Diagnosis

The first thing a doctor does then, is taking a blood sample and sending it to a lab to analyse the different blood cell counts. Probably the lab will find reduced numbers of red blood cells. This reduction is called anaemia. It explains the paleness, as the healthy rosy colour of the skin comes from the circulating blood. Also the tiredness and the breathlessness is caused by the anaemia, as the red blood cells are responsible for the oxygen distribution in the body. But that is not everything, as also the number of platelets will probably be reduced. This reductions is called thrombocytopenia. Platelets are the cells in the blood that are important for wound healing and a lack leads to frequent bleeding in places where the skin is easily torn, like in the mouth or the nose, but also inner organs can be affected. Further, the white blood cell count might additionally be increased, which is also a hint for AML. About the reasons for this I will talk in a later section (Section 1.2.2).
Following these quantitative blood results, the lab will additionally look at the blood cells under the microscope. There, they will see that some of the white blood cells do not look like ordinary white blood cells, but they are larger and have a notably enlarged nucleus. These abnormal white blood cells are called "blasts". The normal white blood cells as well as the blasts are counted and if the proportion of blasts is at least 20% of all cells, AML is diagnosed (normal blood should contain no blasts at all) (Arber, 2019). When the doctor receives these results from the lab, he will send the patient directly to the hospital, where more tests take place. A bone marrow biopsy is done. This sample of the patient's bone marrow is important, as this is the place of origin of the leukaemic cells. The bone marrow sample is also tested for these blast cells, which were already found in the blood. In healthy bone marrow no more than 5% of blasts should be found, but in AML there are more than 20%. The bone marrow is further tested to find out which alterations in the genetic material (gene mutation) are present, as AML is a disease with many genetically different subtypes. That means that the genes that are often mutated in AML are analysed to find which ones harbour a mutation (Döhner et al., 2010). This examination also confirms the diagnosis. Also the genetic material is looked at under the microscope, as there are changes in the chromosomes found in AML blasts for about half of the patients, which can further help to tell if the form of AML is severe or has a

better prognosis.

1.1.3 Treatment

As AML is a disease that leads to a fast worsening of the condition of the patient the treatment is directly started at the day of diagnosis. The treatment itself is unpleasant as the only way to treat AML for nearly 50 years now is a cyclic form of chemotherapy (Lichtman, 2013). The patient is put on a continuous drip feed with D-arabinosyl cytosine (AraC) for 7 days combined with daunorubicin in the first 3 days. This leads to the nickname "7+3" (Longo et al., 2015). You may have heard about the unpleasant side effects of chemotherapy: nausea and vomiting, hair-loss, tiredness, infections and loss of appetite, to name only a few. They occur because of the unspecific effect of these chemotherapeutics. That means, that not only the leukaemic cells are targeted, but also all other cells. AraC inhibits the synthesis of the genetic material, which happens when a cell divides (Furth and Cohen, 1968; Graham and Whitmore, 1970). This inhibition leads also to an unbalance during cell growth and eventually the cell dies (Kim and Eidinoff, 1965). Therefore, the faster dividing cancer cells are killed more rapidly than healthy cells. Also daunorubicin kills faster dividing cells more rapidly by breaking the DNA (Fornari et al., 1994). But the incidental killing of healthy cells leads to immense side effects. Still, if all goes well patients will have less than 5% blasts left and the blood cell counts are back to normal after three of these chemotherapy cycles (induction therapy). This situation is called "complete remission" (CR). They will feel no signs of the disease any more and probably even think they might be cured. Unfortunately, they are most likely not cured, but a small number of remaining AML cells will grow again and that is why they will receive several more chemotherapy cycles to destroy all remaining leukaemia cells (consolidation therapy). The details of this treatment, like the dosage of the chemotherapeutic AraC can vary between patients depending on their age or general state of health. Unfortunately, reaching a CR is not equal to "winning the fight against the disease", because about half of the patients in CR experience a relapse sooner or later.

So, if the patients have a more severe form of AML by means of the symptoms and (following the analyses of the genetic material) also a higher risk of relapsing, then the exclusive treatment with chemotherapy has a low chance of curing the leukaemia. There is another possibility to help these patients: if they are not older than 65 years (a threshold defined on basis of years of experience) and in a good state of health they might get the chance for a stem cell transplantation. The goal of this treatment is to eliminate all leukaemic (stem) cells in

the body of the patients and replace them with healthy stem cells from a person who donates some of his or her stem cells. These new healthy cells (called "graft") will then generate new healthy blood cells for the patient throughout his or her life. And what is even more astonishing is that these cells also help to eliminate remaining leukaemia cells, as the graft cells perceive them as not belonging to the body of the patient. This is called the "graft-versus-leukaemia" effect (Sweeney and Vyas, 2019). Unfortunately, it may also happen, that the donated healthy cells do not recognise the patient's own cells as normal cells, but perceive them as intruders and also destroy them. This so called "graft-versus-host" disease is a high risk for patients receiving a transplantation.

1.2 Molecular Background of AML

1.2.1 Haematopoiesis

The last section was about how the disease looks like from a patient's or clinical point of view. But what is going on in the patient's blood system on a cellular level? How does the disease arise and what does the data mean that I was provided with for my analyses? To find the answer to these questions it is important to understand that the correct function of the cells in the blood is essential for the health of our body. These blood cells arise from a limited number of blood building (haematopoietic) stem cells, which reside in the bone marrow of each person. The bone marrow can be divided into sections, so called niches or micro-environments, each promoting a different kind of cell. One of these niches promote dormant stem cells (Arai and Suda, 2007). That means that these cells nearly never divide. They preserve the genetic material (DNA) as it is, as every division of a cell harbours the risk of making mistakes during the reproduction of the DNA (mutations). These cells are only activated to divide in critical situations, such as infections or chemotherapy, when many new blood cells are needed.
The stem cells that are not in a dormant state divide infrequently to produce less potent precursor cells. These again produce other cells that are even more specific, just like in a family tree, shown in Figure 1.1. This hierarchical production of all kind of blood cells is called "haematopoiesis". The last cells in line are the functional blood cells, that are important to defend the body against viral or bacterial infections and to close wounds. These have no ability to reproduce themselves, whereas stem cells are thought to have an infinite ability to reproduce themselves and that way preserving the whole blood building system. It could be shown, that also leukaemic cells are organized in a hierarchical family tree like the normal

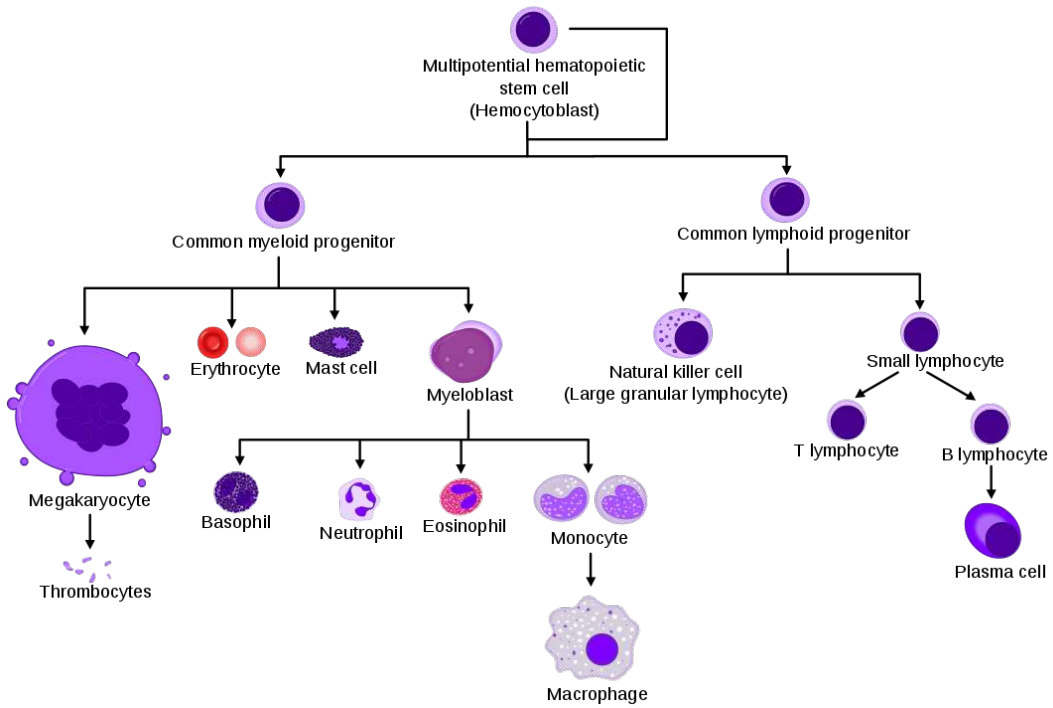

Figure 1.1: Diagram showing the development of different blood cells from haematopoietic stem cell to mature cells. By A. Rad and M. Häggström. CC-BY-SA 3.0 license.

cells (Hope et al., 2004). Non-dividing quiescent and slowly reproducing stem cell-like cells, that are able to generate an AML on their own (leukaemia-initiating cell) were found (Ishikawa et al., 2007; Lapidot et al., 1994). The leukaemic blast cells, however, can only produce a finite number of new cells and, therefore, can not maintain the disease. But now we finally want to face the question where these leukaemic cells come from and what makes them so disruptive.

1.2.2 Clonal Evolution

During a persons life, cells of the blood system divide very often and each division can lead to an error while copying the DNA for the second cell, a genetic mutation (Bertram, 2000). The entirety of all cells originating from one changed cell and therefore, harbouring the same mutation is called a "clone" (Cooper and Young, 2017). Mostly, these mutations have no consequence for the normal function of the cell and hence there coexist several different clones.

But every once in a while, a mutation happens, that changes the function of the cell minimally, giving it an advantage over the other cells of the same type. This competitive advantage can lead to more and more cells with this specific mutation, replacing the other cells. This dominance of one single clone in the blood building system (called "clonal haematopoiesis") can be found in up to 20% of the elderly people (Jaiswal et al., 2014), most of them with no damage to their health. The most common of these mutations are in some selected genes (*DNMT3A*, *TET2*) that were found to be preleukaemic. That means that these cell populations expand, as they have a growth advantage compared to cells without the mutation, but they are not malignant as they still have the ability to produce fully functional blood cells. Only combined with other mutations a preleukaemic clone can develop into a leukaemic clone. Hence, to develop AML, a person needs to have at least two different changes in the genetic material of an early blood cell ("2-hit" model (Gary Gilliland and Griffin, 2002; Welch et al., 2012)). Whether this is a stem cell or an early precursor cell is still not clear (Passegué et al., 2003) and might vary between patients (Yanagisawa et al., 2016). In fact, the genes where AML patients acquire mutations to develop the disease differ widely and many patients have no overlap with others as there are more than 20 genes known to be associated with AML (Cancer Genome Atlas Research Network, 2013). Earlier mutations usually lead to changes that promote the occurrence of more mutations (Suela et al., 2007; Yoshimi et al., 2014), for example by changing the structure of the DNA to destabilise it (epigenetic changes). Later mutations, however, enhance the ability of the cell to multiply faster than others (Corces-Zimmerman et al., 2014) or not to specify further but stay at a multi-potent and highly reproductive state (differentiation block). The mutations in AML usually lead to increased speed of reproduction of the leukaemic blasts (Guan and Hogge, 2000; Minden et al., 1978), as well as an insensitivity to the body's own control mechanisms leading to uncontrolled reproduction (Young et al., 1987). And this is now the explanation for the increased white blood cell count during a diagnosis of AML that I talked about earlier (Section 1.1.2). Because these blasts are still similar to white blood cells they are counted as such and because they reproduce faster than normal white blood cells the white blood cell count is increased. But there are other mutations in AML that can interfere with the ability of blasts to produce functional red blood cells (for oxygen transport), platelets (for wound healing) and white blood cells (for fighting infections), retaining them in an immature state (Bereshchenko et al., 2009; Pabst and Mueller, 2009; Rodriguez-Perales et al., 2015). Leukaemic blasts may also directly interfere and inhibit the healthy stem cells, by removing small signal molecules (growth factors) that are absolutely necessary for cells to survive (Rauch et al., 2016) or by producing inhibiting substances (cytokines) (Cheng et al.,

2015).

1.2.3 Relapse

But what happens to the mixture of healthy and leukaemic cells during chemotherapy? When the patient is treated with chemotherapy this treatment changes the conditions for the competing clones. Now, another characteristic of the cells becomes important: their sensitivity to chemotherapy. But as mentioned earlier, some of the leukaemia-initiating cells can be in a dormant state, without dividing at all. And as the chemotherapy depends on the active division of a cell to effect and kill it (compare explanation of mechanism of action of chemotherapeutics in Section 1.1.3), these non-dividing cells are immune to the treatment and can survive. These cells can later be activated to reproduce again, either by chance or by an external stimulus such as an infection. This will lead to a fast expansion of these malignant cells and a relapse with the same clone as the original clone at diagnosis. A relapse, however, can also appear in another way. As fast division of a cell is usually linked to high chemosensitivity the originally fittest clone becomes eradicated fastest. That can lead to another clone that is more resistant to treatment, outcompeting the original leukaemic clone (Bachas et al., 2012). If this clone still has an advantage in growth compared to the healthy cells it will expand, leading to a relapse of the disease with cells that harbour different mutations (Ding et al., 2012; Krönke et al., 2013).

1.2.4 *NPM1*-Mutation

Some mutations are more likely to reoccur in the relapse clone than others. In this work, I will focus on a subgroup of AML patients that all have a relatively stable mutation in the gene *nucleophosmin 1* (*NPM1*) (Jain et al., 2014). Actually it is not one mutation, there were several different mutations found in the *NPM1* gene in AML patients. But three of these mutations, generally called type A, type B and type D, are the most common ones, one of them found in 88% of the *NPM1*-mutated (*NPM1*-mut) AML patients (Alpermann et al., 2016). This gene encodes a protein that usually shuttles in and out of the nucleus (where the DNA is located) (Borer et al., 1989) but resides more prominently in the nucleus (Cordell et al., 1999). It is a so called tumour suppressor, which means that its normal function is essential to prevent tumour development. Its tumour suppressive function is characterised by involvement in the stabilisation and activation of the protein p53, which is an important regulator of cell division,

preventing the replication of damaged DNA (Colombo et al., 2002). A mutation of *NPM1* leads to a change of a part of the NPM protein, resulting in a nuclear export signal (NES). This not only results in its more prominent residence outside the nucleus (in the cytoplasm of the cell), but also in the inhibition of unmutated NPM by binding to it and preventing its import to the nucleus. That way it interferes twice with the normal function of NPM (Falini et al., 2006). As the mutation results in an interference of p53, the cell is able to reproduce uncontrollably, allowing the tumour cells to expand. In a mouse study it could be shown, that the removal of the NPM in the cytoplasm promotes the differentiation of the cells and prolongs the lives of the mice with a *NPM1* mutation (Brunetti et al., 2018).
There are basically two reasons why I will focus on the patients with this mutation in my work. The first one is that *NPM1*-mut patients are one of the largest subgroups in AML, with about one third of the patients harbouring this mutation (Falini et al., 2005). The second reason is the already mentioned stability of the mutation throughout the disease, making it an excellent target for monitoring the leukaemic burden over time.

1.2.5 *NPM1*-Measurements

The relative amount of *NPM1*-mutated cells (*NPM1* burden) can easily be measured using a standard laboratory method: quantitative polymerase chain reaction (qPCR) (Gorello et al., 2006). In short, during qPCR a copy of the target sequence (here the mutated *NPM1*-gene, including the three most common types of mutation (A,B and D)) is replicated again and again. So even very low occurrences of the gene in the sample are enriched dramatically to exceed the detection limit of the machine and conclusions about the total number of *NPM1* -mut transcript can be drawn. To now derive the leukaemic burden in the sample a reference needs to be measured, to estimate how many cells harbour the mutation compared to the total amount of cells in the sample. Therefore, another gene transcript is quantified in the same sample that is present in all cells in a stable amount. When measuring the *NPM1* burden, the transcript of the so called 'housekeeping gene' *ABL* is used for this purpose. These qPCR measurements of *NPM1/ABL* in the bone marrow of a *NPM1*-mut patient at multiple time points during the course of their disease are used by the doctors to asses the patient's response to the treatment as well as for early identification of a recurrence of leukaemic cells. They make up the data set for this work. We decided on using only measurements from the bone marrow and not from the blood as these are more sensitive (Ivey et al., 2016). The smallest detectable *NPM1* burden is one copy in 10^5 copies of *ABL*. The time points are based

on the recommendation of the European LeukaemiaNet: a measurement at diagnosis, one after two cycles of chemotherapy, one at the end of treatment and every 3 months in the following 2 years (Schuurhuis et al., 2018). But why these time points and not others?

1.2.6 Risk Estimation

The reason why these time points were chosen is that several studies showed that the measurements at these times have prognostic power over the further course of disease, such as the survival for 3 or 4 years or the occurrence of relapse (Balsat et al., 2017; Ivey et al., 2016; Krönke et al., 2011; Shayegi et al., 2013). Therefore, these measurements are part of the assessment of the individual risk of a patient, which basically gives information about how bad the disease is (Döhner et al., 2017). There are three risk categories for AML: favourable, intermediate and adverse risk, which are defined by the genetic abnormalities of the leukaemic cells. Whether a patient with *NPM1* mutation has favourable or intermediate risk depends on another mutation: the *FLT3*-ITD. If this mutation is not present or only in a low frequency (occurring in only a small portion of the blood cells), the patient is categorised as favourable risk and therefore, has an increased chance of survival without relapse. The presence of *FLT3*-ITD in high frequency, however, indicates intermediate risk and therefore, lower chances of recovery. Furthermore, there are general criteria used, that assess the success of treatment after the first cycles of chemotherapy to estimate the severity of disease. Such criteria are the absence of blasts in the blood, less than 5% blasts in the bone marrow and negativity for the genetic marker (e.g. *NPM1*). This risk estimation helps the doctors to estimate the overall survival of a patient as well as supports treatment decisions, such as whether a stem cell transplantation is necessary.

At this point we reached the end of the introductory chapter. So far I talked about AML from a clinical point of view as well as from a molecular point of view. It should be clear now, why the *NPM1* measurements within the data set are conducted, how this is done and what information they harbour. But what additional information about the course of disease is hiding in the data? And how can this be extracted? This will be the topic of our next chapter.

Chapter 2

Data analysis

2.1 Introduction

In this chapter I will lead you through a variety of useful methods for time course data analysis, with the goal to help you understand how this data harbours valuable information about a patient's individual disease progress. We will have a closer look at the data set of time courses of *NPM1*-mut patients and how I analysed it with respect to prognostic potential. In particular, I analysed in what respect the time courses are linked to the later course of disease, such as relapse, death or survival. Such an analysis is, in principle, not new. Shayegi et. al for example, compared different thresholds of *NPM1* burden after treatment in their ability to predict relapse. That means that they divided the patients into two groups concerning their post-treatment *NPM1* level (using different thresholds) and compared the group's risks of suffering a relapse to find the threshold that leads to the biggest difference in relapse risk (Shayegi et al., 2013). Other groups used solely the mutational profile of the patients to estimate their prognostics (Grossmann et al., 2012; Zhang et al., 2019). Another group developed a risk score including information about the patients chromosomal structure and the blast burden (Haferlach et al., 2004). The European LeukaemiaNet used a comprehensive mixture of the mentioned factors in their risk estimation recommendations (Döhner et al., 2017). So, there is already a lot known about the prognostic potential of different measures in AML. But what is so different about my analysis that it is worth to talk about? I focused on an alternative to a comparison of the genetic profile or leukaemic burden at certain time points with the outcome. I derived other factors, such as the speed of tumour burden decrease during therapy. Therefore, I included parameters in the analysis that describe the *dynamics* of

the process. That way, the change of the tumour burden over the entire disease progression is seen as an interdependent process, that can harbour additional information about the course of disease.

2.1.1 Patient Data

At first we should have a detailed look an the data set to understand what I was working with. For this analysis I was provided with the data of 797 *NPM1*-mut patients. While this seems to be a really high number at first glance, the yield when applying some quality criteria is strongly reduced. My requirement for a patient to be included in the analysis was, that I was provided with detailed information about the chemotherapy administration, e.g. the time and duration of each cycle. Furthermore, I set a threshold of at minimum 3 *NPM1*-measurements. These criteria resulted in 340 remaining patients. Details about the patient cohort can be found in the Supplementary Materials and Methods of the Manuscript (Section 2.2.1).

2.1.2 Time-course characteristics

But which parameters can be used to describe the time courses in a more dynamic way than it was done before? Some characteristics need to be defined that capture the individual features of the time courses. This is especially important, as it is not trivial to compare two time courses with each other with samples at different time points. Therefore, I defined five main characteristics of the time courses. To describe the initial decline of the leukaemic burden during therapy I defined the first characteristic as the *elimination slope* (α in Figure 1A in the manuscript, Section 2.2). It reports a measure for the speed of reduction of leukaemic burden in the bone marrow during the treatment period, consisting of several cycles of chemotherapy. It is given in log10(*NPM1/ABL*)% per day. So, an *elimination slope* of -0.5 means an average reduction of the leukaemic burden by $10^{0.5}$% (ca. 3%) every day during the treatment phase. To describe how much the therapy was able to reduce the burden in total, I defined the second characteristic as the *NPM1 level after primary treatment* (n in Figure 1A in the manuscript, Section 2.2). It describes the lowest *measured* leukaemic burden (in %) within the first 9 months after treatment start, which is usually at the end of the last therapy cycle. As the measurement frequency is relatively low this value is usually above the actual lowest value of the leukaemic burden and it was taken at some time point before the end of month 9. To describe the increase of the number of leukaemic cells during relapse I defined the third

characteristic as the *relapse slope* (β in Figure 1A in the manuscript, Section 2.2). It gives the speed of regrowth of the leukaemic burden in case of a relapse, similar to the elimination slope, only that the burden increases, not decreases. A relapse occurs when the leukaemic burden passes the relapse threshold of 1% (defined in (Shayegi et al., 2013)). To describe the duration until a reoccurrence happens I defined the forth characteristic as the *time until relapse* (d in Figure 1A in the manuscript, Section 2.2), providing the time point of the first measurement beyond the relapse threshold in days after treatment start. To describe the duration of time a patient lives with the disease I defined the last characteristic as the time of survival in days after treatment start. Not every characteristic can be derived for every patient, as not all patients have enough measurements in the phase of treatment to estimate the *elimination slope*. Also not every patient experiences a relapse or dies within the observed period. Still, these characteristics are important measures to be able to analyse whether information about the outcome after some time with the disease is already detectable shortly after diagnosis. This information can help to assess the severity of the course of disease. It can be obtained by analysing the relationship of the characteristics to each other. But how is this done?

2.1.3 Rank Correlation

One method to do so is the Spearman rank correlation. In short, this method measures if one characteristic increases while the other characteristic also increases (positive correlation) or decreases (negative correlation). This is done by assigning ranks to each value, where the smallest value has the lowest rank and the biggest value has the highest rank. After doing this with all five characteristics, we have a list of 5 ranks for each patient. Now we can sort the ranks of one characteristic in ascending order. To compare this parameter, we look at the order of the ranks after the sorting of another characteristic. If these ranks are in perfect ascending order, then the Spearman correlation coefficient (r) would be 1. If the ranks of the other parameter were in perfect descending order, r would be -1. If there were an up and down of the ranks of the other characteristic without any tendency r would be 0. Usually r lies somewhere between these values. In biology, values beyond ± 0.8 are usually associated to a strong relationship and values beyond ± 0.6 to a moderate relationship. Let us now have a quick look at the Supplementary Table S1 in the manuscript (Section 2.2.3), where we find the calculated coefficient for the characteristics. We find that the relation between the *NPM1* value after primary treatment n and the time until relapse d was granted a correlation

coefficient $r = -0.48$. This means that there is a weak relation between a deep reduction of leukaemic burden and late relapse (or poor reduction and early relapse).
But there is another value, that is calculated alongside the correlation coefficient. It is the p-value, which can be seen as measure for how certain we can be that the calculated coefficient is different from 0. Just imagine you have only very few patients available to compare two characteristics. Than a perfect correlation of $r = 1$ could as well be pure chance. Therefore, the p-value gives the probability that r is only by chance different from 0 and there is in reality no relationship between the tested characteristics. A widely used threshold for this p-value is 0.05, which means that the probability of reaching a false conclusion (that there is a relationship between the tested parameters, although there is actually none), is only 5% (false positive rate). Hence, a small p-value leads to the conclusion that it is very unlikely that there is actually no relationship between the tested characteristics. All r with a p-value smaller or equal to 0.05 are counted as being "statistically significant". Going back to our example of the time to relapse d and the *NPM1* value after primary treatment n (which had $r = -0.48$) we find a p-value of 0.0001, which leads to the conclusion that the found weak relationship is most likely present and did not occur in the data by pure chance. The 95%-confidence interval is another measure for how certain we can be that the estimated r is different from 0. The interval provides the range where the true r value lies with a probability of 95%. If this includes 0 then the tendency shown by r is not very meaningful.

2.1.4 Cumulative Incidence

So, the rank correlation is a method to analyse the relationship between two parameters. But to better understand if a difference in one of the characteristics has an actual effect on the prognosis of a patient, the patients can be divided into groups, for example in *deep* and *poor* responders, for low and high values of n (*NPM1* value after primary treatment). But how to decide whether one of the groups has a better prognosis than the other? There are more sophisticated ways for such an analysis beyond comparing the times of relapse free survival. Methods that also take into account that for still living patients you can not know the survival time. One of these methods is the cumulative incidence function (CIF). In general this function gives the probability of an event to occur until a certain time. There is always a target event, for example death due to AML and there might be competing events and censoring events. A competing event is an event, that, when happening, changes the probability of the target event to occur. A standard example is the stem cell transplantation, as this is performed to prolong

the survival of an AML patient and therefore changes the probability of death due to AML. A censoring event, however, is for example the end of the study before death occurred. We can not tell when this patient might die. But it is assumed that the probability of dying has not changed and the patient is still at risk of dying due to AML. So, when estimating the probability of the target event using a set of patient data, the CIF takes into account the occurrence of competing and censoring events. This method can be used to analyse the impact of the value of one characteristic, such as n, on the survival. Therefore, we split the patients in two groups, *deep* and *poor* responders, and calculate the CIF for each of the groups. When plotting them we see whether there is a difference between the curves of the two groups. Looking at Figure 1C in the manuscript (Section 2.2) we can see a clear difference between the two groups. And yet again we can estimate the probability that this difference occurred just by chance, although the groups are actually similar with respect to the characteristic, as the p-value of the Gray test (Gray, 1988)). This p-value was calculated to be 0.1, so, there is a 10 % probability that the difference occurred just by chance and the groups are not actually different from each other. To determine the best threshold between poor and deep responders with respect to n, leading to the biggest difference between the groups, the p-value can help, by selecting the threshold that leads to the smallest p-value.

2.1.5 Group-wise Comparison

To be sure that one result is really correct it is not uncommon to use another method to confirm earlier findings. To confirm the afore mentioned relation between a low n and a later time of relapse d I used the splitting of the patients into the two groups of poor and deep responders and analysed how these groups differ in their time of relapse. This leads us to a method for group-wise comparisons. Therein, boxplots are a simple graphical method to get an impression whether a characteristic is different in two groups. In a boxplot the median of the data is marked. A box around the median shows the lower and upper quartile of the data, that means that it includes all values except one quarter of the higher values and one quarter of the lower values. The minimum and maximum (excluding very extreme values) are then marked with the whiskers. The boxplots for the poor and deep responders for their time until relapse can be found in Figure 2B in the manuscript (Section 2.2, with a clearly visible difference between the groups. There is again a test principle (U-test) to estimate the probability (p-value), that a possible difference between two groups with respect to the observed parameter occurred just by chance. In our example the p-value was calculated to

be 0.001, leading to the conclusion that this difference is truly there.

2.1.6 Logistic regression

Now, we talked about how to find relationships and differences in two groups. But if you know some specific characteristic of a patient, e.g. the NPM1 level after primary treatment, would it not be exciting to be able to estimate the chance of surviving the next 5 years, just from this single characteristic? There is a method that can be used to do exactly that: the logistic regression. For this, a binary dependent variable is needed, a variable that has exactly two outcomes, such as dying in or surviving the next 5 years (coded as 0 and 1). Every patient is categorised in one of these outcomes. To analyse the impact of one specific characteristic (here NPM1 level after primary treatment), a logistic function can be fitted to this dependency between the characteristic and the outcome. A logistic function is a monotone function that can attain values between 0 and 1. The extreme case would be a step function where the function is 0 until some threshold is reached and than becomes 1. In practice, this is a more subtle transition. The steepness of this transition, however, gives a hint about how good the used variable can differentiate between the two outcomes, here how suitable the NPM1 level after primary treatment is to estimate the 5-year survival probability. The logistic function itself then estimates the probability of the 5-year survival depending on a patient's NPM1 level after primary treatment. Further, the fitted function can give information about the odds of dying within 5 years when the NPM1 level after primary treatment increases by 1. An example of such a logistic regression can be found in Figure 1E in the manuscript (Section 2.2), where the steepness of the curve indicates that a perfect prediction of the 5-year survival using only the *NPM1* level after primary treatment is not possible.

2.1.7 Summary of results

As already mentioned, I used the characteristics derived from the time course data to find connections between them. Using the Spearman rank correlation, I found that an earlier relapse and a faster growing relapse have a negative impact on the survival time (compare Supplementary Materials 2.2.3). A higher residual *NPM1* level after treatment was furthermore connected with an earlier relapse. To further investigate prognostic potentials of the characteristics, I calculated the CIF to first compare patients with a slow and a fast therapy response (measured by the elimination slope α) and second to compare poor and deep responders

(measured by the minimal *NPM1* level). This lead to the conclusion, that both characteristics have prognostic potential, as the groups clearly differ in their cumulative incidence of death (see Figure 1B and 1C in the manuscript), where deep responders and those with a fast therapy response show the better prognosis. An analysis of different time points for the evaluation of the minimum *NPM1* level showed, that the optimal time is 9 months after therapy start as the 5-year cumulative incidence of death shows a visible difference when comparing deep and poor responders (see Figure 1D in the manuscript). A logistic regression (Figure 1E in the manuscript) to analyse how the 5-year survival depends on the minimal *NPM1* after treatment n showed, that if a patient has an n that is about one $\log_{10}$-scale lower than that of another patient, the chance of surviving 5 years is nearly twice as high as that of the other patient. When further comparing the poor and the deep responders using boxplots, I could show that deep responders tend to have a faster elimination of the leukaemic burden during therapy (smaller elimination slope α), as well as a later time of relapse (see Figure 2A and 2B in the manuscript). Further, I could show that higher *FLT3-ITD* burden is connected to higher n and earlier death (see Figure 2C and 2D in the manuscript). Finally, I showed, that deep responder have a visibly more efficient reduction of leukaemic burden during the treatment period compared to poor responders (see Figure 2E in the manuscript). Concluding from all these results I found that looking at the dynamics of the leukaemic burden during the course of disease can substantially add to risk assessment for *NPM1*-mut AML patients.

2.2 Manuscript

Authors: H. Hoffmann, C. Thiede, I. Glauche, M. Kramer, C. Röllig, G. Ehninger, M. Bornhäuser and I. Roeder
Title: The prognostic potential of monitoring disease dynamics in *NPM1*-positive acute myeloid leukemia
Journal: Leukemia 33, 1531 - 1534
Year: 2019
DOI: 10.1038/s41375-018-0371-y

This paper is not available as Open Access and cannot be included in this document.

2.2.1 Supplementary Materials and Methods

This paper is not available as Open Access and cannot be included in this document.

2.2.2 Supplementary Figures

This paper is not available as Open Access and cannot be included in this document.

2.2.3 Supplementary Table

This paper is not available as Open Access and cannot be included in this document.

Chapter 3

Mechanistic modelling

3.1 Introduction

In the previous chapter I introduced you to the descriptive data analysis, where the pure data was used to gain insights into the disease dynamics. But this approach is neglecting all the further knowledge that was derived by biologists in many years of hard work. Would it not be great to be able to combine background knowledge about disease mechanisms with the time-course patient data? The characteristics derived in the previous chapter are solely descriptive, but what if we could derive patient-specific parameters that give insights about the underlying biological process, such a the speed of reproduction of the leukaemic cells? So, in this chapter I will focus on the development, adjustment and analysis of a mechanistic computer model that is able to simulate the time courses of the leukaemic burden in AML patients in the bone marrow. The aim of this part is to show how such a model: (i) is able to reproduce the patients' individual dynamic course of molecular leukaemic burden (fraction of leukaemic cells in the bone marrow), (ii) to gain a better understanding of the source of the differences between individual courses of disease, e.g. why some patients relapse earlier than others and (iii) to use this information to improve risk classifications.

3.1.1 Model Development

But how can it be done to describe biological mechanisms mathematically, such as the effect of chemotherapy or of the normal reproduction of cells on the cell numbers? Therefore, it is especially important to be able to describe how the cell numbers change over time, taking into

account the current cell numbers. This surmounts the abilities of statistical models, such as I introduced in the last chapter. But there is a class of equations that are perfectly suitable for this purpose: the ordinary differential equations (ODEs). ODEs are a powerful tool for the description of dynamic processes. They can describe how a variable x (here number of cells) changes with time t within the considered system (here a patient's bone marrow). This can be mathematically written as: $\frac{dx}{dt}$, which means *change of* x (dx) over *change of* t (dt). To describe the dynamics of AML I chose a model structure that is similar to a model for a different type of leukaemia, chronic myeloid leukemia (CML)(Roeder et al., 2006). The department I am working in has many years of experience with this model. It is characterized by two different compartments, which model an active stem cell niche where the cells can reproduce themselves and an quiescent stem cell niche, where no reproduction takes place. The change of the cell numbers of a specific type (leukaemic or healthy) in a specific niche (actively reproducing ar dormant) is influenced by a mixture of influx and efflux. Influxes are the reproduction of the cells and the entering of cells from another niche, whereas the effluxes are chemotherapeutic kill, differentiation or exit to another niche. The fluxes are characterized by specific rates that describe how many cells of the type leave or enter the niche in one way or another, within one time step. The rates of influx and efflux in this system are the transition rates between the two niches, the differentiation rate, the reproduction rate and the chemotherapeutic kill rate. The kill rate of chemotherapy can be switched on and off, to mimic the cyclic treatment in AML patients. A schematic overview of the system can be found in Figure 1a in the manuscript (Section 3.2). An important factor for the dynamics of the system is, that every niche has a finite capacity and if a it is already completely filled with cells there is no space, or there are no resources for more cells to enter. Therefore, the movement in the system is influenced by a total capacity of each niche, as well as the current number of cells of the niche. Concretely, this means that if the niche is completely full (the total number of cells in the niche equals the maximum capacity) then the influx term in the ODE becomes zero. This is only a theoretical consideration as there is always influx and efflux happening at the same time. If all influx equals the efflux we have no change in the number of cells any more ($\frac{dx}{dt} = 0$). This state is called the steady state.

All this influx and efflux knowledge leads to the equations describing the changes in the bone marrow of a patient, which can be found on page 4 in the manuscript (Section 3.2). But what we actually want is to find a way to simulate the leukaemic burden of a patient. Therefore, we need to know how many cells of each type are in the system at each time point. To get this information we need two things: values for the parameters of the equations (the rates

and capacities) and a solution of the ODEs. The solution of the ODE is another equation that describes the absolute number of cells x in dependence of the time t, also written as $x(t)$. Unfortunately, finding an exact solution by solving the equations is only possible in rare cases and, therefore, we use a numerical method (Runge-Kutta method) to find an approximation of the solution. The model's parameters are mostly taken from other studies or from the earlier mentioned model, I derived this model from (see also the Materials and Methods Section in the manuscript). Using the approximated solution of the ODEs and the parameter values we can now compute the number of each cell type at each point in time, given the number of the cells at the beginning. From the so gained absolute cell numbers we can calculate the relative leukaemic burden in the bone marrow, which is then comparable to the *NPM1/ABL* values of real patients.

3.1.2 Model fit to patient data

With this model I want to reproduce every patient's individual course of disease. How can that be done? To do so, I decided to keep all model parameters for all patients fixed to the same value, except for two of them, which describe the features with the largest impact on the AML dynamics. One of them is the proliferation rate of the leukaemic cells, which gives the speed of reproduction of the leukaemic cells in the bone marrow. The other is the transition rate from dormancy to active reproduction of the leukaemic cells, which basically gives the speed for leukaemic cells to enter the state, where they can not only reproduce themselves but can also be killed by chemotherapy. Before trying to find the best parameter combination for a patient the treatment information of this patient is also given to the model. That means that we tell the model when to switch on the effect of the chemotherapeutic kill and when to switch it off, to mimic the cyclic treatment of the AML patient. To then find the values for the two parameters that lead to the best possible match for this patient, the parameters are varied and the resulting solution by the model is compared to the *NPM1/ABL* measurements of this patient. This is repeated systematically with different parameter values until the optimal values are found that lead to the smallest possible difference between the approximated model solution and the patient data. This way I get a pair of values of the fitted parameters for each individual patient, that I than analyse for their relation to the patient's time of relapse.

3.1.3 Existing AML models

At this point the question arises, why such a personalized model for AML patients was not developed earlier. I guess it is because of the lack of such a relative unique and large data set as I have it available, which took years to gather. But I am not the first person to think of a mathematical model describing AML. So, what kind of models did others develop before me and with which goal? Such models date already back until the 70's where S. Rubinov and J. Lebowitz developed the first model, describing the disease dynamics in a similar way as I did (Rubinow and Lebowitz, 1976a,b). They describe a dormant state and an active state, a competition between healthy and leukaemic cells, as well as cyclic chemotherapy administration, although they used different regulatory mechanisms and had no patient data available. They used this model to make general statements about ideal time points for treatment. This model was followed by other models also analysing different treatment regimens (Afenya, 1996; Andersen and Mackey, 2001). These models were then followed by more models with different aims. One part of the models was aimed to elucidate possible mechanisms in AML development (Cucuianu and Precup, 2010; Getto et al., 2013; Jäkel et al., 2018; Liso et al., 2008; Rodriguez-Brenes et al., 2013), whereas others focused on regulatory mechanisms, features of leukaemic cells or clonal evolution (Crowell et al., 2016; Dingli et al., 2007; Jiao et al., 2018; Stiehl et al., 2014; Wang et al., 2017). There are basically two other models with a similar aim, as the model I developed: the description of the individual course of disease and a correlation of the individual parameters with the outcome. The first was developed by Stiehl et al. and it includes regulatory mechanisms by signalling molecules and a possible independence of leukaemic cells from these signals (Stiehl et al., 2018). This model was fitted to the time courses of 41 exemplary patients to analyse the heterogeneity in the model parameters and their link to the survival of the patients. Chemotherapy, however, is not taken into account, so that the therapy response is assumed to be the same for all patients. The other model is a very detailed model by Sarker et al. (Sarker et al., 2017) which considers the whole blood linage by modelling the cell mobility between the bone marrow and different blood compartments in the body as well as the interactions amongst mature blood cells and their progenitors. Regulatory mechanisms with signalling molecules and chemotherapy treatment are also included. Patient data, however, was not used for individual parameter estimation, but to validate the model by testing different criteria found in the data. To correlate the parameters with the outcome theoretical patients (in form of a parameter set) were used to draw conclusions. So, the crucial and unique feature of my AML model is that it is able to reproduce

the entire time course of leukaemic burden of individual patients from diagnosis until relapse, including the heterogeneity in treatment and therapy response.

3.1.4 Survival analysis

This model, fitted to the patient's time courses provided me with a pair of patient-specific parameters. But what can these parameters tell us about the patient's course of disease? How can I compare the time until relapse of different subgroups of the patients? To analyse the results I gained using the mechanistic model and thereby answer these questions I used a statistical method, the Kaplan-Meier (KM) estimator and the logrank-test, which is a common method to compare the outcome of different groups of patients. This is also called survival or "time to event" analysis. The KM estimator is used to compute the survival function, which gives the fraction of patients that survive until a certain time point. Survival functions always decrease in time. Here, however, I did not analyse the survival time, but the relapse-free survival. These KM-plots can be found in Figure 3c and d in the manuscript (Section 3.2). That means that the curve shows the fraction of patients, that did not die or experience a relapse depending on the time. The KM estimator also takes into account that you do not always know when a patient relapses or dies, respectively. There are events that make it impossible to know what happened to the patients, such as the end of a study or a patient moves to a different place and is not included in the monitoring any more. These events are handled as censoring events, as I do not know what happens afterwards. I assume that the probability of experiencing the target event (here death or relapse) stays the same after a censoring event. Often looking at two of these survival curves, here to compare different combinations of the fitted parameters, gives already a good notion whether one group has a better survival than the other. Additionally, to confirm this notion, there is the logrank-test, which tests how likely it is that there is actually no difference between the two curves (again giving a p-value as probability). Often in relation to the survival analysis the hazard ratio is computed to estimate the advantage of one group over the other. Therefore, a hazard function for each group is estimated, which gives the rate of how many events of interest can happen per time unit (also called hazard rate). The ratio of the rates of the two groups is the hazard ratio and it tells you how much higher the risk for someone in one group is to experience the event than it is for someone of the other group. Hence, a hazard ratio of 2 means that one group's risk is twice as high as the other group's risk. Other methods for the analysis of the relation of the fitted parameters and the time of relapse are the group-wise comparison using

box plots and the logistic regression, both already explained in detail earlier (Section 2.1.6 and Section 2.1.5). Now, that we talked about how risks of relapse can be estimated and what the fitted parameters of the model can tell us about the patient, for me the question arises, whether I can actually estimate the exact time of relapse for each single patient. To find out if this works, I used the individually fitted model and estimated from the resulting dynamic of the leukaemic burden a time of relapse for each patient.

3.1.5 Concordance correlation coefficient

But can I know how accurate my estimate is for the whole cohort? There is the so called Lin's concordance correlation coefficient, which measures the accordance of the estimated relapse time from the fitted model and the relapse time inferred from the data over all patients. If this coefficient is 1, then for all patients the relapse times from the model is equal to the relapse times from the data. The lower the coefficient is, the further the two relapse times are apart from each other.

3.1.6 Summary of results

Now, as all methodological background is clear I can start to explain the main findings of the second manuscript in this work. Therein, I developed a mathematical model that describes the dynamics of healthy and leukaemic cells in the bone marrow of an AML patient. The model consists only of four ordinary differential equations, and assumes that the chemotherapy has the same effect on healthy cells as it has on leukaemic cells and, thereby, not including a faster elimination of leukaemic cells just by treatment administration. Nonetheless, I could show, that despite these simplifying assumption the model is able to generally reproduce the reduction of the leukaemic burden during treatment (Figure 1 in the manuscript 3.2). To fit this model to the individual time courses of the patients, I varied two of the model parameters for each individual patient. The first parameter is the reproduction rate of the leukaemic cells. It describes how fast the leukaemic cells can reproduce themselves. The other parameter is the leukaemic activation rate, which describes how fast the leukaemic cells switch from a dormant, chemo-resistant state into an active chemo-sensitive state. I could show, that these model fits are able to mimic the individual time courses of most of the 275 patients included in this study. The model only failed for about 5% of the patients. I could show, that the two patient-specific parameters are related to the patient's time of relapse. Using

an appropriate threshold of the ratio of these two individual parameters the patients could be separated into two different groups. These two groups showed distinct relapse-free survival characteristics with a better separation than the currently used risk classification (Figure 3 in the Manuscript 3.2). The hazard ratio for the groups is more than 5, meaning that the chance of dying or experiencing a relapse is more than 5 times higher in one group than in the other. Therefore, it would be extremely helpful to reliably identify the two parameters early during the disease to derive improved risk stratification schemes for treatment improvement. The main shortcoming, however, is that when trying to predict the relapse time, based only on the data points of the first 9 months and neglecting all the points afterwards the concordance coefficient as measure of accuracy is with 0.37 much lower than 1 (perfect prediction). For more than 40% of the patients the predicted relapse time diverges more than half a year from the actual time. Still the findings in this manuscript suggest that the usage of a mathematical model can not only give insights into the patients-specific differences between treatment responses, but further harbour the potential to improve the assessment of the severity of a patient's individual course of disease.

3.2 Manuscript

Authors: H. Hoffmann, C. Thiede, I. Glauche, M. Bornhaeuser and I. Roeder
Title: Differential response to cytotoxic therapy explains treatment dynamics of acute myeloid leukaemia patients: insights from a mathematical modelling approach
Journal: Journal of Royal Society Interface 17
Year: 2020
DOI: 10.1098/rsif.2020.0091

This paper is not available as Open Access and cannot be included in this document.

Chapter 4

Relapse Prediction

4.1 Introduction

And now we come to the third and last part of this work[1]. In the previous chapter we saw, that a prediction of the relapse time based on the available data using the mechanistic model is not possible. What came to my mind then was a method, that was used in a wide range of different prediction tasks, in some cases clearly outperforming traditional methods: neural networks. The question that instantly came to my mind and defined further strategies for this work was: could the predictability of relapse for AML patients also profit from this approach and hence, improve the treatment of these patients?

4.1.1 Neural networks

But what are neural networks and how can they be used in this case of predicting the regrowth of leukaemic cells in a patient's bone marrow? The idea for neural networks came, as the name already suggests, from our own brain and its network of neurons. The idea is to build an artificial construct that has similar abilities as the human brain. In some points computers may already outperform humans without neural networks, for example when solving difficult mathematical operations in a fraction of a second or solving differential equations. But in other points, such as image processing the human brain is much better. You only need a fraction of a second to recognise a dear friend of yours from across the street, who is not

[1] This is my favourite part, as I developed the ideas on how to proceed here and I did what really inspired me. Additionally, this is the part where I learned most new things, which is for me the greatest motivation ever.

even looking in your direction. This is only possible because you had years and even decades to gather experiences in real life situations. Those experiences are stored in your brain and that way you can connect different experiences, which is really difficult for computers. But these neural networks enable computers to learn from examples they were exposed to and draw conclusions, similar to the way human brains work.

To understand how a neural network is working, I will first explain a bit about how the human brain is working. The brain is full of neurons, which are interconnected forming one huge network. When a stimulus comes in, such as the visual impression perceived by the eyes when reading a book, the neurons of the visual nerve are activated to forward this stimulus to the brain. The signal transmission between neurons happens mostly unidirectional, that means that they receive a signal from one or multiple cells and transmit this to one or more other cells. You can picture this as electrical pulses that go through these cells, coming in at the input (the dendrites) and being forwarded to the cells connected to the output (the axon) of the cell, as visualized in Figure 4.1. This is done in a way that only signals that surpass a certain threshold are transmitted. Therefore, the information is processed by the connectivity of the neurons as well as by the transformation of the signal. Signals that come in at the same time from different neurons are added up and only important signals are being transmitted. In our example with the image from the eyes while reading, that would mean that this signal is processed, so letters are interpreted as words. Additionally, all other objects that are in the current field of vision, such as your hand holding the book or the surroundings, are not perceived by the brain. They have no meaning while you are fully concentrated on understanding the written words.

And this is basically how neural networks work. Simple neural networks have an input, where the data (the time series data of leukaemic burden in my case) comes in, which is connected to a bunch of artificial neurons that are all connected to the input, but not to each other, as shown in Figure 4.2. This connection can be strong or weak, which is represented by multiplying the input value with a weight. Every input node is connected with every node in the following layer. In each artificial neuron the weighted signals from all inputs are summed up and modified by a so called activation function. The most frequently used activation function is the rectified linear unit (ReLU), which does what the neurons do and only transmits a signal if it surpasses a certain threshold. For the artificial neuron this threshold is "0". All nodes in this layer are then connected to the output, which is the desired result (here the decision whether a relapse will happen or not), again with weights applied to set the importance of the connection.

The purpose of the neural network is to learn how the time courses of patients look like that

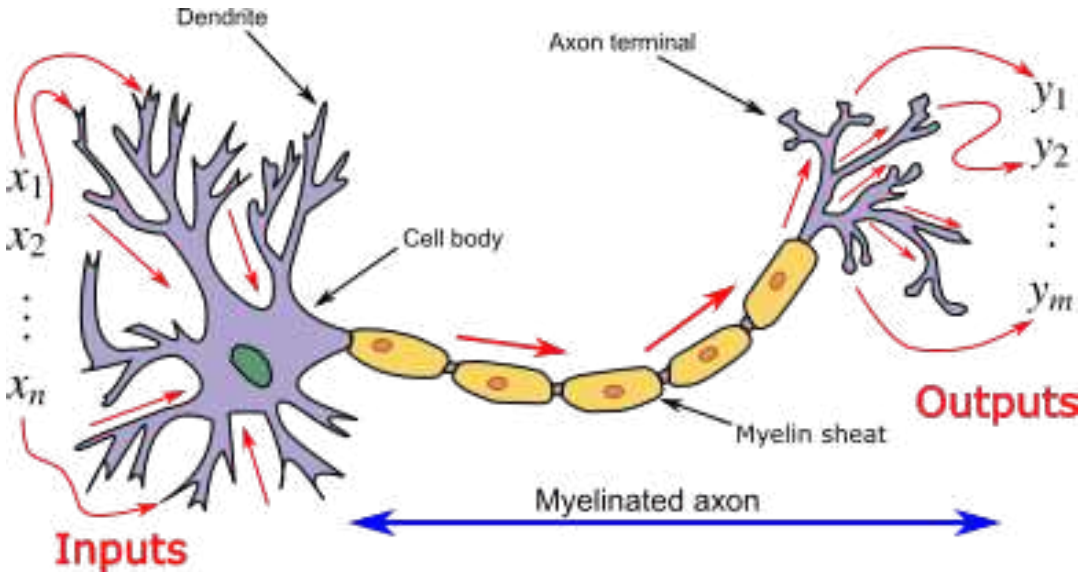

Figure 4.1: Neuron and myelinated axon, with signal flow from inputs at dendrites (x_1 to x_n) to outputs (y_1 to y_m) at axon terminals. By Egm4313.s12 at English Wikipedia, CC BY-SA 3.0 , via Wikimedia Commons

suffer a relapse compared to the time courses of patients that do not. So, we need to provide a data set of time courses for which we know whether a relapse occurred or not. This can be used for training the network. During training the neural network learns how to set the weights at all connections to be able to correctly predict the outcome. This process to find the best weights is called back-propagation. The idea behind it is that the weights are arbitrarily set at the beginning and for all training samples the output, 'relapse' or 'no relapse', is calculated. This output is then compared to the ground truth of this data. Therefore, it is essential for this type of training that we know the ground truth of training data (supervised learning). The distance between the truth and the predicted outcome (also called the loss) is calculated using the so-called loss-function. There are different kinds of loss-functions and it depends on the problem you want to solve, which loss-function you will use. For a problem, such as mine, where I want to classify the data into two different categories (relapse and no relapse) this loss function is called "binary cross-entropy ". Trying to reduce this loss the weights in the network are adjusted by going back through the network from the output to the input (the back-propagation). Calculating the loss and then adjusting the weights is repeated multiple times until the network is good enough or reached an optimum. Now, the networked is trained. When you input an unknown sequence it will predict the outcome based on what it has learned

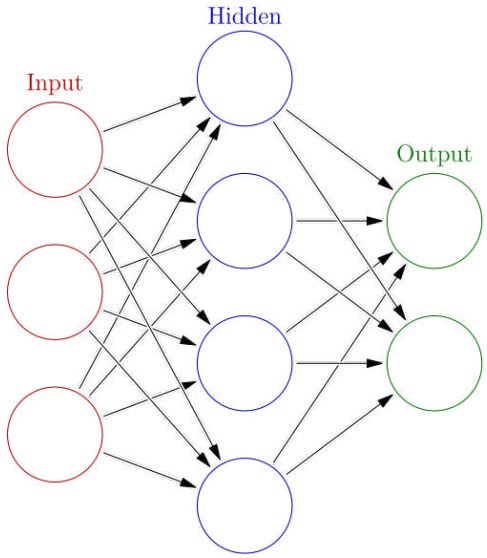

Figure 4.2: An artificial neural network is an interconnected group of nodes, inspired by a simplification of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another. This is a 3-layer neural network. Glosser.ca, CC BY-SA 3.0, via Wikimedia Commons

during the training.

But how can this network grab the meaning of time series, where all input values depend on the ones before that and maybe even on future values? This connection between the values is not easily learned by a simple neural network. Fortunately, a neural network architecture was developed that is able to remember things about the earlier values of the time series (Hochreiter and Schmidhuber, 1997). This architecture is called long-short-term memory (LSTM), as it has a memory to be able to find connection on short, but also on long distances within a sequence. I said earlier, that the neurons within one layer are not connected to each other, but only to the previous and the next layer (as in Figure 4.2). But for this kind of architecture, the neurons within a layer are interconnected and the information going out from one neuron goes also into the next neuron in the same layer. These kind of networks are called recurrent neural networks (Yu et al., 2019).

4.1.2 Data generation

But would that not need a huge amount of data for the network to learn all different facets of AML time courses? We already saw earlier that it is not trivial to detect differences in the first months of the time courses between the patients as the mechanistic model could not clearly find them. So, I needed a huge amount of data, which I simply did not have. But where to get this data from? It definitely is impossible to generate real clinical AML data on short notice. But I had another idea: I could use the mechanistic AML model, which I introduced in the previous chapter, to generate data. As the model is meant to mimic real patient data as closely as possible, it was basically made for this task. To set the model parameters as close as possible to the ones of real patients I sampled them from the parameters fitted to the patient data. Those I then varied slightly to generate a higher number of different parameter sets. Also the information of chemotherapies I could take from the patient cohort. So the only further information needed was the time points of the measurements. I could take those also from the real patients and that way generating data that was very close to real data. But I already showed, that this data has not much potential for good predictions. Therefore, I decided to generate several different data sets with more or less measurements, more or less noise, with and without detection limit or strategically set measurement time points. With these different data sets, I could analyse the influence of factors, such as the number and time points of measurements, measurement error and detection limit on the predictability of relapse, to find out what requirements exist to get good results in this prediction task. Maybe these findings could motivate later studies generating more suitable data for relapse predictions.
All of these generated data sets had data points in the first 9 months after start of therapy, including the information whether this virtual patient would experience a relapse within the following 15 months (2-year relapse) or not. This "ground truth" information is essential to verify how good the method works in predicting the 2-year relapse. So, I had an extensive simulated data set available. But I did not just train a neural network, get the results and say, well that is how it works. To be able to understand if using this neural network has an actual advantage (or disadvantage) compared to other conventional methods I wanted to compare different methods. So in the end I wanted to tell which method is the most suitable for the prediction task. But what additional methods come to mind?

4.1.3 Mechanistic models

First of all I wanted to test how good the mechanistic model works on the artificial data, to have a complex method with much background knowledge incorporated in my list of tested methods and simply because this was my actual starting point. But is it not trivial to use a model to predict the relapse that was used to generate the data in the first place? I know that the model perfectly describes the dynamics underlying the data. But this does not necessarily result in perfect predictions, but gives me a clear advantage of interpreting my results. Because, in that way I can analyse how explicitly the data quality impacts the predictability with the model, as the model is known to be correct (to describe the simulated data). So, the model should be able to perfectly predict the relapse if the data is flawless. But how does the reduction of measurements and the introduction of a measurement error influence the predictability with this method? This could be answered using the model that perfectly describes the underlying processes.
But how exactly does a prediction with the model work? To predict the 2-year relapse of a patient, the model is fitted to the data (with measurements of 9 months) to find the individual parameters for this patient. Then simulating a timespan of two years with the model, using these parameters and evaluating the proportion of leukaemic cells at the end of the simulation will give us the prediction. Is the proportion of leukaemic cells below the relapse threshold (of 1 %) no relapse is predicted, but if it is above this threshold this patient is predicted to experience a relapse within 2 years after therapy start.

4.1.4 Generalized linear models

So, now I have two methods, one is a model that knows exactly how the leukaemic burden of AML patients develop over time and another that knew nothing to begin with, but which learns from simply looking at the time courses. I wanted to include another method, that would lie in-between these methods concerning their need of prior knowledge about the data. But what does "in-between these methods" mean in this case? I thought that this means that I already interpret the data somehow and than give these handcrafted interpretations (or features) to a model that can learn from them the differences between relapser and non-relapser. An ideal and widely used method for this is a generalized linear model (GLM). When fitting this model to the data, it estimates, how much each of the features influences the outcome, similar to the weight estimation for the neural network. As our outcome is either "1" for relapse or "0" for no

relapse this model needs to end up with a number between 0 and 1 (for the tendency for one of the outcomes) and hence, we are using a logistic regression method, which was already introduced in Section 2.1.6. And what are these handcrafted features? These features are values that I constructed to describe the shape of the time series, such as the steepness of the slope of decrease of the leukaemic burden during a chemotherapy cycle or the lowest measured leukaemic burden. I could also use the earlier derived time course characteristics I introduced in section 2.1.2. So, after the training of the model, the estimated parameters tell me how likelier the patient has a relapse if it has a steeper decrease of the leukaemic burden during chemotherapy. Then, taking the features of new time series that are so far unknown to the model it can estimate the probability of each outcome, predicting the patient to have the outcome that is more likely. So, the accuracy of the model (and also of the other models) is the proportion of correctly predicted outcomes using a test data set, that was not used during model training.

4.1.5 Summary of results

Using the artificial data sets (generated with the mechanistic model for AML), I could confirm that the quality of the data has a major impact on the prediction accuracy. Especially, the sparseness of the data reduces the predictability of relapse. Comparing the three proposed methods I found, that the difference in performance between the methods was only minor. But how general are these results? Are they also true for other use-cases? To answer these questions, I teamed up with a colleague, to test all these methods also for another kind of leukaemia, the chronic myeloid leukaemia (CML). I will not go into details about the differences of the two diseases here, as this is not my focus. When comparing the predictability of relapse using four artificial data sets for CML patients we saw, that not the sparsity of the measurements, as in AML, but the measurement error has the largest interfering impact. Also deviating from the results for AML, in the case of CML we saw clear differences between the performance of the different methods. The GLM and the neural network were found to be more promising for predictions than the mechanistic model.
But in the end what I observed was, that the accuracies for data sets that were close to real AML patient data was not very high (ranging between 60 and 75%). Is it possible to improve this limited predictability by selecting the time points of measurement more carefully? To find an answer to this, I generated a data set with a more sophisticated measurement scheme that is oriented at the time points of the chemotherapy cycles. Using this to train the models for

relapse prediction I could see that the predictability is visibly increased for all three methods, suggesting to rethink the time points of measurements for AML patients.
In conclusion, we showed in this study that all three methods are suitable for relapse predictions, but depending on the data quality and the underlying dynamics of the disease sone methods might lead to better results than others. Further research should be conducted concerning the time points of measurement.

PLOS ONE

RESEARCH ARTICLE

How to predict relapse in leukemia using time series data: A comparative in silico study

Helene Hoffmann[1☯], Christoph Baldow[1☯], Thomas Zerjatke[1], Andrea Gottschalk[1], Sebastian Wagner[1], Elena Karg[1], Sebastian Niehaus[1,2], Ingo Roeder[1,3], Ingmar Glauche[1]*, Nico Scherf[1,4]*

1 Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, School of Medicine, TU Dresden, Dresden, Germany, **2** AICURA Medical GmbH, Berlin, Germany, **3** National Center of Tumor Diseases (NCT), Partner Site Dresden, Dresden, Germany, **4** Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

☯ These authors contributed equally to this work.
* nscherf@cbs.mpg.de (NS); ingmar.glauche@tu-dresden.de (IG)

Citation: Hoffmann H, Baldow C, Zerjatke T, Gottschalk A, Wagner S, Karg E, et al. (2021) How to predict relapse in leukemia using time series data: A comparative in silico study. PLoS ONE 16(11): e0256585. https://doi.org/10.1371/journal.pone.0256585

Editor: Mohamed A Yassin, Qatar University, QATAR

Received: May 21, 2021

Accepted: August 10, 2021

Published: November 15, 2021

Data Availability Statement: Patient data for AML patients was published in [23] and can be found at https://doi.org/10.6084/m9.figshare.12871777.v1 The CML patient data was published in [31]. Source code is available at https://zenodo.org/record/4293490#.X8DznMtKg-Q. The computational study presented in this manuscript is based on simulated data. Our simulation data and also the source code to generate such simulated datasets can be freely accessed at: https://zenodo.org/record/4293490#.X8DznMtKg-

Abstract

Risk stratification and treatment decisions for leukemia patients are regularly based on clinical markers determined at diagnosis, while measurements on system dynamics are often neglected. However, there is increasing evidence that linking quantitative time-course information to disease outcomes can improve the predictions for patient-specific treatment responses. We designed a synthetic experiment simulating response kinetics of 5,000 patients to compare different computational methods with respect to their ability to accurately predict relapse for chronic and acute myeloid leukemia treatment. Technically, we used clinical reference data to first fit a model and then generate de novo model simulations of individual patients' time courses for which we can systematically tune data quality (i.e. measurement error) and quantity (i.e. number of measurements). Based hereon, we compared the prediction accuracy of three different computational methods, namely mechanistic models, generalized linear models, and deep neural networks that have been fitted to the reference data. Reaching prediction accuracies between 60 and close to 100%, our results indicate that data quality has a higher impact on prediction accuracy than the specific choice of the particular method. We further show that adapted treatment and measurement schemes can considerably improve the prediction accuracy by 10 to 20%. Our proof-of-principle study highlights how computational methods and optimized data acquisition strategies can improve risk assessment and treatment of leukemia patients.

Introduction

Myeloid leukemias are characterized by aberrations affecting the proliferation and maturation of myeloid progenitor cells, leading to the progressive displacement of functional blood cells by immature and dysfunctional *leukemic* cells. Depending on the time scale of the displacement process, myeloid leukemias are further divided in chronic and acute leukemias.

Q or at our repository https://gitlab.com/imb-dev/aml-cml-relapse-prediction-method-comparison.

Funding: This work was supported by the Technische Universität Dresden, Faculty of Medicine Carl Gustav Carus, MeDDrive (Grant no. 60470) awarded to NS and by the Federal Ministry of Education and Research of Germany (BMBF), e:Bio Module III - Junior Research Group "MessAge" (Grant no. 031A315) awarded to IG. AICURA medical GmbH provided support in the form of salary for author SN. The specific roles of these authors are articulated in the 'author contributions' section. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have read the journal's policy and have the following competing interests: SN is a paid employee of AICURA medical GmbH. There are no patents, products in development or marketed products associated with this research to declare. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Patients with chronic myeloid leukemia (CML) typically carry a disease-specific chromosomal translocation forming the *BCR-ABL1* fusion gene [1–4]. Tyrosine kinase inhibitors (TKI) have been established as a targeted therapy leading to molecular remission in most patients under continuous drug administration [5]. Molecular monitoring of disease-specific *BCR-ABL1* mRNA in peripheral blood is the established strategy to quantify the leukemic burden under ongoing therapy. Current therapeutic challenges include the cessation of TKI treatment, upon which about 50% of CML patients develop a molecular recurrence and do not maintain treatment-free remission [6–8].

Acute myeloid leukemia (AML) is a highly heterogeneous disease with a variety of mutational profiles involved [9]. Commonly, a cyclic induction therapy with cytotoxic drugs such as cytarabine and anthracyclines aims to achieve sustainable remission, while a subsequent consolidation therapy supports the maintenance of the remission status. Molecular detection of mutated oncogenes or their transcripts is increasingly used to monitor leukemic burden in treated AML patients and can help to prospectively identify patients at the onset of disease recurrence [10, 11].

Disease recurrence after treatment-induced remission is a significant risk for all leukemia patients. Although the reappearance of CML after TKI cessation can be targeted well by restarting the treatment, physical and psychological side effects of retreatment can be minimized if a prospective identification of ineligible patients can be achieved. AML relapse usually occurs after completion of intensive chemotherapy treatment [12] and is associated with a poor prognosis [13]. In those case, the ability to prospectively predict the risk and timing of relapse or molecular recurrence is of highest importance to optimize and adjust the individual treatment strategy.

Currently, treatment decisions are based on the recommended risk stratification schemes. Those risk assessments are commonly based on *static* measurements from single time points, often at diagnosis [14, 15]. In contrast, treatment response dynamics, such as the speed of initial remission, are only rarely evaluated for risk stratification [16]. However, it was shown that molecular disease dynamics indeed correlate with therapy response and future relapse occurrence [17–22]. We reason that the direct integration of molecular response dynamics in the form of time-series data, which are increasingly available from standard disease monitoring, is a crucial element to improve the patient-specific risk stratification.

Assessing this question from a technical point of view, there are several, conceptually different approaches to integrate time-series data from molecular disease monitoring into an improved risk assessment. It is so far not clear how well these approaches are suited for time course data of hematological malignancies, and what their particular strengths and weaknesses are in this context. In order to address this question, we study three methods representing typical examples of the methodological spectrum:

- *Mechanistic models* (MM) describe the molecular disease dynamics as a functional consequence resulting from the interaction between relevant system components (such as cell types, drugs, cytokines etc.). They are commonly implemented as systems of ordinary differential equations (ODE) or as stochastic models. While some model parameters might be directly measurable, other model-specific parameters are obtained by optimally fitting the simulated time course to the available patient data. Evolving the model further in time allows to simulate the expected future behavior. Although MMs require considerable expert knowledge about the underlying mechanisms, the results of these models are readily interpretable as the model parameters typically carry explicit biological meaning.
- On the other end of the spectrum, deep learning approaches [23–25] use generic *neural network models* (NN) to adapt them on a training data set for which time-series data and the

corresponding future behavior is known. Roughly speaking, the NN implicitly identifies characteristic features within the time course data that correlate with future outcomes. Those methods require no *a priori* knowledge about the underlying mechanisms, but they are not suitable to directly interpret underlying biological mechanisms. Moreover, the training of NN requires a sufficient amount of annotated data.

- Classical statistical models like logistic regression classifiers can be used to correlate characteristic, predefined features of the time course data (such as speed of remission or remission level) with the known outcome. Such statistical models are summarized as *generalized linear models* (GLM) [26]. Herein, prior knowledge about general treatment dynamics is directly incorporated as an explicit feature of the GLM, while no understanding of the underlying biological mechanisms is required. Although GLMs are typically easier to interpret than neural networks (as the influence of parameters on the prediction can be assessed [27]) this probabilistic approach does not allow for explicit mechanistic interpretations as it is the case for MMs.

In this work, we systematically compare these three methods. In particular, we study the influence of data size, sampling density and measurement error on their prediction accuracy. As available data sets of relevant molecular time courses for AML and CML are currently limited, we first generate an artificial patient cohort (*synthetic data*) using different established mathematical models of those diseases [20, 28] (Fig 1). This artificial data set closely mimics the features of a smaller sample of real patient time courses, while the number of measurements and the particular noise level can be varied systematically and consistently. Based on this reference simulations, we are further able to suggest alternative disease surveillance schemes that may enhance the predictive power.

Materials and methods

Mechanistic models

To generate the synthetic data, we used two of our recently published mechanistic models for AML [20] and CML [28], both implemented as systems of ordinary differential equations (ODE). For the AML scenario, four ODEs are used to describe both leukemic and healthy stem cells. Two out of 11 model parameters are optimized to account for patient-specific differences in the disease characteristics, while the others were chosen to account for the general treatment dynamics. For the CML models, three ODEs represent active and inactive leukemic cells plus a population of interacting immune cells. In this case, we estimate 7 of 13 model parameters to optimally describe a patient's response. Details of the model setup are provided in the S6 and S7 Figs.

Patient data

For the generation of a set of realistic parameters, we fitted the respective mechanistic model to previously published time course data reflecting the patient's leukemia remission during and after therapy. In particular, we used the time courses of 275 NPM1-mut AML patients, in which the level of NPM1-mut/ABL abundance is used as a measure of leukemia load (median follow-up time of 10 months, the median number of 5 measurements [20]). Furthermore, we integrated data sets from 21 CML patients reflecting both their BCR-ABL1/ABL1 remission levels under TKI therapy and after therapy cessation (median follow-up time of 84 months, the median number of 28 measurements [28]). Examples of model fits to patient data, and the mean absolute error for each fitted patient can be found in S1 Fig.

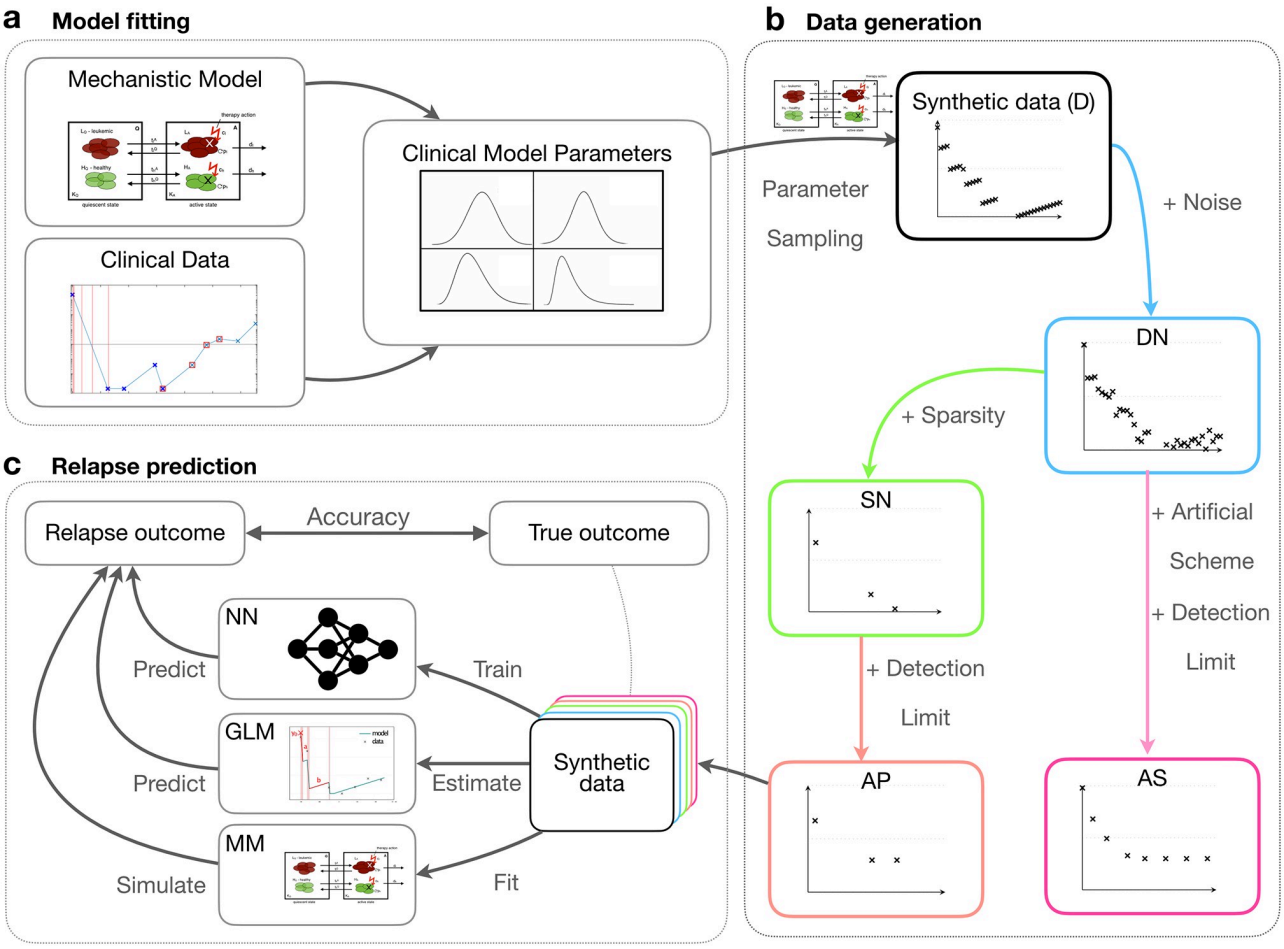

Fig 1. Conceptual overview of our methodological approach. (a) We developed mathematical models for both AML and CML from mechanistic and empirical knowledge [20, 28]. The models are first fitted to actual patient data to obtain realistic parameter distributions. (b) We sampled from these empirical parameter distributions to simulate dense, synthetic data (D). We gradually reduced the data quality to mimic actual clinical measurements by introducing noise (dense-noisy, DN), introduce sparsity (sparse-noisy, SN) and a minimum detection limit (artificial patient data, AP). Additionally, we introduced a more informative scheme (artificial scheme, AS), in which the temporal measurements are optimally spaced (AML) or a period of reduced treatment dose precedes therapy cessation (CML). (c) We systematically compared the performance of our mechanistic model (MM), a generalized linear model (GLM) and a neural network (NN) to predict the outcome (relapse/no relapse) of our virtual patient data with varying quality.

https://doi.org/10.1371/journal.pone.0256585.g001

Parameter fitting

Both, the AML and the CML model are initially fitted to the available patient data. Technically, we vary possible configurations of free parameters of the model such that the difference to the data is minimized (measured in terms of the sum of squares of the residuals on the logarithmic axis). While a simple optimization routine (*sequential quadratic programming*) is sufficient for the AML models, we apply a genetic algorithm combined with a gradient-based method for the CML scenarios which is better suited to avoid local minima. For further details we refer to the S1 File.

The same optimization routines are applied when the MMs are fitted to the artificial reference data for which we can tune data density and measurement noise (see below).

Generation of artificial data

To generate artificial patient data, we take random samples from the sets of parameters that were initially derived from fitting the mechanistic models to the available patient data.

In the case of AML, it was sufficient to randomly sample new parameter combinations from the set of empirically observed parameters plus adding a small, normally distributed variation to prevent the generation of identical duplets (see S1 File). For the treatment regime (namely the number and timing of induction cycles) we sampled one particular clinical chemotherapy schedule which we observed in the given patient data. Only artificial patients that reached remission (i.e. leukemic burden fell below the threshold of 1%) were included in the data sets. Using this parameterization and the corresponding schedules, we simulated artificial time courses of 24 months length. In analogy to the clinical situation, AML relapse is assigned if the fraction of leukemic increases above the threshold of 1% within 2 years after treatment start.

For the corresponding artificial CML time-courses, we sampled the seven model parameters from the distribution of empirical estimates in the available data basis under the condition that their mutual correlations are maintained (for details see S1 File). The time of therapy cessation was sampled based on kernel density estimates from the cessation time of the given patients (avg of 92 months with a standard deviation of 28.2 months). This information was then used in de novo forward simulations to generate artificial time-courses of varying duration until treatment stop plus 10 years thereafter. CML recurrence was defined as leukemia abundance $> 0.1\%$ (corresponding to BCR-ABL1/ABL1 = 0.1%, MR3).

In order to study how the data quality influences the prediction quality, we generated the following five reference data sets for both disease scenarios (examples in S2 and S3 Figs):

- Dense data (D): with weekly (AML) or monthly (CML) exact measurements, respectively.
- Dense-noisy data (DN): where white noise was added to each measurement, according to the noise level found in the given clinical patient data.
- Sparse-noisy data (SN): generated from the DN data set by reducing the number of data points to reflect the measurement frequency in clinical patient data.
- Artificial-Patient data (AP): by adding a detection limit to the SN data as found in the clinical patient data.
- Artificial scheme data (AS): Similar to AP data but using an improved sampling scheme compared to the clinical patient data. For AML measurements are made at the end of each chemotherapy cycle and every six weeks afterwards. For CML, the treatment dose is reduced to half of the usual dose 12 months before therapy cessation with frequent measurements during this period.

Using this synthetic reference data, we use the following setup to evaluate the correctness of predictions. For AML, all measurements from the initial treatment phase to 9 months after diagnosis are provided to the three methods and a corresponding relapse prediction within the subsequent 15 months is derived. For CML, we use all measurements up to the treatment stop to predict whether a patient will present with disease recurrence within ten years thereafter. The long timespan has been chosen to reflect the slow evolution of CML. To obtain the corresponding model predictions from the MM, we fitted the model parameters to the initial time course data (see above) and then simulated the future behavior using the fitted model parameters for each dataset individually. In contrast, both GLM and NN are optimized using a 10-fold cross validation on labelled data sets for which the respective outcome of relapse occurrence is provided.

Explicit features of time series for GLM analysis

As the Generalized Linear Model, we use a logistic regression classifier. The model uses explicit features that describe characteristics of the time-course data. We took the two characteristics

of AML time-courses defined in our previous work [19]: the elimination slope α, describing the speed of decrease of leukemic burden over the time of treatment and the lowest measured leukemic burden after treatment n. In this work, we further added three additional features obtained from a segmented regression approach: the leukemic burden at diagnosis (y_0, the following decreasing slope during the times of treatment (a) and the increasing slope of the leukemic burden in between treatment cycles (b) (S4A Fig).

For CML, we defined seven features from fits of a bi-exponential function that described the decrease of the leukemic burden after treatment start. These features include the bi-exponential parameters (A, α, B, β), the corresponding deviation of the fit and the data (σ), the cessation time and the BCR-ABL1 value before cessation or half dose. For the AS data, we expand these features with the behavior of the leukemic burden during the time of dose reduction including linear function parameter (γ), the deviation during half dose (C) and the last measured value before cessation (S4B Fig).

Neural network

NN were only trained on the raw time course data with no explicit features provided. To predict the occurrence of relapse, we used a bidirectional Long-short-term-memory (LSTM) network as a default architecture to handle sequence data with varying length. The model consists of a bidirectional LSTM layer followed by a fully connected feature extractor and a binary classification output. We use the respective cross-entropy loss to train the network. We implemented the network in Python using the Keras library [29]. To get a robust estimate of the model performance, we conducted 10 training runs on the same dataset and chose the network with the highest validation accuracy. We then did 10-fold cross-validation for the entire experiment to assess the average and the variability of the results. Further details about the network architecture and training can be found in the S1 File.

Accuracy

We use the traditional definition of accuracy as the ratio of the number of correct predictions over the total number of predictions: $acc = \frac{\#correct}{\#total} = \frac{TP+TN}{TP+FP+TN+FN}$ where TP, TN, FP, and FN are true positives, true negatives, false positives and false negatives respectively.

Results and discussion

Artificial patient data provide a suitable basis to systematically analyze the performance of predictive, computational models

We apply two mechanistic, mathematical models to simulate the dynamics of AML and CML [20, 28] thereby creating sets of artificial response data. To make sure that the artificial data resemble real patient time-courses as closely as possible, we fitted the models to respective data sets obtained from 275 AML patients carrying a traceable NPM1-mutation (consisting of a total of 1567 measurements quantifying the relative amount of NMP1-mut transcript [20] over time on a log10-scale) and 21 CML patients (with in total 478 measurements [28] quantifying the relative amount of BCR-ABL transcripts over time on a log10-scale). We report on the overall fitting quality in S1 Fig The fitted model parameters are used to simulate synthetic time courses (Fig 1a and 1b). To assess the influence of data quality, we gradually degraded the fully sampled, noise-free time series. We used estimates of the measurement frequencies and measurement errors obtained from the patient data to adjust the corresponding sampling density and noise level for the synthetic data (see S1 File). In total, we created four different datasets with 5000 time-courses from each model to systematically study the influence of data quantity

and quality: (i) a dense (D) data set consisting of weekly (AML) or monthly (CML) measurements of the leukemic burden free of any measurement error. (ii) For the dense-noisy (DN) data we added a normally distributed "technical" noise (see S1 File) to all data points of D to match the measuring error (AML) or the residuals observed between real data and their corresponding model fits (CML). (iii) In a third step, we reduced the total number of measurements per patient, creating a sparse-noisy (SN) data set that matches the measurement frequency in the real data. (iv) Finally, to make the data as realistic as possible, we also added a detection limit for very low measurements, called artificial patient (AP) data. Example time courses for all data sets can be found in S2 and S3 Figs.

To verify that the created artificial patient data (AP) sets are indeed similar to the real patient data, we derived characteristic features to quantitatively compare them. Those characteristic features refer to typical time scales and remission levels of the patient's response (see S4 Fig, Materials and methods). The features are computed separately for the AP data and the given patient data. The visual comparison in S5 Fig indicates that the median values of the characteristic features are very similar between AP and real data. It appears, that especially for the case of CML, the synthetic data sets yield a larger variance compared to the real data. A closer look at the data reveals that this is effect, at least partially, results from a sampling effect, as the variance measurement is only based on a small data set (n = 21) of real patients.

Data quality has a strong influence on prediction accuracies, but the drop in performance considerably differs between models and use-cases

Similar to the clinical presentation, we classified the synthetic time-courses as whether they show a relapse or not. For both CML and AML, we define disease recurrence by an increase of the leukemic burden (measured in terms of relative transcript abundance) within a predefined period above a given threshold (AML: leukemic burden increasing $> 1\%$ after treatment termination; CML: leukemic burden $> 0.1\%$ for at least one month).

We then systematically compared the accuracy of relapse predictions between the three general methods (namely MM, GLM, NN). To do so we provide each method with data from the initial treatment phase and compare the resulting predictions with the ground truth from the artificial data sets. For AML, we provide all measurements from the initial treatment phase until 9 months after diagnosis and derive a corresponding prediction on whether a relapse is expected within the subsequent 15 months. For CML, we use all measurements up to treatment stop to predict whether a patient will present with disease recurrence within ten years thereafter. We use the following strategy to derive predictions for the three methods: MM: fitting the mechanistic model to the initial treatment data only and further simulating the future time course, GLM: feeding the explicit features of the initial time-course (see Materials and methods) into a GLM classifier and NN: using an end-to-end learning approach with a neural network model applied to the initial time-course, which has been trained previously on an annotated reference data set (Fig 1c).

Next, we analyzed how well the different approaches (MM, GLM, NN) can predict the outcome for the artificial patient data and how model performance changes with varying data quality (Fig 1b and 1c, and Experimental Procedures). The results of the 10-fold cross-validation of the model performance are depicted in Fig 2. As expected, the prediction accuracy (see Materials and methods) declines for all approaches when the data quality decreases. We point out that the decrease in data quality differs between use-cases and models. In the case of AML, the introduction of sparsity leads to a relatively sharp drop in model performance. This drop illustrates the strong dependency on the number of measurements per time series: as in the given patient we only have a median of 4 measurements in the SN and AP data, compared to

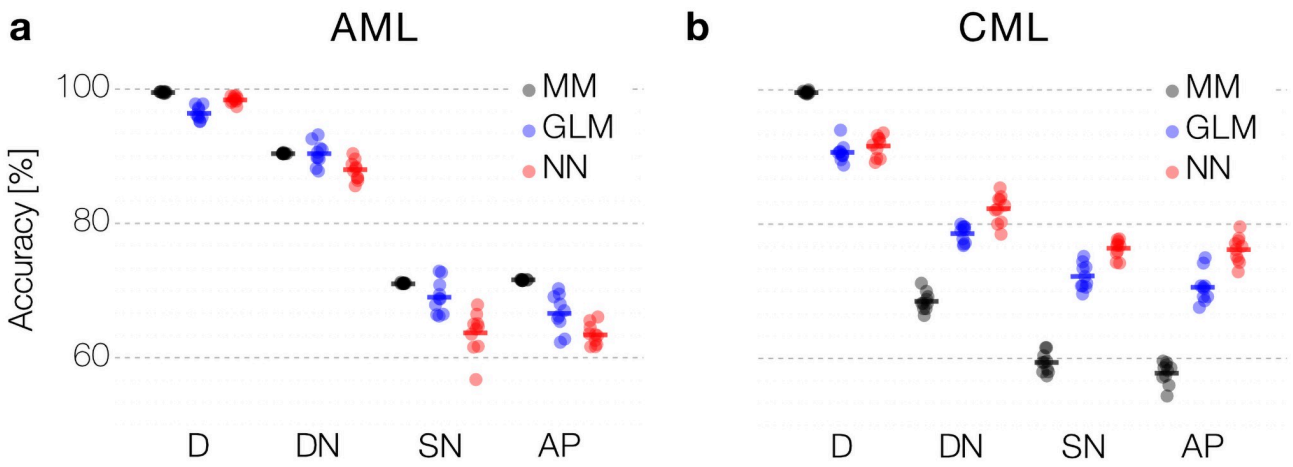

Fig 2. Prediction accuracy across data quality and computational models. (a, b) Comparison of performance between mechanistic model (MM), generalized linear model (GLM) and neural network (NN) to predict relapse in synthetic data for AML (a) and CML (b) using 10-fold cross-validation. Data quality gradually decreases from fully sampled, noise-free data (D), to noisy (DN), sparse and noisy (SN), and artificial patient data (AP) (see main text for details).

https://doi.org/10.1371/journal.pone.0256585.g002

39 weekly measurements in the dense data set (D) set. In line with this argument, we observe a more gradual decline in performance when comparing the effect of introducing noise and sparsity in the CML case. Here, we face a median of 25 measurements in the SN and AP data, compared to 93 monthly measurements in the dense data (D).

Interestingly, the difference in model performance is not consistent across the two use-cases. For the sparser AML data, all models perform similarly on the dense (D) and noisy data (DN). However, when introducing more sparsity into the data, a mechanistic model performs more robustly than the generic NN model (a difference in the accuracy of 6.3 and 7.4 percentage points for the SN and AP) and the GLM model performance is in between MM and NN. This result reflects the importance of introducing prior knowledge (or inductive bias) when dealing with very few data (Fig 2a).

We observe a different situation in the CML case. Here, the prediction accuracy for the mechanistic model drops down substantially more compared to the statistical GLM model and the generic NN when data quality decreases (a difference in accuracy between MM and NN of 19.7% for SN and 19.8% for AP, respectively). We recall that the noise-free data (D) was generated by the very same mechanistic model (compare Fig 2b). The high prediction accuracy for this data indicates that the correct (generative) MM can truly be identified. However, given the higher number of free parameters (n = 7) in the CML case, a reduction of data quality (either resulting from noisy or sparse measurements) more strongly affects the identifiability of the correct MM, while the GLM and the NN appear more robust.

Focusing on the artificial patient samples (AP), which best mimic the available patient data sets, the suggested models reach an accuracy of up to 70% (compare Fig 2a and 2b). These findings shows that predictive computational methods can indeed support risk assessment in myeloid leukemias based on nontrivial patterns in time series data obtained during treatment. However, the resulting prediction accuracy might not adhere to the expected standards for clinical decision support. Our systematic analysis shows how data characteristics, in particular the measurement schedule, effects the performance. Data scarcity and limited accuracy of available measurements per patient appears as a limiting factor for the overall prediction accuracy for relapse occurrence. Given those constrains on the data side, we are skeptical that structural changes to the computational methods (e.g. by refining the neural network architecture)

can substantially improve the overall performance. However, below we outline the potential in optimizing the measurement process to yield more informative sampling schemes.

Refined measurement and treatment schemes lead to improved prediction accuracies

We demonstrated that a significant limitation for the prediction accuracy results from the sparsity of the available data, in particular for the case of AML. Here, molecular diagnostics and especially bone marrow aspirates are limited resources in the clinical setting. As only increasing the sampling frequency is not an option in many cases, we wondered whether an optimized timing of the measurements could lead to better predictions while the overall number of measurements remains the same. To investigate this question, we created an additional set of artificial patients (AS) with consistent measurement intervals during the nine-month treatment period (i.e. the first day of each therapy cycle and every six weeks during the treatment-free phase). This typically results in 4 to 8 (median = 7) measurements per patient, which is only a moderate increase to the reported median of 5 measurements in the clinical sample. Fig 3a indicates that for this amended sampling regimen, we can already increase the accuracy of all prediction approaches (MM and NN by up to 12%, less pronounced for GLM). This finding strongly suggests that an adapted sampling scheme can considerably contribute to better relapse predictions, e.g. using methods from an optimal experimental design [30–33].

Owing to the establishment of regular BCR-ABL measurements in TKI-treated CML patients, available time courses are usually sufficient to monitor treatment response and remission status. It is still controversial, to which extend treatment free remission correlates with the observed time course of initial response [22, 28]. However, results from the DESTINY trial [34] suggest that dynamics of BCR-ABL increase during TKI dose reduction correlates with the remission status after treatment cessation [18]. The DESTINY trial differs from other TKI stop trials as patients in molecular remission reduced their TKI dose to 50% of the original dose for 12 months before TKI was finally stopped [34]. Motivated by this study, we simulated a corresponding data set in which a 12-month dose reduction is explicitly added to the model simulation (AS dataset). Training the prediction approaches to explicitly integrate this additional 12 month perturbation period, we found a substantial increase in the prediction accuracy of up to 19.1% (Fig 3b). We argue that probing the system's response to perturbation

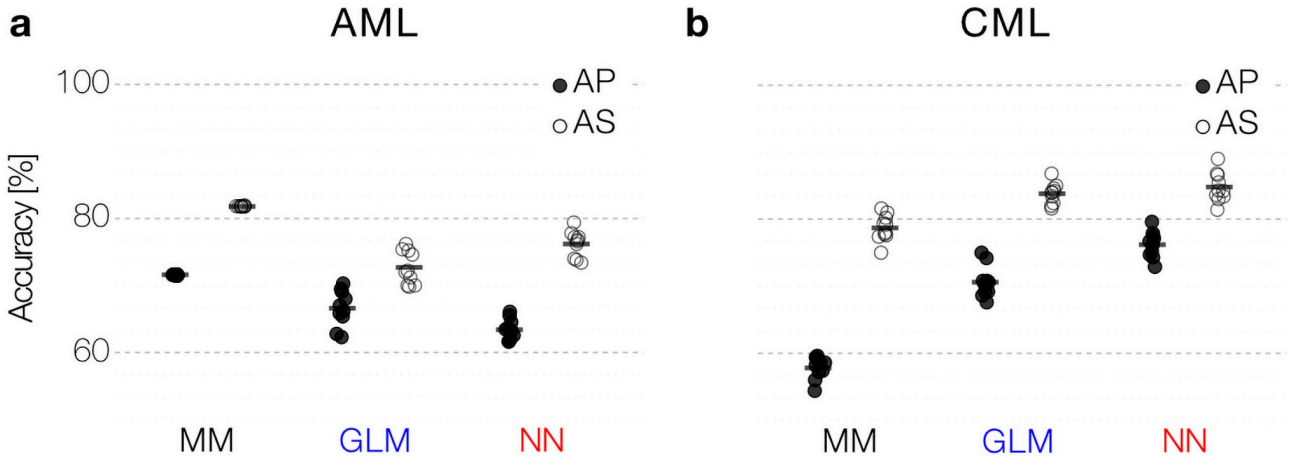

Fig 3. Dedicated measurement schemes. (a, b) A dedicated measurement scheme (AS) improves prediction performance with the same number of data points for all models compared to the AP data both for AML (a) and CML (b) data.

https://doi.org/10.1371/journal.pone.0256585.g003

(such as dose reduction) provides additional information about control mechanisms that cannot be obtained from ongoing monotherapy [16, 18, 28].

Our analysis demonstrates that optimized measurement schedules or systematic treatment alterations can substantially improve the accuracy of relapse predictions.

Conclusions

We showed that qualitatively different computational approaches, ranging from machine learning approaches to mechanistic models, are in principle suited to support relapse prediction based on time-series data of leukemia remission levels. To this end, we employed simulated time course data generated by mechanistic mathematical models, which we previously developed to describe disease and treatment dynamics in CML and AML. It is the advantage of this approach that we obtain highly controlled, although idealized, remission curves as a reference set from which we can abstract different levels of sampling density and measurement error. The simulated data allows us to refer to the *ground truth* of the underlying generative model. Using this artificial reference data, we could demonstrate that data quality in terms of measurement frequency and measurement error has a more substantial influence on the accuracy of the prediction than the employed prediction method, which is particularly evident in the AML data. Our results for the CML case indicate that fitting a more complex mechanistic model (in terms of the number of model parameters) to noisy data yields a greater uncertainty compared to a statistical predictor like a GLM or a NN.

Our analysis illustrates that generic methods, such as NN work well for the prediction of disease recurrence if frequent measurements are available (as in the CML data). For diseases with sparse measurements and limited data on the other hand (exemplified in the AML data), neural networks (and representation learning in general) is less suited for identifying the critical factors underlying the disease dynamics. In such cases, it is beneficial to incorporate prior knowledge to yield better predictions using either mechanistic models of the disease, if available, or statistical approaches based on explicit (phenomenological) features. In our current study, we used a long-short-term-memory (LSTM) NN as the standard approach for analyzing sequential data. An interesting next step is to assess if more complex neural network models [35, 36] can even improve the LSTM results, although we suspect that data quality is the major limiting factor.

Overfitting is a known problem of all machine learning approaches and applies to both the GLM and the NN method we presented. In order to minimize this risk, we applied a 10-fold cross validation which was also used to estimate the variation of the estimated accuracies. Our general approach is limited by the generation of time course data from generative models which intrinsically do not reflect "unexpected" behaviors. As long as the true data basis of clinical time courses is limited, only the additional consideration of alternative generative models could help to address this issue.

Regardless of the exact choice for a predictive computational method, our study indicates that the optimization of measurement schemes and clinical protocols is a promising strategy to improve the overall prediction accuracy without necessarily requiring more measurements per patients. In our predictions for AML recurrence, we could reach a level of accuracy of about 80% for the prognosis of relapse occurrence within two years after diagnosis. This result would already exceed the prediction accuracy for relapse-free survival after 12 months in the study by [15]. As our results are based on synthetic data which most likely does not reflect the full heterogeneity that could be seen in larger patient data sets, our comparison should be treated with caution and needs to be validated using independent clinical data obtained in a comparable context. Still, our findings indicate that standardized measurement schedules add

critical leverage to improve the ability for predicting relapse no matter what computational methods are used. Our artificial measurement schemes indicated a clear improvement, while we did not even apply formal optimization criteria to obtain most suitable regimes that maximizes accuracy while minimizing the number of measurements. This finding opens a clear perspective for future research on optimized measurement strategies that balance a maximized gain of information from clinical data with an economical use of resources. We argue that such refined schedules can contribute to reaching a level of prediction accuracy, which indeed supports clinical decision making.

In this work, we focused on the accuracy of relapse prediction employing three different, prototypic computational approaches working on time-series data. However, their implementation in a decision-making context also requires an intuitive understanding of how the method works. Although NN do not require any prior knowledge and can achieve excellent prediction accuracies, it is not trivial to identify which aspects of the data are causative for a particular prediction [37, 38]. In other words, the "black box" nature of NN does intrinsically not reveal the key features of the data on which a decision is based. There is a general, ongoing scientific discussion whether this intrinsic limitation of NN should prevent its application for particular questions, especially in health care [39, 40]. Currently, decision-makers and regulatory authorities hardly consider such methods for integration into clinical routines, although this might change in the future. Orthogonal developments in the field of "explainable AI" are currently pushing towards interpretability and the identification of causal relations between different system components [41–43]. As for now, MM represent the other side of the "interpretability spectrum" as they superimpose a principal understanding of the causal interactions onto the final observations. It appears tempting to favor this type of approach. However, it comes with other limitations: such models are highly specific and not easily transferable to other disease entities, and it cannot be guaranteed that all essential interactions are indeed mapped (compare [20]). The extent to which the non-representation of potential interactions effects the model predictions is hard to quantify and most likely highly disease specific. GLMs represent a middle ground and balance several aspects of NN and MM approaches. They can be helpful if detailed mechanistic knowledge is missing while important features of the response characteristics can readily be named, estimated and also interpreted. However, their overall performance depends strongly on the choice of those hand-crafted features and is also vulnerable to missing critical aspects.

The increasing availability of diagnostic methods to track molecular remission in different cancer types over extended time periods will establish a rich data source to explore further how this dynamic information can be correlated with the future course of treatment and disease [16]. Obtaining a systematic understanding of how different computational methods can be used to exploit this data is of crucial importance to provide usable predictions. Sufficient model validation within the particular domain is the prerequisite to integrate such computational models into decision making in a clinical context.

Supporting information

S1 Fig. Mechanistic model fit to patient data. (a) Example time-course of an AML patient (measured in terms of NPM1-mut abundance relative to reference gene ABL; blue dots) from start of chemotherapy at time point 0 until molecular relapse and the respective model fit (solid line; leukemic burden, rescaled by a factor 100 to match the clinical NPM1-mut/ABL ratios [20]). Red lines indicate time of chemotherapy administration. (b) Mean absolute error (MAE) for the fit of the mechanistic model to all 275 AML patients time-courses. (c) Example time-course of a CML patient (measured in terms of BCR-ABL/ABL abundance; black dots;

triangles indicate undetectable BCR-ABL levels with the corresponding detection threshold) from start of TKI treatment at time point 0 until disease recurrence after treatment stop (grey region) and respective model fit (solid line). (d) MAE of all 21 fitted CML patients.
(TIFF)

S2 Fig. Generation of the artificial AML data sets. We use a sample patient for which we obtain weekly and precise measurements, referred to as *dense* data (D). Adding a technical, normally distributed noise to each measurement on the log-scale, we obtain *dense-noisy* data (DN). *Sparse-noisy* data (SN) was generated from the DN data set, by reducing the number of data points to meet the measurement frequency in real patients. *Artificial patient* data (AP) is the data set most similar to the real patient data, which differs from the SN data set only by the inclusion of a detection limit (dashed red line), as it is found in the real data. *Artificial scheme* data (AS) is a data set, close to real data, with a measurement scheme, where measurements are made at the end of each chemotherapy cycle and every 6 weeks afterwards.
(TIFF)

S3 Fig. Overview of artificial CML data sets. *Dense* data (D) was simulated with monthly exact measurements. *Dense-noisy* data (DN) was obtained by adding normally distributed noise to each measurement. *Sparse-noisy* data (SN) was generated from the DN data set, by reducing the number of data points to meet the measurement frequency in real patients. *Artificial-Patient* data (AP) is the data set most similar to the real patient data, which differs from the SN data set only by the inclusion of a detection limit, as it is found in the real data. *Artificial scheme* data (AS) is a data set, close to real data, with an additional 12-month period of half-dose TKI treatment (shown in grey).
(TIFF)

S4 Fig. Derived features of the time-courses. (a) Features describing AML time courses: y_0 the leukemic burden at diagnosis, a the decreasing slope during treatment cycles, b the increasing slope in treatment free intervals (where y_0, a and b are obtained from a segmented regression approach), α the overall decreasing slope during treatment (shown as dashed line, separately fitted to the measurements) and n the minimal leukemic burden after treatment. (b) Features describing CML time courses: A, B and C being the intercepts of the straight lines fitted to the first and the second part of the bi-exponential approximation and to the increase of the leukemic burden during half-dose periods, respectively. α, β and γ are the respective slopes.
(TIFF)

S5 Fig. Similarity of artificial patients and real patients. A distribution comparison of statistical parameters: Comparison of distribution of parameters describing the course characteristics between artificial patient data (AP, blue) and real data (RD, black). (a) Parameters characterizing AML response: a—decreasing slope during chemotherapy cycle, b—increasing slope during treatment-free periods, y_0—initial burden on log scale, α—elimination slope, n—minimal measured leukemia burden after primary treatment. (b) Parameters characterizing CML response: the intercepts A and B (on a log scale) as well as the slope parameters α (on log scale) and β.
(TIFF)

S6 Fig. Schematic overview of the mathematical model for AML. Both, leukemic L and healthy H stem cells can reversibly change between two states (according to the rates t): the quiescent state Q with carrying capacity K_Q and the active state A with carrying capacity K_A. Cells in A undergo proliferation with rate p, differentiation with rate d and are subject to

chemotherapy with kill rate *c*.
(TIFF)

S7 Fig. Schematic overview of the mathematical model for CML. Leukemic stem cells (LSC) can reversibly change between two states X and Y (according to the rates p_{XY} and p_{YX}, respectively): X defines the quiescent, non-replicating cells, Y defines the active, proliferating cells. LSC in Y proliferate according to a logistic growth model with maximal proliferation rate p_Y and carrying capacity K_Y. The TKI-effect is described by a constant rate e_{TKI} affecting the leukemic cells in Y. Immune cells in Z are activated by cells in Y (immune recruitment), following an immune window approach (see Supporting information). At the same time the immune cells kill proportional target cell in Y. Immune cells in Z are generated with rate r_z and decay with rate a.
(TIFF)

S1 File. Supporting information.
(PDF)

Acknowledgments

We thank clinical partners for providing the patient data within the originals works [19, 20, 28] on which this simulation study is based.

Author Contributions

Conceptualization: Helene Hoffmann, Christoph Baldow, Ingmar Glauche, Nico Scherf.

Data curation: Helene Hoffmann, Christoph Baldow.

Formal analysis: Helene Hoffmann, Christoph Baldow, Thomas Zerjatke, Ingmar Glauche.

Funding acquisition: Ingo Roeder, Ingmar Glauche, Nico Scherf.

Investigation: Christoph Baldow, Thomas Zerjatke, Andrea Gottschalk, Elena Karg.

Methodology: Helene Hoffmann, Christoph Baldow, Sebastian Wagner, Sebastian Niehaus, Ingo Roeder, Ingmar Glauche, Nico Scherf.

Resources: Elena Karg, Ingo Roeder.

Software: Helene Hoffmann, Christoph Baldow, Sebastian Wagner.

Supervision: Ingo Roeder, Ingmar Glauche, Nico Scherf.

Visualization: Helene Hoffmann, Christoph Baldow, Nico Scherf.

Writing – original draft: Helene Hoffmann, Christoph Baldow, Sebastian Wagner, Sebastian Niehaus, Ingo Roeder, Ingmar Glauche, Nico Scherf.

Writing – review & editing: Helene Hoffmann, Christoph Baldow, Thomas Zerjatke, Andrea Gottschalk, Sebastian Wagner, Elena Karg, Sebastian Niehaus, Ingo Roeder, Ingmar Glauche, Nico Scherf.

References

1. Thijsen S, Schuurhuis G, van Oostveen J, Ossenkoppele G. Chronic myeloid leukemia from basics to bedside. Leukemia. Springer Nature; 1999; 13:1646–74.
2. Melo JV, Barnes DJ. Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nat Rev Cancer. Springer Science and Business Media LLC; 2007; 7:441–53. https://doi.org/10.1038/nrc2147 PMID: 17522713

3. Chereda B, Melo JV. Natural course and biology of CML. Ann Hematol. Springer Science and Business Media LLC; 2015; 94 Suppl 2:S107–21. https://doi.org/10.1007/s00277-015-2325-z PMID: 25814077

4. Zhou H, Xu R. Leukemia stem cells: the root of chronic myeloid leukemia. Protein Cell. 2015; 6:403–12. https://doi.org/10.1007/s13238-015-0143-7 PMID: 25749979

5. Hochhaus A, Larson RA, Guilhot F, Radich JP, Branford S, Hughes TP, et al. Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia. N Engl J Med. 2017; 376:917–27. https://doi.org/10.1056/NEJMoa1609324 PMID: 28273028

6. Cerveira N, Loureiro B, Bizarro S, Correia C, Torres L, Lisboa S, et al. Discontinuation of tyrosine kinase inhibitors in CML patients in real-world clinical practice at a single institution. BMC Cancer. 2018; 18:1245. https://doi.org/10.1186/s12885-018-5167-y PMID: 30541488

7. Mahon F-X, Réa D, Guilhot J, Guilhot F, Huguet F, Nicolini F, et al. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. Lancet Oncol. 2010; 11:1029–35. https://doi.org/10.1016/S1470-2045(10)70233-3 PMID: 20965785

8. Nagafuji K, Matsumura I, Shimose T, Kawaguchi T, Kuroda J, Nakamae H, et al. Cessation of nilotinib in patients with chronic myelogenous leukemia who have maintained deep molecular responses for 2 years: a multicenter phase 2 trial, stop nilotinib (NILSt). Int J Hematol. 2019; 110:675–82. https://doi.org/10.1007/s12185-019-02736-5 PMID: 31538327

9. Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368:2059–74. https://doi.org/10.1056/NEJMoa1301689 PMID: 23634996

10. Schuurhuis GJ, Heuser M, Freeman S, Béné M-C, Buccisano F, Cloos J, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2018; 131:1275–91. https://doi.org/10.1182/blood-2017-09-801498 PMID: 29330221

11. Shayegi N, Kramer M, Bornhäuser M, Schaich M, Schetelig J, Platzbecker U, et al. The level of residual disease based on mutant NPM1 is an independent prognostic factor for relapse and survival in AML. Blood. American Society of Hematology; 2013; 122:83–92. https://doi.org/10.1182/blood-2012-10-461749 PMID: 23656730

12. Oliva EN, Franek J, Patel D, Zaidi O, Nehme SA, Almeida AM. The Real-World Incidence of Relapse in Acute Myeloid Leukemia (AML): A Systematic Literature Review (SLR). Blood. American Society of Hematology; 2018; 132:5188–5188.

13. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N Engl J Med. New England Journal of Medicine (NEJM/MMS); 2015; 373:1136–52.

14. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017; 129:424–47. https://doi.org/10.1182/blood-2016-08-733196 PMID: 27895058

15. Othus M, Wood BL, Stirewalt DL, Estey EH, Petersdorf SH, Appelbaum FR, et al. Effect of measurable ('minimal') residual disease (MRD) information on prediction of relapse and survival in adult acute myeloid leukemia. Leukemia. 2016; 30:2080–3. https://doi.org/10.1038/leu.2016.120 PMID: 27133827

16. Roeder I, Glauche I. Overlooking the obvious? On the potential of treatment alterations to predict patient-specific therapy response. Exp Hematol [Internet]. Elsevier; 2020; https://doi.org/10.1016/j.exphem.2020.11.006 PMID: 33246016

17. Branford S, Yeung DT, Parker WT, Roberts ND, Purins L, Braley JA, et al. Prognosis for patients with CML and >10% BCR-ABL1 after 3 months of imatinib depends on the rate of BCR-ABL1 decline. Blood. American Society of Hematology; 2014; 124:511–8.

18. Gottschalk A, Glauche I, Cicconi S, Clarke RE, Roeder I. Molecular dynamics during reduction of TKI dose reliably identify molecular recurrence after treatment cessation in CML. Blood. 2020;

19. Hoffmann H, Thiede C, Glauche I, Kramer M, Röllig C, Ehninger G, et al. The prognostic potential of monitoring disease dynamics in NPM1-positive acute myeloid leukemia. Leukemia. 2019; 33:1531–4. https://doi.org/10.1038/s41375-018-0371-y PMID: 30675011

20. Hoffmann H, Thiede C, Glauche I, Bornhaeuser M, Roeder I. Differential response to cytotoxic therapy explains treatment dynamics of acute myeloid leukaemia patients: insights from a mathematical modelling approach. J R Soc Interface. 2020; 17:20200091. https://doi.org/10.1098/rsif.2020.0091 PMID: 32900301

21. Saussele S, Richter J, Guilhot J, Gruber FX, Hjorth-Hansen H, Almeida A, et al. Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. Lancet Oncol. 2018; 19:747–57. https://doi.org/10.1016/S1470-2045(18)30192-X PMID: 29735299

22. Shanmuganathan N, Pagani IS, Ross DM, Park S, Yong AS, Braley JA, et al. Early BCR-ABL1 kinetics are predictive of subsequent achievement of treatment-free remission in chronic myeloid leukemia. Blood [Internet]. 2020; http://dx.doi.org/10.1182/blood.2020005514

23. Zhang X. Time series analysis and prediction by neural networks. Optim Methods Softw. Taylor & Francis; 1994; 4:151–70.

24. Goodfellow I, Bengio Y, Courville A. Deep Learning. The MIT Press; 2016.

25. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller P-A. Deep learning for time series classification: a review [Internet]. arXiv [cs.LG]. 2018. http://arxiv.org/abs/1809.04356

26. McCullagh P, Nelder JA. Generalized Linear Models (Chapman & Hall/CRC Monographs on Statistics and Applied Probability). 2 edition. Chapman and Hall/CRC; 1989.

27. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems 30. Curran Associates, Inc.; 2017. p. 4765–74.

28. Hähnel T, Baldow C, Guilhot J, Guilhot F, Saussele S, Mustjoki S, et al. Model-Based Inference and Classification of Immunologic Control Mechanisms from TKI Cessation and Dose Reduction in Patients with CML. Cancer Res. 2020. p. 2394–406. https://doi.org/10.1158/0008-5472.CAN-19-2175 PMID: 32041835

29. Chollet F. keras [Internet]. Github; [cited 2020 Nov 10]. https://github.com/keras-team/keras

30. Chaloner K, Verdinelli I. Bayesian Experimental Design: A Review. Stat Sci. Institute of Mathematical Statistics; 1995; 10:273–304.

31. Goodwin. Dynamic System Identification: Experiment Design and Data Analysis. Academic Press; 1977.

32. Seeger MW. Bayesian Inference and Optimal Design for the Sparse Linear Model. J Mach Learn Res. 2008; 9:759–813.

33. Walter E, Pronzato L. Identification of Parametric Models: from Experimental Data. Springer London; 2010.

34. Clark RE, Polydoros F, Apperley JF, Milojkovic D, Rothwell K, Pocock C, et al. De-escalation of tyrosine kinase inhibitor therapy before complete treatment discontinuation in patients with chronic myeloid leukaemia (DESTINY): a non-randomised, phase 2 trial. Lancet Haematol. 2019; 6:e375–83. https://doi.org/10.1016/S2352-3026(19)30094-8 PMID: 31201085

35. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. Neural Ordinary Differential Equations [Internet]. arXiv [cs.LG]. 2018. http://arxiv.org/abs/1806.07366

36. De Brouwer E, Simm J, Arany A, Moreau Y. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series. arXiv [csLG] [Internet]. 2019; https://arxiv.org/abs/1905.12374

37. Arras L, Montavon G, Müller K-R, Samek W. Explaining Recurrent Neural Network Predictions in Sentiment Analysis [Internet]. arXiv [cs.CL]. 2017. http://arxiv.org/abs/1706.07206

38. Shrikumar A, Greenside P, Kundaje A. Learning Important Features Through Propagating Activation Differences [Internet]. arXiv [cs.CV]. 2017. http://arxiv.org/abs/1704.02685

39. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019; 25:24–9. https://doi.org/10.1038/s41591-018-0316-z PMID: 30617335

40. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019; 1:206–15.

41. Tjoa E, Guan C. A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI [Internet]. arXiv [cs.LG]. 2019. http://arxiv.org/abs/1907.07374

42. Pearl J, Mackenzie D. The book of why: the new science of cause and effect. Basic books; 2018.

43. Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. Towards Causal Representation Learning [Internet]. arXiv [cs.LG]. 2021. http://arxiv.org/abs/2102.11107

Chapter 5

Discussion

After reading all these finding a big question arises: what is the impact of these finding on the world, the patients, the doctors or the scientists? By looking at the time course data from AML patients I could show you that the dynamics of a disease harbour additional interesting insights beside just the measurements at fixed time points. So, I could show the relation between the decrease of the leukaemic burden during chemotherapy and the level of leukaemic cell numbers after therapy. This finding can be used from other scientists to conduct similar studies for other diseases or in other areas of research to generally change the strategy of looking at data. Data collection should be reconsidered to be more focused on capturing the dynamics of a process instead of concentrating on fixed measures. Ideally, my findings might have an influential role for this change in perspective, enabling further advances in all medical research fields and beyond.

More I could show, that a flexible view on the time point when a characteristic is measured can yield more meaningful results. This was shown with the level of leukaemic cell numbers after primary treatment, which is most meaningful, when evaluated 9 months after therapy start and not measured at a fixed time point for every patient. This can help to establish new schemes for patient monitoring in AML and, in the long run, improve the treatment of these patients. Although my work does not give concrete recommendation for the perfect measurement scheme it invites to investigate further.

And then I showed you that a mathematical model is able to describe the dynamics of AML at its place of origin, the bone marrow, in a patient-specific manner. But what is the advantage of using such a mechanistic model for the description of these time courses and why not using just a model that simply describes the ups and downs of the leukaemic burden, just like the

characteristics that I defined in Chapter 2? Well, a descriptive model is nice and useful, when you are interested in which of the components of the time course has the biggest impact on the outcome. Statistical methods such as correlation, as well as linear models such as a GLM can yield information about that. But a mechanistic model, that describes each single process in detail can yield additional information. It not only results in patient-specific descriptive parameters, but these values are usually associated with a biological meaning. So, the values of a parameters from the mechanistic model do not only tell you that the leukaemic burden goes down nicely during therapy, but also whether this is because of the slow growth of the leukaemic cells or because they are easily killed by the therapeutics. Therefore, we can get notions on what exactly is different on the cellular level between two patients, who seem to react in a similar way to therapy, but only one of them experiences a relapse. Using this model we can, however, not proof that there is really happening what we think. But such a model is a good start to get an idea on how each part of the system is interacting with the other parts. Hence, using mechanistic models your hypothesis can be tested theoretically, you can establish new ones and design experiments or clinical studies to verify these. Nonetheless, it is important to be clear that these models are not telling you exactly what is really happening within the system they are describing, but only giving interesting hints and indications. They are meant to be a simplified pictures of reality, as it would be impossible to describe all influencing factors in a model. It is not only not possible, but also not desired, as the goal of such a mechanistic model is to describe a clearly defined process (here the dynamics of leukaemic burden in AML patients) with a model that is as simple as possible. This simplification reduces the problem to the most important factors, making it easier to interpret. The fact, that the proposed model was able to reproduce most of the patients time courses leads to the conclusion that the factors that were included in the model (i.e. a quiescent state of the stem cells, equal chemotherapy effect on all cells, etc.) indeed are the most important ones in AML dynamics for the majority of patients. Still, there are patients whose time-courses could not be reproduced. For them further factors play an important role that are not included in the model. Therefore, it is always a trade-off between the simplicity and interpretability of a model and its transferability to a wide range of different patients. The decision that the model as proposed here is the best compromise for this case, was mostly based on the fact, that reducing the complexity resulted in bad fitting to the data (not shown here) and the majority of patients could be reproduced rather good, so I had no incentive to further inflate the model.
The resulting insights in a patient's individual disease characteristics from this model can contribute to better understanding why patients react to their chemotherapy as they do. This

ability of describing the course of disease of each patient in a very personalized manner exceeds the current practices of dividing the patients into groups that are thought to react similar to the same treatment. Not only for AML, but also for other diseases these models can convey an important step away from levelling down all patients to an average patient but seeing them as diverse and unique as they actually are. Personalized treatment will rely on mathematical descriptions of all aspects of a disease and their dynamic behaviour and entanglements.

The here shown insights into the mechanisms of AML help to estimate the severity of a patient's course of disease and obtain clues about the unique characteristics of a patient's disease. The estimated parameters obtained by fitting the model to a patient's individual course of disease, give hints about two important factors for disease progression: the aggressiveness of the disease and the sensitivity to chemotherapy. Using this information the treatment can be improved to better meet a patients individual needs. And not only the treatment can be more personalized using these methods, but also the prognosis of the further development of the disease. Both aspects are surely connected to each other and can not only help the patient to get better with personalised treatment but he or she also gets the chance to live on with more reliable statements about the further progress of the disease. Therefore, my hope is that it will be more and more established to integrate computational methods in medical research and clinical routines, to support doctors in their decision making. That does not mean that such models or other computer programs are able to replace doctors in the future, as the emotional contact with a well trained doctor plays a major role for the patient to build trust and has a positive impact on his/her outcome (Riedl and Schüßler, 2017; Shuaib et al., 2020). Further, I personally do not think that it is wise to hand over all decisions about the life of a patient to a machine, without complete understanding the decision-making process. I see the role of models in medicine as a supporting structure, to be able to handle all available data at once and to enable the doctor to focus more on the patient itself than on all the numbers in the patient record.

And what about the relapse prediction? The main finding here was, that there are different methods that are suitable for this task, but none of these is able to make reliable predictions on real patient data. For me the most important message from this part of my work is, that if we want to use the whole power computers can offer to improve the treatment of patients it is absolutely essential to generate high quality data. Using a simple example I could show, that the choice of measurement time points can hugely increase the accuracy of predicting a patient's relapse. In my example the measurement scheme was oriented on the time points

of therapy, resulting in different time points of measurement for every patient, as they are not treated at the exact same time points. This scheme makes the measured values more comparable between patients than they are using the same time point for everyone. That would be feasible to implement into clinical routine without increasing the cost by much. It would be an improvement to invest some time into finding an optimal schedule, where as few as possible measurements are taken with still generating a reliable prediction. Such an optimisation problem can also be solved using computational methods, such as reinforcement learning, where a neural network is trained to find the optimal schedule.

Still, there is much to be done in this field as my results also showed, that the noise has a large impact on the predictability of relapse, which can not be reduced by adapted measurement schemes, but by better measurement techniques. Furthermore, so far, the predictions only concentrated on the question whether a patient will relapse within the first two years after diagnosis. A more detailed prediction of the time-point of relapse was not possible with the given methods and data. To improve this I am sure it would be necessary to integrate multiple sources of data into one model, beyond the measurements of leukaemic burden and the chemotherapy times points. Other important indicators in AML are the blood values, as they show the downstream effects of the disease on the patients immune system. Also gene-expression data was shown to have predictive value in AML (Warnat-Herresthal et al., 2020). Also the emerging technique of liquid biopsy (Hocking et al., 2016), where the leukaemic burden is derived from DNA snippets of the leukaemic cells in the blood, could improve the data shortage, by easier sampling compared to bone marrow biopsy.

So, following this work, doctors will hopefully see that taking regular and optimally timed samples to assess the leukaemic burden of a patient can have a thrust in better judging the state of a patient. Hence, this might lead to following projects, where computational scientists collaborate closely with clinical doctors to underpin my findings with studies, developing and testing improved schedules.

Maybe the work I have done will play a part in reforming the handling of patients into a direction of treating a patient more following his or her individual health-characteristics than over his general diagnosis, leaving the path of treating a disease and starting to treat a patient.

Contribution to Publications

The prognostic potential of monitoring disease dynamics in NPM1-positive acute myeloid leukemia The following parts were mainly completed by me: Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Visualization
I received support for the following parts: Conceptualization, Data Curation, Writing - Review and Editing
Other team members contributed to: Resources, Supervision, Project administration, Funding acquisition

Differential response to cytotoxic therapy explains treatment dynamics of acute myeloid leukaemia patients: insights from a mathematical modelling approach The following parts were mainly completed by me: Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Visualization
I received support for the following parts: Conceptualization, Data Curation, Writing - Review and Editing
Other team members contributed to: Resources, Supervision, Project administration, Funding acquisition

How to predict relapse in leukemia using time series data: A comparative in silico study For this publication, the following is a list of authors' contributions as stated in the original publication:
Conceptualization: **Helene Hoffmann**, Christoph Baldow, Ingmar Glauche, Nico Scherf.
Data curation: **Helene Hoffmann**, Christoph Baldow.
Formal analysis: **Helene Hoffmann**, Christoph Baldow, Thomas Zerjatke, Ingmar Glauche.
Funding acquisition: Ingo Roeder, Ingmar Glauche, Nico Scherf.

Investigation: Christoph Baldow, Thomas Zerjatke, Andrea Gottschalk, Elena Karg. Methodology: **Helene Hoffmann**, Christoph Baldow, Sebastian Wagner, Sebastian.
Niehaus, Ingo Roeder, Ingmar Glauche, Nico Scherf.
Resources: Elena Karg, Ingo Roeder.
Software: **Helene Hoffmann**, Christoph Baldow, Sebastian Wagner.
Supervision: Ingo Roeder, Ingmar Glauche, Nico Scherf.
Visualization: **Helene Hoffmann**, Christoph Baldow, Nico Scherf.
Writing – original draft: **Helene Hoffmann**, Christoph Baldow, Sebastian Wagner, Sebastian Niehaus, Ingo Roeder, Ingmar Glauche, Nico Scherf.
Writing – review and editing: **Helene Hoffmann**, Christoph Baldow, Thomas Zerjatke, Andrea Gottschalk, Sebastian Wagner, Elena Karg, Sebastian Niehaus, Ingo Roeder, Ingmar Glauche, Nico Scherf.
I was mainly involved in the AML-related parts of this publication. The third publication was created in close collaboration with my colleague who took on the CML part.

Bibliography

Afenya, E. (1996). Acute leukemia and chemotherapy: A modeling viewpoint. *Mathematical Biosciences*, 138(2):79–100.

Alpermann, T., Schnittger, S., Eder, C., Dicker, F., Meggendorfer, M., Kern, W., Schmid, C., Aul, C., Staib, P., Wendtner, C. M., Schmitz, N., Haferlach, C., and Haferlach, T. (2016). Molecular subtypes of npm1 mutations have different clinical profiles, specific patterns of accompanying molecular mutations and varying outcomes in intermediate risk acute myeloid leukemia.

Andersen, L. K. and Mackey, M. C. (2001). Resonance in periodic chemotherapy: a case study of acute myelogenous leukemia. *Journal of theoretical biology*, 209(1):113–130.

Arai, F. and Suda, T. (2007). Maintenance of quiescent hematopoietic stem cells in the osteoblastic niche. *Annals of the New York Academy of Sciences*, 1106:41–53.

Arber, D. A. (2019). The 2016 WHO classification of acute myeloid leukemia: What the practicing clinician needs to know. *Seminars in Hematology*, 56(2):90–95.

Bachas, C., Schuurhuis, G. J., Assaraf, Y. G., Kwidama, Z. J., Kelder, A., Wouters, F., Snel, A. N., Kaspers, G. J. L., and Cloos, J. (2012). The role of minor subpopulations within the leukemic blast compartment of AML patients at initial diagnosis in the development of relapse. *Leukemia*, 26(6):1313–1320.

Balsat, M., Renneville, A., Thomas, X., de Botton, S., Caillot, D., Marceau, A., Lemasle, E., Marolleau, J.-P., Nibourel, O., Berthon, C., Raffoux, E., Pigneux, A., Rodriguez, C., Vey, N., Cayuela, J.-M., Hayette, S., Braun, T., Coudé, M. M., Terre, C., Celli-Lebras, K., Dombret, H., Preudhomme, C., and Boissel, N. (2017). Postinduction Minimal Residual Disease Predicts Outcome and Benefit From Allogeneic Stem Cell Transplantation in Acute Myeloid

Leukemia With NPM1 Mutation: A Study by the Acute Leukemia French Association Group. *Journal of Clinical Oncology*, 35(2):185–193.

Bereshchenko, O., Mancini, E., Moore, S., Bilbao, D., Månsson, R., Luc, S., Grover, A., Jacobsen, S. E. W., Bryder, D., and Nerlov, C. (2009). Hematopoietic Stem Cell Expansion Precedes the Generation of Committed Myeloid Leukemia-Initiating Cells in C/EBPα Mutant AML. *Cancer Cell*, 16(5):390–400.

Bertram, J. S. (2000). The molecular biology of cancer. *Molecular Aspects of Medicine*, 21(6):167–223.

Borer, R. A., Lehner, C. F., Eppenberger, H. M., and Nigg, E. A. (1989). Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell*, 56:379–390.

Brunetti, L., Gundry, M. C., Sorcini, D., Guzman, A. G., Huang, Y.-H., Ramabadran, R., Gionfriddo, I., Mezzasoma, F., Milano, F., Nabet, B., Buckley, D. L., Kornblau, S. M., Lin, C. Y., Sportoletti, P., Martelli, M. P., Falini, B., and Goodell, M. A. (2018). Mutant NPM1 Maintains the Leukemic State through HOX Expression. *Cancer Cell*, 34(3):499–512.e9.

Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine*, 368(22):2059–74.

Cheng, H., Hao, S., Liu, Y., Pang, Y., Ma, S., Dong, F., Xu, J., Zheng, G., Li, S., Yuan, W., and Cheng, T. (2015). Leukemic marrow infiltration reveals a novel role for Egr3 as a potent inhibitor of normal hematopoietic stem cell proliferation. *Blood*, 126(11):1302–1313.

Colombo, E., Marine, J.-C., Danovi, D., Falini, B., and Pelicci, P. G. (2002). Nucleophosmin regulates the stability and transcriptional activity of p53. *Nature cell biology*, 4:529–533.

Cooper, J. N. and Young, N. S. (2017). Clonality in context: hematopoietic clones in their marrow environment. *Blood*, 130(22):2363–2372.

Corces-Zimmerman, M. R., Hong, W.-J., Weissman, I. L., Medeiros, B. C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7):2548–53.

Cordell, J. L., Pulford, K. A., Bigerna, B., Roncador, G., Banham, A., Colombo, E., Pelicci, P. G., Mason, D. Y., and Falini, B. (1999). Detection of normal and chimeric nucleophosmin in human cells. *Blood*, 93:632–642.

Crowell, H. L., MacLean, A. L., and Stumpf, M. P. (2016). Feedback mechanisms control coexistence in a stem cell model of acute myeloid leukaemia. *Journal of Theoretical Biology*, pages 1–11.

Cucuianu, A. and Precup, R. (2010). A hypothetical-mathematical model of acute myeloid leukaemia pathogenesis. *Computational and Mathematical Methods in Medicine*, 11(1):49–65.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D., McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V. J., Cook, L., McGrath, S. D., Vickery, T. L., Wendl, M. C., Heath, S., Watson, M. A., Link, D. C., Tomasson, M. H., Shannon, W. D., Payton, J. E., Kulkarni, S., Westervelt, P., Walter, M. J., Graubert, T. A., Mardis, E. R., Wilson, R. K., and DiPersio, J. F. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing.

Dingli, D., Traulsen, A., and Michor, F. (2007). (A)symmetric stem cell replication and cancer. *PLoS Computational Biology*, 3(3):0482–0487.

Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Büchner, T., Dombret, H., Ebert, B. L., Fenaux, P., Larson, R. A., Levine, R. L., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G. J., Sanz, M., Sierra, J., Tallman, M. S., Tien, H.-F., Wei, A. H., Löwenberg, B., and Bloomfield, C. D. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*, 129(4):424–447.

Döhner, H., Estey, E. H., Amadori, S., Appelbaum, F. R., Büchner, T., Burnett, A. K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R. A., Lo-Coco, F., Naoe, T., Niederwieser, D., Ossenkoppele, G. J., Sanz, M. A., Sierra, J., Tallman, M. S., Löwenberg, B., and Bloomfield, C. D. (2010). Diagnosis and management of acute myeloid leukemia in adults: Recommendations from an international expert panel, on behalf of the European LeukemiaNet.

Falini, B., Bolli, N., Shan, J., Martelli, M. P., Liso, A., Pucciarini, A., Bigerna, B., Pasqualucci, L., Mannucci, R., Rosati, R., Gorello, P., Diverio, D., Roti, G., Tiacci, E., Cazzaniga, G.,

Biondi, A., Schnittger, S., Haferlach, T., Hiddemann, W., Martelli, M. F., Gu, W., Mecucci, C., and Nicoletti, I. (2006). Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. *Blood*, 107:4514–4523.

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., Bigerna, B., Pacini, R., Pucciarini, A., Liso, A., Vignetti, M., Fazi, P., Meani, N., Pettirossi, V., Saglio, G., Mandelli, F., Lo-Coco, F., Pelicci, P.-G., and Martelli, M. F. (2005). Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *The New England journal of medicine*, 352:254–266.

Fornari, F. A., Randolph, J. K., Yalowich, J. C., Ritke, M. K., and Gewirtz, D. A. (1994). Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells. *Molecular pharmacology*, 45(4):649–56.

Furth, J. J. and Cohen, S. S. (1968). Inhibition of mammalian DNA polymerase by the 5'-triphosphate of 1-β-d-arabinofuranosylcytosine and the 5'-triphosphate of 9-β-d-arabinofuranosyladenine. *Cancer research*, 28(10):2061–2067.

Gary Gilliland, D. and Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia.

Getto, P., Marciniak-Czochra, A., Nakata, Y., and dM Vivanco, M. (2013). Global dynamics of two-compartment models for cell production systems with regulatory mechanisms. *Mathematical Biosciences*, 245(2):258–268.

Gorello, P., Cazzaniga, G., Alberti, F., Dell'Oro, M. G., Gottardi, E., Specchia, G., Roti, G., Rosati, R., Martelli, M. F., Diverio, D., Coco, F. L., Biondi, A., Saglio, G., Mecucci, C., and Falini, B. (2006). Quantitative assessment of minimal residual disease in acute myeloid leukemia carrying nucleophosmin (NPM1) gene mutations. *Leukemia*, 20(6):1103–1108.

Graham, F. L. and Whitmore, G. F. (1970). Studies in mouse L-cells on the incorporation of 1-β-d-arabinofuranosylcytosine into DNA and on inhibition of DNA polymerase by 1-β-d-arabinofuranosylcytosine 5'-triphosphate. *Cancer research*, 30(11):2636–2644.

Gray, R. J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*, 16(3):1141–1154.

Grossmann, V., Schnittger, S., Kohlmann, A., Eder, C., Roller, A., Dicker, F., Schmid, C., Wendtner, C. M., Staib, P., Serve, H., Kreuzer, K. A., Kern, W., Haferlach, T., and Haferlach, C. (2012). A novel hierarchical prognostic model of AML solely based on molecular mutations. *Blood*, 120(15):2963–2972.

Guan, Y. and Hogge, D. E. (2000). Proliferative status of primitive hematopoietic progenitors from patients with acute myelogenous leukemia (AML). *Leukemia*, 14(12):2135–2141.

Haferlach, T., Kern, W., Schoch, C., Schnittger, S., Sauerland, M. C., Heinecke, A., Büchner, T., and Hiddemann, W. (2004). A new prognostic score for patients with acute myeloid leukemia based on cytogenetics and early blast clearance in trials of the German AML Cooperative Group. *Haematologica*, 89(4):408–418.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hocking, J., Mithraprabhu, S., Kalff, A., and Spencer, A. (2016). Liquid biopsies for liquid tumors: emerging potential of circulating free nucleic acid evaluation for the management of hematologic malignancies.

Hope, K. J., Jin, L., and Dick, J. E. (2004). Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nature immunology*, 5(7):738–43.

Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N., Fukata, M., Miyamoto, T., Lyons, B., Ohshima, K., Uchida, N., Taniguchi, S., Ohara, O., Akashi, K., Harada, M., and Shultz, L. D. (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nature Biotechnology*, 25(11):1315–1321.

Ivey, A., Hills, R. K., Simpson, M. A., Jovanovic, J. V., Gilkes, A., Grech, A., Patel, Y., Bhudia, N., Farah, H., Mason, J., Wall, K., Akiki, S., Griffiths, M., Solomon, E., McCaughan, F., Linch, D. C., Gale, R. E., Vyas, P., Freeman, S. D., Russell, N., Burnett, A. K., and Grimwade, D. (2016). Assessment of Minimal Residual Disease in Standard-Risk AML. *New England Journal of Medicine*, 0(0):160120140023007.

Jain, P., Kantarjian, H., Patel, K., Faderl, S., Garcia-Manero, G., Benjamini, O., Borthakur, G., Pemmaraju, N., Kadia, T., Daver, N., Nazha, A., Luthra, R., Pierce, S., Cortes, J., and

Ravandi, F. (2014). Mutated NPM1 in patients with acute myeloid leukemia in remission and relapse. *Leukemia and Lymphoma*, 55(6):1337–1344.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz, G., Correa, A., Banahan, B. F., Gabriel, S., Kathiresan, S., Stringham, H. M., Mccarthy, M. I., Boehnke, M., Tuomilehto, J., Haiman, C., Groop, L., Atzmon, G., Wilson, J. G., Neuberg, D., Altshuler, D., Ebert, B. L., and Chavez, A. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N Engl J Med December*, 25(37126):2488–2498.

Jäkel, F., Worm, O., Lange, S., and Mertelsmann, R. (2018). A stochastic model of myeloid cell lineages in hematopoiesis and pathway mutations in acute myeloid leukemia. *PLOS ONE*, 13(10):e0204393.

Jiao, J., Luo, M., and Wang, R. (2018). Feedback regulation in a stem cell model with acute myeloid leukaemia. *BMC Systems Biology*, 12(S4):43.

Kim, J. H. and Eidinoff, M. L. (1965). Action of 1-β-D-arabinofuranosylcytosine on the nucleic acid metabolism and viability of HeLa cells. *Cancer research*, 25(5 Part 1):698–702.

Krönke, J., Bullinger, L., Teleanu, V., Tschürtz, F., Gaidzik, V. I., Kühn, M. W. M., Rücker, F. G., Holzmann, K., Paschka, P., Kapp-Schwörer, S., Späth, D., Kindler, T., Schittenhelm, M., Krauter, J., Ganser, A., Göhring, G., Schlegelberger, B., Schlenk, R. F., Döhner, H., and Döhner, K. (2013). Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. *Blood*, 122:100–108.

Krönke, J., Schlenk, R. F., Jensen, K.-O., Tschürtz, F., Corbacioglu, A., Gaidzik, V. I., Paschka, P., Onken, S., Eiwen, K., Habdank, M., Späth, D., Lübbert, M., Wattad, M., Kindler, T., Salih, H. R., Held, G., Nachbaur, D., von Lilienfeld-Toal, M., Germing, U., Haase, D., Mergenthaler, H.-G., Krauter, J., Ganser, A., Göhring, G., Schlegelberger, B., Döhner, H., and Döhner, K. (2011). Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group. *J. Clin. Oncol.*, 29(19):2709–2716.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., and Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*, 367(6464):645–648.

Lichtman, M. A. (2013). A historical perspective on the development of the cytarabine (7days) and daunorubicin (3days) treatment regimen for acute myelogenous leukemia: 2013 the 40th anniversary of 7+3. *Blood Cells, Molecules, and Diseases*.

Liso, A., Castiglione, F., Cappuccio, A., Stracci, F., Schlenk, R. F., Amadori, S., Thiede, C., Schnittger, S., Valk, P. J. M., Döhner, K., Martelli, M. F., Schaich, M., Krauter, J., Ganser, A., Martelli, M. P., Bolli, N., Löwenberg, B., Haferlach, T., Ehninger, G., Mandelli, F., Döhner, H., Michor, F., and Falini, B. (2008). A one-mutation mathematical model can explain the age incidence of acute myeloid leukemia with mutated nucleophosmin (NPM1). *Haematologica*, 93(8):1219–1226.

Longo, D. L., Döhner, H., Weisdorf, D. J., and Bloomfield, C. D. (2015). Acute Myeloid Leukemia. *New England Journal of Medicine*, 373(12):1136–1152.

Minden, M. D., Till, J. E., and McCulloch, E. A. (1978). Proliferative state of blast cell progenitors in acute myeloblastic leukemia (AML). *Blood*, 52(3):592–600.

Pabst, T. and Mueller, B. U. (2009). Complexity of CEBPA dysregulation in human acute myeloid leukemia.

Passegué, E., Jamieson, C. H. M., Ailles, L. E., and Weissman, I. L. (2003). Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proceedings of the National Academy of Sciences of the United States of America*, 100 Suppl:11842–9.

Rauch, P. J., Ellegast, J. M., Widmer, C. C., Fritsch, K., Goede, J. S., Valk, P. J. M., Löwenberg, B., Takizawa, H., and Manz, M. G. (2016). MPL expression on AML blasts predicts peripheral blood neutropenia and thrombocytopenia. *Blood*, 128(18):2253–2257.

Riedl, D. and Schüßler, G. (2017). The influence of doctor-patient communication on health outcomes: A systematic review. *Zeitschrift fur Psychosomatische Medizin und Psychotherapie*, 63(2):131–150.

Rodriguez-Brenes, I. A., Wodarz, D., and Komarova, N. L. (2013). Stem cell control, oscillations, and tissue regeneration in spatial and non-spatial models. *Frontiers in oncology*, 3(April):82.

Rodriguez-Perales, S., Torres-Ruiz, R., Suela, J., Acquadro, F., Martin, M. C., Yebra, E., Ramirez, J. C., Alvarez, S., and Cigudosa, J. C. (2015). Truncated RUNX1 protein generated by a novel t(1;21)(p32;q22) chromosomal translocation impairs the proliferation and differentiation of human hematopoietic progenitors. *Oncogene*, 35(February):1–10.

Roeder, I., Horn, M., Glauche, I., Hochhaus, A., Mueller, M. C., and Loeffler, M. (2006). Dynamic modeling of imatinib-treated chronic myeloid leukemia: Functional insights and clinical implications. *Nature Medicine*, 12(10):1181–1184.

Rubinow, S. I. and Lebowitz, J. L. (1976a). A mathematical model of the acute myeloblastic leukemic state in man. *Biophysical Journal*, 16(8):897–910.

Rubinow, S. I. and Lebowitz, J. L. (1976b). A mathematical model of the chemotherapeutic treatment of acute myeloblastic leukemia. *Biophysical journal*, 16(11):1257–71.

Sarker, J. M., Pearce, S. M., Nelson, R. P., Kinzer-Ursem, T. L., Umulis, D. M., Rundell, A. E., and Rundell, A. E. (2017). An Integrative multi-lineage model of variation in leukopoiesis and acute myelogenous leukemia. *BMC systems biology*, 11(1):78.

Schuurhuis, G. J., Heuser, M., Freeman, S., Béné, M.-C., Buccisano, F., Cloos, J., Grimwade, D., Haferlach, T., Hills, R. K., Hourigan, C. S., Jorgensen, J. L., Kern, W., Lacombe, F., Maurillo, L., Preudhomme, C., van der Reijden, B. A., Thiede, C., Venditti, A., Vyas, P., Wood, B. L., Walter, R. B., Döhner, K., Roboz, G. J., and Ossenkoppele, G. J. (2018). Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood*, 131(12):1275–1291.

Shayegi, N., Kramer, M., Bornhäuser, M., Schaich, M., Schetelig, J., Platzbecker, U., Röllig, C., Heiderich, C., Landt, O., Ehninger, G., and Others (2013). The level of residual disease based on mutant NPM1 is an independent prognostic factor for relapse and survival in AML. *Blood*, 122(1):83–92.

Shuaib, A., Arian, H., and Shuaib, A. (2020). The increasing role of artificial intelligence in health care: Will robots replace doctors in the future? *International Journal of General Medicine*, 13:891–896.

Stiehl, T., Baran, N., Ho, A. D., and Marciniak-Czochra, A. (2014). Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation

patterns at diagnosis and relapse. *Journal of the Royal Society, Interface / the Royal Society*, 11(94):20140079.

Stiehl, T., Ho, A. D., and Marciniak-Czochra, A. (2018). Mathematical modeling of the impact of cytokine response of acute myeloid leukemia cells on patient prognosis. *Scientific Reports*, 8(1).

Suela, J., Alvarez, S., Cifuentes, F., Largo, C., Ferreira, B. I., Blesa, D., Ardanaz, M., García, R., Marquez, J. a., Odero, M. D., Calasanz, M. J., and Cigudosa, J. C. (2007). DNA profiling analysis of 100 consecutive de novo acute myeloid leukemia cases reveals patterns of genomic instability that affect all cytogenetic risk groups. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 21(6):1224–1231.

Sweeney, C. and Vyas, P. (2019). The Graft-Versus-Leukemia Effect in AML. *Frontiers in oncology*, 9.

Wang, W., Stiehl, T., Raffel, S., Hoang, V. T., Hoffmann, I., Poisa-Beiro, L., Saeed, B. R., Blume, R., Manta, L., Eckstein, V., Bochtler, T., Wuchter, P., Essers, M., Jauch, A., Trumpp, A., Marciniak-Czochra, A., Ho, A. D., and Lutz, C. (2017). Reduced hematopoietic stem cells frequency predicts outcome in acute myeloid leukemia. *Haematologica*, page haematol.2016.163584.

Warnat-Herresthal, S., Perrakis, K., Taschler, B., Becker, M., Baßler, K., Beyer, M., Günther, P., Schulte-Schrepping, J., Seep, L., Klee, K., Ulas, T., Haferlach, T., Mukherjee, S., and Schultze, J. L. (2020). Scalable Prediction of Acute Myeloid Leukemia Using High-Dimensional Machine Learning and Blood Transcriptomics. *iScience*, 23(1):100780.

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., Lu, C., Zhang, Q., Lin, L., O'Laughlin, M. D., McMichael, J. F., Delehaunty, K. D., Fulton, L. A., Magrini, V. J., McGrath, S. D., Demeter, R. T., Vickery, T. L., Hundal, J., Cook, L. L., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T. N., Walker, J. R., Watson, M. A., Heath, S. E., Shannon, W. D., Varghese, N., Nagarajan, R., Payton, J. E., Baty, J. D., Kulkarni, S., Klco, J. M., Tomasson, M. H., Westervelt, P., Walter, M. J., Graubert, T. A., Dipersio, J. F., Ding, L., Mardis, E. R., and Wilson, R. K. (2012). The origin and evolution of mutations in acute myeloid leukemia. *Cell*, 150:264–278.

Yanagisawa, B., Ghiaur, G., Smith, B. D., and Jones, R. J. (2016). Translating Leukemia Stem Cells into the Clinic: Harmonizing the Heterogeneity. *Experimental Hematology*.

Yoshimi, A., Toya, T., Kawazu, M., Ueno, T., Tsukamoto, A., Iizuka, H., Nakagawa, M., Nannya, Y., Arai, S., Harada, H., Usuki, K., Hayashi, Y., Ito, E., Kirito, K., Nakajima, H., Ichikawa, M., Mano, H., and Kurokawa, M. (2014). Recurrent CDC25C mutations drive malignant transformation in FPD/AML. *Nature communications*, 5:4770.

Young, D. C., Wagner, K., and Griffin, J. D. (1987). Constitutive expression of the granulocyte-macrophage colony-stimulating factor gene in acute myeloblastic leukemia. *The Journal of clinical investigation*, 79(1):100–6.

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, 31(7):1235–1270.

Zhang, N., Chen, Y., Lou, S., Shen, Y., and Deng, J. (2019). A six-gene-based prognostic model predicts complete remission and overall survival in childhood acute myeloid leukemia. *OncoTargets and Therapy*, 12:6591–6604.

Anlage 1

Technische Universität Dresden
Medizinische Fakultät Carl Gustav Carus
Promotionsordnung vom 24. Juli 2011

Erklärungen zur Eröffnung des Promotionsverfahrens

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten:
……
……
……
……
……
……
……...

3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

5. Die Inhalte dieser Dissertation wurden in folgender Form veröffentlicht:
……
……
……
6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.
……

7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät der Technischen Universität Dresden anerkenne.

8. Ich habe die Zitierrichtlinien für Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.

Ort, Datum H. Hoffmann

Unterschrift des Doktoranden
(Diese Erklärungen sind an das Ende der Arbeit einzubinden) Formblatt 1.2.1, Seite 1-1, erstellt 18.10.2013

Anlage 2

Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation

☒ das zustimmende Votum der Ethikkommission bei Klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetz betreffen
Aktenzeichen der zuständigen Ethikkommission

EK98032010

☐ die Einhaltung der Bestimmungen des Tierschutzgesetzes
Aktenzeichen der Genehmigungsbehörde zum Vorhaben/zur Mitwirkung

nicht zutreffend

☐ die Einhaltung des Gentechnikgesetzes
nicht zutreffend
Projektnummer .....

☒ die Einhaltung von Datenschutzbestimmungen der Medizinischen Fakultät und des Universitätsklinikums Carl Gustav Carus.

Dresden, den 02.02.2022

Unterschrift des Doktoranden