Doctorado en Física y Matemáticas
Universidad de Castilla-La Mancha

Modelos matemáticos en cáncer, hematopoyesis y análisis de datos en leucemia

Mathematical models in cancer, haematopoiesis and leukaemia data analysis

Salvador Chulián García

Tesis doctoral (Ph. D. Thesis) Diciembre 2020

Directores (Advisors): Víctor M. Pérez-García - María Rosa Durán

To my grandma, Manolita
A mi abuela, Manolita

Summary

In this thesis we explore mathematical methods and models as a tool to unravel the unknowns of tumour dynamics and relapse. We focus on leukaemia, a cancer of the white blood cells which develops in the bone marrow. Leukaemic tumours, together with brain cancer, are the leading cause of cancer death in children and young adults.

We review several mathematical models, based on ordinary differential equations and partial differential equations, describing the dynamics of cancer, specifically leukaemia, one of the most common cancers in childhood. We develop a model which describes the dynamics of healthy B lymphocytes, the white cells causing B acute lymphoblastic leukaemia, the most frequent paediatric leukaemia. We consider patients' flow cytometry data, not only to describe the development of B cells, but also to characterize biomarkers able to predict patients relapse factors. We analyse them with machine learning methods as well as topological ones, describing the differences between intensities of surface cell markers as well as shape features predicting relapse in leukaemia patients. Finally, we resort to the Lie classical method to obtain solutions of equations describing tumour dynamics in general.

This study has been the result of the collaboration efforts between clinicians, haematologists, immunologists and mathematicians. This multidisciplinary research aims to understand tumour features in order to improve risk stratification, and therefore the associated therapy protocols.

Resumen

En esta tesis se exploran métodos y modelos matemáticos como herramientas para descifrar interrogantes acerca de la dinámica tumoral y la recaída. Nos centramos en la leucemia, cáncer de las células blancas sanguíneas que se desarrolla en la médula ósea. Los tumores asociados a la leucemia, así como los tumores cerebrales, son la primera causa de fallecimiento por cáncer en niños y jóvenes adultos.

Se examinan múltiples modelos matemáticos, basados en ecuaciones diferenciales ordinarias o parciales, dada su capacidad para describir dinámicas tumorales, especialmente la de la leucemia, el cáncer más frecuente en la infancia. Se desarrolla un modelo que describe la dinámica de los linfocitos B sanos, células blancas causantes de la leucemia linfoblástica aguda B, una de las más comunes en la edad pediátrica. Se consideran datos de citometría de flujo de pacientes, no sólo para describir el desarrollo de las células B, sino también para caracterizar biomarcadores capaces de predecir factores asociados a la recaída en el cáncer. Para ello, se analizan las diferencias en la intensidad de marcadores de superficie celular, tanto con técnicas de "machine learning", como con métodos topológicos, para así describir características de forma que predigan la recaída en pacientes con leucemia. Para finalizar, se recurre al método clásico de Lie para obtener soluciones de ecuaciones que describen de manera general la dinámica tumoral.

Este estudio ha sido el resultado del esfuerzo colaborativo entre profesionales clínicos, hematólogos, inmunólogos y matemáticos. Esta investigación multidisciplinar tiene como objetivo el de entender las características tumorales, para así mejorar la clasificación de factores de riesgo y, por lo tanto, los protocolos terapéuticos asociados.

Acknowledgements

I would like to sincerely thank my two thesis supervisors, Víctor M. Pérez-García and María Rosa Durán, for the great pieces of advice, not only during the work process of this thesis, but also for their scientific and career example that I am proud to follow.

I am also grateful for all the collaborations and friends I have made through this work in the medical world. Both Drs. Cristina Blázquez and J. Francisco Gutiérrez (Kiko) from Jerez Hospital were the first to believe in our work together, and this thesis' results are as theirs as mine. I want to thank Drs. Lourdes Hermosín, Manuel Ramírez, Águeda Molinos, Ana Castillo and Teresa Caballero for their willingness to work with us and for all the biology discussions we had during this process.

I would like to thank all the mathematical collaborators from abroad and their teams for my time working with them: Helen Byrne, Bernadette J. Stolz from the University of Oxford, Anna Marciniak-Czochra and Thomas Stiehl from the University of Heidelberg, Rita Tracinà and Mariano Torrisi from the University of Catania and Masood Khalique from the North-West University. It has been my pleasure to discuss with you, learn from all of you and I am happy for all the future work that lies ahead thanks to you.

I cannot forget my colleagues from the Mathematical Oncology Laboratory: Ode, Jesús, Carmen, Juan J., David, Julián, Gabriel, Juan B. and María Jesús, who gave me a helping hand both scientifically and a break when needed. As for my colleagues at the University of Cádiz, I would like to specially thank Mariluz Gandarias, for her guidance and cheer, as well as my partner in crime Álvaro Martínez. You have been and will be the greatest coworkers during and after this experience. I want to share my happiness with all my coffee break

colleagues from the University of Cádiz: Adri, Bosco, David, Roberto, Rafa, Tamara, Antonio, Patri, Marta B., Gema and Marta S. Thanks for handling me.

My most sincere thanks go to my friends and family beyond the scientific world. To Capde, because life is brutal; to Marta, for our reconnection; to the Room of Requirements, you already know why; to Alba, Irene and Carmen, my constants; and to Cri and Ana Rosa, because our laughs are like medicine to me.

The last is for my most beloved ones: to my family, for their unconditional love. To Pili and Carlos, for their guidance and love as if I were your son. An special place is for my grandparents: Veli, for the love of learning; Salvador, for the importance of a good laugh and a good meal with you; Antonio, for the strength to keep us together; Manolita, because your warm love and strength will always be remembered. I cannot end without telling Paco (and of course your family) the special place you have in my heart. You make me happier every single day.

Finally, to my parents, the most amazing people you will ever find. You taught me love and how to handle life with a smile. To my father, because life put you on test, and you were and are as strong as steel. To my mother, because you are not only an example to your son, but to everybody who ever meets you. You are the kindest people I will ever know. I love you.

This work has been partially supported by Ministery of Science and Technology, Spain (grant number PID2019-110895RB-I00), the Spanish Foundation for Science and Technology (FECYT project PR214), the Asociación Pablo Ugarte (APU, Spain), Junta de Andalucía (Spain) group FQM-201, Junta de Comunidades de Castilla-La Mancha (grant number SBPLY/17/180501/000154), the Programme of Research and Transfer Promotion from the University of Cádiz (grant number EST2020-025, EST2019-169, MV2019-296, MV2018-503) and Health and Family department of Junta de Andalucía/FEDER (grant number ITI-0038-2019).

Agradecimientos

Quiero agradecer enormemente a mis dos directores, Víctor M. Pérez-García y María Rosa Durán, por sus maravillosos consejos, no solo durante el proceso de esta tesis, sino también por el ejemplo, tanto científico como académico, que me enorgullece seguir.

Agradezco a todos los colaboradores y amigos que hecho en el mundo médico durante este trabajo. Los Drs. Cristina Blázquez y J. Francisco Gutiérrez (Kiko) fueron los primeros en creer en nuestro trabajo juntos, y los resultados de esta tesis son tan suyos como míos. Quiero agradecer a los Drs. Lourdes Hermosín, Manuel Ramírez, Águeda Molinos, Ana Castillo y Teresa Caballero, sus ganas de trabajar con nosotros y por las discusiones biológicas que hemos tenido durante este proceso.

Querría agradecer a todos los colaboradores matemáticos del extranjero y a su equipo por el tiempo que he estado trabajando con ellos: Helen Byrne, Bernadette J. Stolz de la Universidad de Oxford, Anna Marciniak-Czochra y Thomas Stiehl de la Universidad de Heidelberg, y Rita Tracinà y Mariano Torrisi de la Universidad de Catania y Masood Khalique de la Universidad North-West. Ha sido un placer para mí tanto debatir como aprender con vosotros. Gracias por todo el futuro trabajo que tenemos por delante.

No puedo olvidar a mis compañeros del Mathematical Oncology Laboratory: Ode, Jesús, Carmen, Juan J., David, Julián, Gabriel, Juan B. y María Jesús, que me han ayudado científicamente y también me han ofrecido un buen descanso cuando se ha necesitado. En cuanto a mis compañeros de la Universidad de Cádiz, quiero dar las gracias especialmente a Mariluz Gandarias, por su guía y alegría, así como a mi cómplice Álvaro Martínez. Habéis sido (y seréis) los mejores compañeros durante y tras esta experiencia. Quiero compartir mi

alegría con mis compañeros de desayuno de la Universidad de Cádiz: Adri, Bosco, David, Roberto, Rafa, Tamara, Antonio, Patri, Marta B., Gema y Marta S. Gracias por aguantarme.

Mi más sincero agradecimiento es para mis amigos y familia más allá del mundo científico. A Capde, porque life is brutal; a Marta, por nuestra reconexión; a la Sala de los Menesteres, ya sabéis por qué; a Alba, Irene y Carmen, mis constantes; y a Cri y Ana Rosa, porque cuando nos reímos así curamos todos los males.

Para el final dejo a los que más quiero: a mi familia, por su amor incondicional. A Pili y a Carlos, por su gran guía y amor como si fuera su hijo. Un lugar especial es para mis abuelos: a Veli, por su amor por aprender; a Salvador, por la importancia de una buena risa y una buena comida contigo; a Antonio, por la fuerza por mantenernos unidos; a Manolita, por tu cálido amor y la fuerza por los que siempre te recordaremos. No puedo terminar sin decirte, Paco (y a tu familia, por supuesto), el gran lugar que tenéis en mi corazón. Si soy cada día más feliz, es gracias a ti.

Finalmente, a mis padres, que son las personas más maravillosas que encontraré jamás. Me habéis enseñado a querer y a llevar la vida con una sonrisa por delante. A mi padre, porque la vida te puso a prueba, y has sido y eres más fuerte que nadie. A mi madre, porque no eres solo un gran ejemplo para tu hijo, sino para cualquiera que te conoce. Sois las personas más buenas que hay. Os quiero.

Este trabajo ha sido financiado por el Minsterio de Ciencia y Tecnología de España (proyecto PID2019-110895RB-I00), la Fundación Española para la Ciencia y Tecnología (FECYT UCA PR214), la Asociación Pablo Ugarte (APU, España), el grupo FQM-201 de la Junta de Andalucía (España), Junta de Comunidades de Castilla-La Mancha (proyecto SBPLY/17/180501/000154), el programa de Investigación y Transferencia de la Universidad de Cádiz (referencias EST2020-025, EST2019-169, MV2019-296, MV2018-503) y el Departamento de Salud y Familia de la Junta de Andalucía/FEDER (proyecto ITI-0038-2019).

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Leukaemia is a cancerous disease in which blood cells display abnormal proliferation and invade other tissues. It is one of the biggest health issues globally. Almost half a million new leukaemia cases were diagnosed in 2018 [1].

Blood cancers affect the production and function of blood cells. They are the most common cancer types in children from birth to 14 years of age and account for around 3% of all cancers diagnosed in developed countries. Blood cancer survival in adults is about 50%. Although survival in children is higher and improving, blood cancer is still the major cause of cancer death in paediatric patients [2, 3, 4].

Most types of blood cancer start in the bone marrow, which is where blood is produced. In most blood cancers, the normal development process, starting from stem cells and leading to a hierarchy of more differentiated cells, is interrupted by the uncontrolled abnormal growth of specific types of blood cell.

There are three major types of blood cancer. Leukaemias are caused by the rapid production of abnormal white blood cells. Lymphomas are a type of blood cancer comprising abnormal lymphocytes, a type of white blood cells that fight infections. These cells multiply and collect in lymph nodes and other tissues and impair the lymphatic system's functionality to remove unnecessary fluids from the body and fight infections. Finally, myeloma is a cancer of the plasma cells, which produce disease- and infection-fighting antibodies.

It is well understood how blood cells differentiate from stem cells into more specialised cells (the so-called haematopoiesis process), as represented schematically in Fig. 1.1. At the top of the hierarchy governing normal haematopoiesis there are the haematopoietic stem cells (HSCs) [5, 6]. Pluripotent haematopoietic stem cells can give rise to either lymphoid or myeloid progenitors. Lymphoid progenitors can generate either lymphoblasts, which will become B

or T lymphocytes, or Natural Killer cells. These are all part of the specific immune system. Myeloid progenitors can also lead to a broad variety of cells, including erythrocytes, thrombocytes, or other cells of the non-specific immune system.

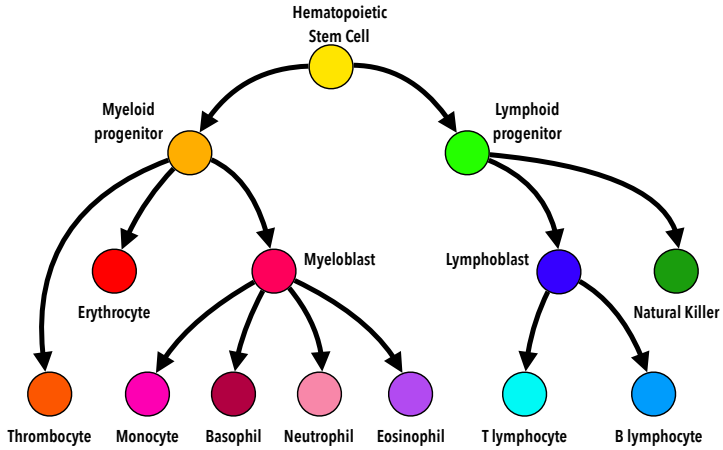


Figure 1.1: **Differentiation tree for blood cells**.

Although the classical understanding of haematopoiesis has considered cell types to be discrete compartments, current knowledge of the process [7] considers the evolution of cell types as a continuum process (see Fig. 1.2). This is because haematopoietic cells acquire lineage features through a continuous process involving the expression of different characteristic molecules [8].

In this framework, the type of cell that becomes cancerous determines the specific type of blood cancer. For instance, leukaemia can be either myeloid (or myelogenous), or lymphoid (or lymphoblastic, or lymphocytic). Also, leukaemias can be distinguished by the maturation stage of the transformed cells. Acute leukaemias affect blast cells (immature blood cells), and grow very fast. Chronic leukaemias cause an accumulation of mature cells, leading to slowly growing cancers. Thus, there are four different classes of leukaemias: Acute Lymphoblastic leukaemia (ALL), Chronic Lymphocytic leukaemia (CLL), Acute Myelogenous leukaemia (AML) and Chronic Myelogenous leukaemia (CML) [9]. However, it is not completely clear whether the hierarchical organisation is preserved in blood cancers. Myeloid leukaemias seem to be hierarchically

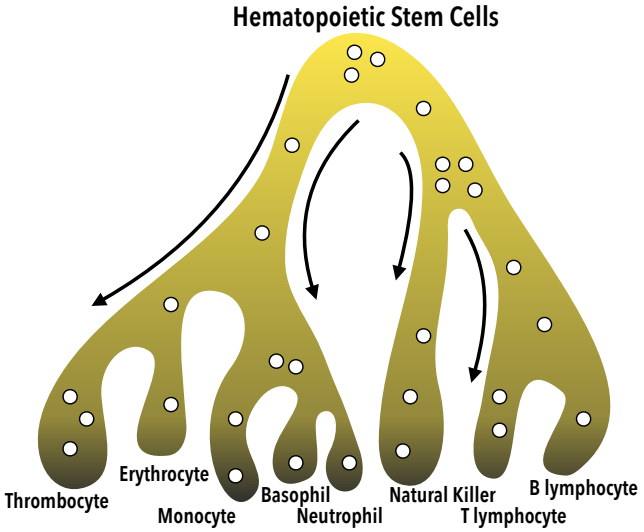


Figure 1.2: **Representation of continuum differentiation model**. Each dot represents a single cell and its location on a differentiation trajectory. Fig. adapted from Ref. [7].

organised, whereas acute lymphoblastic leukaemias are not [10, 11, 12].

The origins of the mutations are essential to understanding self-renewal and differentiation fractions for cancer cells [10, 13]. These probabilities could be explained by some of the basic hallmarks of cancer [14], such as sustaining proliferative signalling, resisting cell death, immortality, evading growth suppressors, and metastasis. The detection of these hallmarks is essential to tailoring treatments, which depend on classifying each patient within risk groups [15]. Currently, patients are assigned to a risk group depending on several factors, including the cell's morphology, the results of molecular or biochemical analysis, and the so-called flow cytometry techniques [16, 17]. This is done by taking samples of the bone marrow (where the haematopoiesis process occurs) which are then characterised in terms of immunophenotypic patterns [18], which can be standardised [19].

Mathematical modelling may offer a new perspective in Oncology, specifically in blood cancers, with a huge potential to develop new strategies to characterise

tumours and personalise treatments [20, 21, 22]. The cancer hallmarks relevant to each specific stage of development for each tumour type can be accounted for in mathematical models, usually as parameters to be estimated or as equations which model the dynamics of blood cancer development.

Blood cancers, and specifically leukaemias, have been one of the first types of cancers that has been thoroughly studied by applied mathematicians. It should be noted that there are many mathematically-grounded studies published in this field in high impact medical and general-science journals.

Leukaemias are a 'global' disease of the bone marrow, and as such spatial effects are usually ignored. They can be modelled mathematically, in an initial approach, using ordinary differential equations. More complex models have used partial differential equations, but to describe the evolution of some kind of trait or subpopulation, rather than spatial variables. Furthermore, blood cell counts are an easy way to gather information about the evolution of the disease. Putting the data together has led to substantial interest in the disease from modellers and clinicians managing the disease.

Only a small fraction of the data available during routine clinical procedures is used for diagnosis and incorporated into the models developed so far. In this thesis we will focus on the role of mathematical models based on differential equations and in flow cytometry data analysis. Mathematical techniques for (big-)data analysis [23] also have huge potential for providing answers to specific questions of relevance to leukaemia, whether alone or in combination with other mathematical methods. For instance, some studies have pointed out their potential use in avoiding expert manual gating of the data to identify leukaemic clones [24], analysing mass cytometry data [25, 26], or predicting treatment response [27].

Considering all of the exposed above, the main objectives of this thesis read as follows:

1. To review the current literature respecting leukaemia mathematical modelling.
2. To develop a mathematical model describing the healthy development of B lymphocytes, one of the main white blood cells causing leukaemia.
3. To find differences in the immunophenotypical distribution of flow cytometry data in patients recurring in leukaemia with respect to those who do not relapse.
4. To apply topological and machine learning methods in order to classify patients in terms of their relapse risk.

5. To obtain analytical solutions of tumour-related mathematical models via the Lie classical method.

Thus, this thesis aims to expand the current methods and mathematical approaches regarding leukaemia and tumour modelling. In Chapter 2 and Chapter 3 (Part I) we include mathematical models based on differential equations, while in Chapter 4 and Chapter 5 (Part II) we incorporate methods relying on leukaemia patients data. Moreover, in Chapter 6 (Part III) we consider a theoretical background for the obtention of solutions of differential equations describing tumour dynamics. Finally, in Chapter 7 and Chapter 8 (Part IV) we gather all conclusions and results exposed and include new approaches related to this thesis and considered as open problems.

Thus, this thesis is organised as follows:

Chapter 2 contains a review of mathematical models based on differential equations describing the growth and treatment of the most common leukaemias.

Chapter 3 develops a mathematical model which describes the normal behaviour of B lymphocytes, one of the common cells causing acute lymphoblastic leukaemia.

Chapter 4 finds a collection of biomarkers showing significant differences in expression levels between relapsed and non-relapsed leukaemia patients on diagnosis.

Chapter 5 anticipates the risk of relapse in paediatric ALL patients by combining methods from topological data analysis and machine learning.

Chapter 6 presents three applications of the Lie classical method to obtain solutions of differential equations relevant in terms of its applicability in cell dynamics and tumour invasion.

Chapter 7 and Chapter 8 summarise the most important conclusions of the research and analyses the open problems generated by this study.

Chapter 9 presents the publications related to this thesis.

Appendix A features additional material of stability conditions for the model presented in Chapter 3.

Appendix B consists of results related to the topological analysis performed in Chapter 5.

Appendix C includes additional results for the generators of the models presented in Chapter 6.

PART I

Mathematical models of haematopoiesis and leukaemia

CHAPTER 2

Mathematical modelling of leukaemia

This chapter is intended to expand the available literature on blood cancers [9, 28] to incorporate more studies and greater detail, by focusing on leukaemia. Our plan is as follows. Firstly, we summarise mathematical models based on differential equations describing the growth of myeloid leukaemias. This focus reflects the fact that myeloid leukaemias are the commonest among adults. The only models that exist for lymphoblastic leukaemia concern treatment. We then review mathematical models for different types of leukaemia treatment. Finally, we discuss the results and summarise our conclusions.

2.1 Mathematical models of myeloid leukaemias

Myeloid leukaemia arises from alterations of cells of the myeloid lineage, and is considered a clonal disorder of the haematopoietic stem cells (HSCs). The condition may lead to an increase in myeloid cell, erythroid cell or platelet counts, not only in peripheral blood but also in the bone marrow. As described above, the two general types are chronic myeloid leukaemia (CML) and acute myeloid leukaemia (AML), depending on the maturation stage of the cells. In CML cells mature during the chronic phase, while in AML blast cells fail to mature, generating large amounts of blasts, i.e. immature cells [29, 30].

2.1.1 Stem-cell based models of myeloid leukaemias

Stem-cell based models for myeloid leukaemia are based on mathematical models of the normal blood generation process, called haematopoiesis. The role of stem

cells in cancer was recently reviewed in [31] in terms of mathematical models which can characterise cell behaviour in normal cell development. For blood cells, an important haematopoiesis model was proposed by Marciniak et al [32]. The main assumption of this model was that the process of differentiation, i.e., the ability of a cell to change from one type to another, was described in several discrete maturation stages, beginning with stem cells as the first stage of maturation.

As cells mature, their proliferation rate increases, while the self-renewal fraction lowers, where self-renewal was understood as the probability of having the same properties and fates as their parent cell. This process is summarised in Figure 2.1. The model includes different cell subpopulations with n different maturation stages and feedback signalling to regulate haematopoiesis.

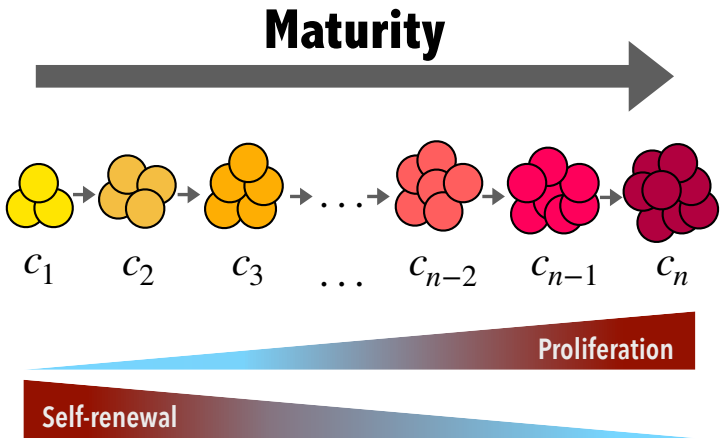


Figure 2.1: **Schematic representation of the assumptions behind the model** (2.1). Cells were grouped into n different maturation stages, c_i with $i = 1, ..., n$. As cells mature, and the j index increases, their proliferation rates p_j^c increase, whereas the self-renewal fractions a_j^c decrease.

The mathematical model describing the dynamics comprises a set of ODEs

for the several compartments of normal cells (c_i)

$$\frac{d}{dt}c_1(t) = (2a^c_{1,\max}s(t) - 1)p^c_1 c_1(t) - d^c_1 c_1(t), \tag{2.1a}$$

$$\frac{d}{dt}c_i(t) = 2(1 - a^c_{i-1,\max}s(t))p^c_{i-1}c_{i-1}(t) + \tag{2.1b}$$
$$+ (2a^c_{i,\max}s(t) - 1)p^c_i c_i(t) - d^c_i c_i(t),$$

$$\frac{d}{dt}c_n(t) = 2(1 - a^c_{n-1,\max}s(t))p^c_{n-1}c_{n-1}(t) - d^c_n c_n(t), \tag{2.1c}$$

and another set of ODEs for the leukaemic cells (l_j)

$$\frac{d}{dt}l_1(t) = (2a^l_{1,\max}s(t) - 1)p^l_1 l_1(t) - d^l_1 l_1(t), \tag{2.1d}$$

$$\frac{d}{dt}l_j(t) = 2(1 - a^l_{j-1,\max}s(t))p^l_{j-1}l_{j-1}(t) \tag{2.1e}$$
$$+ (2a^l_{j,\max}s(t) - 1)p^l_j l_j(t) - d^l_j l_j(t),$$

$$\frac{d}{dt}l_m(t) = 2(1 - a^l_{m-1,\max}s(t))p^l_{m-1}l_{m-1}(t) - d^l_m l_m(t), \tag{2.1f}$$

where $c_i = c_i(t)$ denotes the density (or number) of healthy cells in each maturation stage $i = 1, .., n$, p^c_i are the proliferation rates of healthy haematopoietic cells in mitosis, $a^c_{i,\max}$ are the self-renewal fractions, and d^c_i the death rates for every cell maturation stage. The notation is analogous for the leukaemic cells, $l_j = l_j(t)$ for $j = 1, ..., m$, and the constants p^l_j, $a^l_{j,\max}$ and d^l_j for $j = 1, ..., m$.

Feedback signalling was described in that study using the cytokine effect function $s(t)$. Cytokines are small proteins which assist in regulating fraction chemical signalling in cells. Cytokine concentration is modelled by the equation

$$s(t) = \frac{1}{1 + k^c\, c_n(t) + k^l\, l_m(t)}, \tag{2.1g}$$

where k^c and k^l are the signalling regulation strength, for both normal and leukaemic cells, respectively. These parameters are sensitive to the number of mature healthy and leukaemic cells, $c_n(t)$ and $l_n(t)$. This signalling was assumed to control the dynamics of cell proliferation and differentiation in the mathematical model. Figure 2.2 shows an example of evolution towards the homoeostatic equilibrium of the healthy haematopoietic cell compartments for $n = 6$.

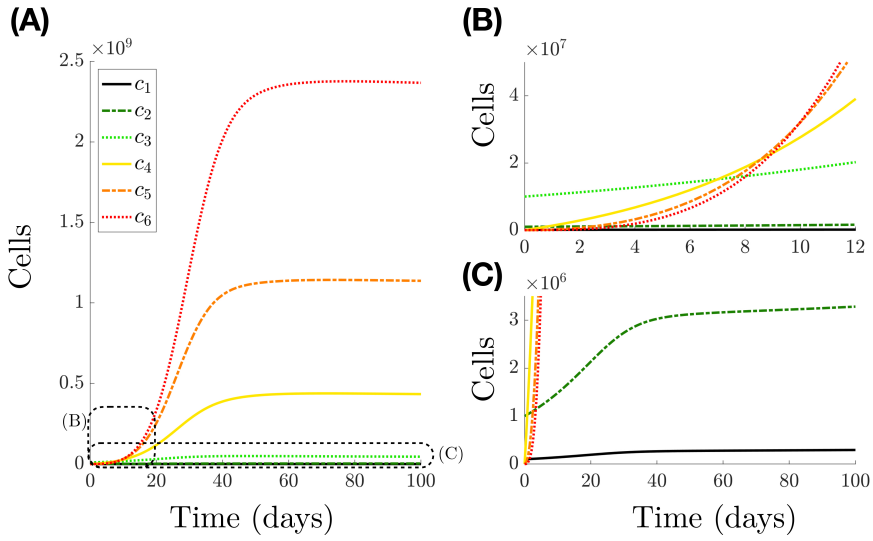


Figure 2.2: **Simulations of the healthy cell dynamics in model** (2.1). Simulations of the evolution of a set $n = 6$ compartments accounting for six maturation stages according to the model (2.1). The insets A, B show more the details of the dynamics of the same simulation. Following [32], the parameter values were $a_1 = 0.0865$, $a_2 = 0.1155$, $a_3 = 0.1735$, $a_4 = 0.3465$ and $a_5 = 0.693$ for the self-renewal fractions. For the proliferation rates, $p_1 = 0.7$, $p_2 = p_3 = p_4 = 0.65$ and $p_5 = 0.55$ were considered. The death rate was $d = 0.3$ and signal strength $k^c = 1.6 \cdot 10^{-10}$ cells^{-1}. Cell initial values were $c_1(0) = 10^5$, $c_2(0) = 10^6$ and $c_3(0) = 10^7$ and null for the other initial values.

The model of Eqs. (2.1) in [33] was built on the basis of the haematopoiesis model of [32]. The main conclusion of the mathematical study of [33] was that both self-renewal fractions and proliferation rates could be indicators of poor prognosis. Similar models were also studied in [34], where some mathematical properties, including linear stability analysis, and necessary and sufficient conditions for the expansion of malignant cell clones, were studied for related models.

A similar model by the same group [35] described the differentiation process as a two-stage process, but considered instead the multi-clonal nature of

leukaemia, the feedback processes and the role of treatment. The study performed numerical simulations for 'in-silico' virtual patients, and obtained estimated parameters from the tumour growth data of two real patients. The researchers concluded that self-renewal might be a key mechanism in the clonal selection process. It was also stressed that late relapses could originate from clones that were already present at diagnosis, a question that has been the subject of discussion in the biomedical research literature. Stem cell self-renewal has been reviewed in terms of their impact on the dynamics of cell populations in [36], concluding that a high self-renewal fraction can lead to faster cancer growth.

A similar model, [37], accounted for genetic instability through the inclusion of the possibility of mutations, an essential hallmark in cancer evolutionary dynamics. Through comparison of patient data and simulations, the authors highlighted the fact that the self-renewal potential of the first emerging leukaemic clone would have a major impact on the emergence of clonal heterogeneity so that it might serve as a biomarker of patient prognosis. A recent study of the group [38] on acute leukaemias formalised the clonal selection dynamics via integro-differential equations. They concluded that clonal selection was driven by the self-renewal fraction of Leukaemic Stem Cells (LSCs), constructing numerical solutions based on patient data parameters from the existing literature. These simulations showed that high self-renewal for LSC clones was a marker of stability in the presence of interclonal heterogeneity.

The model set out in Eq. (2.1) was further used in [39] to study feedback signals from myelodysplasic syndrome (MDS) clones and their effect on normal haematopoiesis. The model was fitted using serum samples from 57 MDS patients and five healthy controls. On the basis of the numerical simulations, the authors reached the conclusion that a high self-renewal fraction of MDS-initiating cells may be critical for the development of the disease. It was conjectured that remission could be achieved if this parameter could be lowered.

Considering the dependence of leukaemic cell to cytokines, the model (2.1) is compared in [40] to a mathematical model including cytokine-independent leukaemic cell proliferation. In it, leukaemic cells are not controlled by cell signalling as in Eq. (2.1g), but instead a death rate is included that increases with the number of cells in the bone marrow, and acts on all cell types residing in bone marrow. This allows the authors to explain unexpected responses in some patients, such as blast crises or remission without chemotherapy. This was done by assigning patient data to two different groups that differ with respect to overall survival: those with cytokine-dependent or cytokine-independent leukaemic cell populations.

2.1.2 Cell-cycle-based mathematical models of myeloid leukaemias

In some CML patients, symptoms may recur [41]. This is why periodicity is specifically studied for this disease. Thus, several authors considered the cell cycle in order to explain periodicity.

The cell cycle is the process regulating cell division. It is a multi-stage process including, firstly, mitosis (M), the process of nuclear division; and a stage called interphase, the interlude between two M phases. In the interphase, three different substages occur: the G_1 phase, in which the cell prepares DNA synthesis; the S phase, where DNA replicates; and the G_2 phase, where the cell prepares for mitosis. Any cell, before going the S phase, can enter a resting state called G_0, where the cell becomes quiescent and remains in a non-proliferating stage. This process is summarised in Figure 2.3.

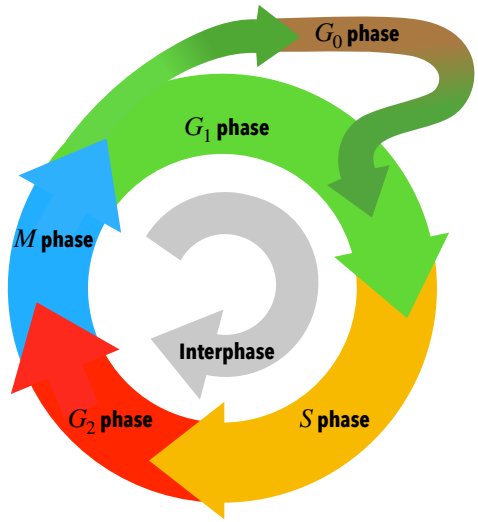


Figure 2.3: **Overview of the cell cycle**. The G_1 phase prepares for DNA replication and synthesis in the S phase. G_2 prepares cells for the mitosis phase M. While in G_1, cells can become quiescent, entering a resting phase G_0.

Many mathematical models have considered different aspects of the cell cycle [42]. However, many of those models, arising from the so-called systems

biology approach, are quite complex. Due to the periodic nature of the cell cycle in proliferating cell populations, several mathematical models have tried to account for this cycling behaviour in a simplified form. Periodicity and other dynamic behaviours of haematological diseases are reviewed in [43].

Specifically, the model in [44] described the dynamics of blood pluripotential stem cells. Their approach was to write equations for a population $N(t)$ of cells in the resting phase G_0 and another $P(t)$ of proliferating cells, described by

$$\frac{dN}{dt} = -\delta N - \beta(N)N + 2\beta(N_\tau)N_\tau e^{-\gamma\tau}, \quad \tau < t \tag{2.2a}$$

$$\frac{dP}{dt} = -\gamma P + \beta(N)N - \beta(N_\tau)N_\tau e^{-\gamma\tau}, \quad \tau < t, \tag{2.2b}$$

where $N_\tau = N(t-\tau)$, τ being the cell cycle time. The function $\beta(N) = \beta_0/\left(1+(N/N_*)^n\right)$ is the mitotic re-entry rate, i.e. the rate of cell entry into proliferation, where β_0, N_*, n are parameters. The parameter δ is the total differentiation fraction from the G_0 phase, and γ is the fraction of irreversible cell loss from all portions of the proliferating-phase stem-cell population. Taking values for these parameters from the literature, the authors concluded that the origin of aplastic anaemia and periodic haematopoiesis could be related to irreversible cell loss from the blood pluripotential stem compartment.

In [45], Eqs. (2.2) were studied to describe the existence and stability of long-period oscillations of stem cell populations in periodic chronic myelogenous leukaemia. This was made possible by studying a contractive return map, such that a fixed point of the return map gave a stable periodic solution of the model equation. This was computed in such a way that there was no analytic formula for the periodic solution in the limiting case $n \to \infty$.

Other work based on the (2.2) model, such as [46], gives estimates of the model parameters for a typical normal human, and explored the changes in some of these parameters necessary to account for the quantitative data on leukocyte, platelet and reticulocyte cycling in 11 patients with Periodic Chronic Myelogenous leukaemia (PCML). Their results indicated that the critical model parameter changes required to simulate the PCML patient data were an increase in the amplification in the leukocyte line, an increase in the differentiation fraction from the stem cell compartment into the leukocyte line, and the rate of apoptosis in the stem cell compartment. In a companion study [47], they found that the parameter changes that mimic untreated cyclical neutropenia correspond to a decreased amplification (increased apoptosis) within the proliferating neutrophil precursor compartment, and a decrease in the maximal rate of re-entry into the proliferative phase of the stem cell

compartment. The case of granulocyte colony stimulating factor treatment was also studied. Safarishahrbijari and Gaffari [48] used the equations for red blood cells and platelets from [46] and for leukocytes from [43] to identify parameters in PCML. The inclusion of new parameters resulted in a better fit of clinical data and from the data extracted from both platelet and leukocyte models.

Pujo-Menjouet and Mackey [49], performed a local stability analysis of the model (2.2) and found the conditions for Hopf bifurcation to occur. Periodic oscillations were studied depending on five haematopoietic stem cell parameters: the mitotic rate sensitivity, the maximal rate of cell entry into proliferation from the resting G_0 phase, the differentiation and apoptosis rate and the time to entry into mitosis. Extensions of this work [50], have proven that, under periodic treatment, there is a periodic solution with the same period. This could be related to the observed oscillatory behaviour of blood cells' counts under treatment in CML.

A different type of models to describe myeloblastic leukaemias have been constructed on the basis of the work of Rubinow and Lebowitz [51]. The model itself was based on granulocytopoiesis, also studied by these authors in [52]. In this first work, qualitative analysis was performed, supporting evidence for alterations which presumably occurs in cyclic neutropenia. For both models they considered four compartments for healthy cells as shown schematically in Figure 2.4: the active A and G_0 cell compartments, representing the proliferative pools, and the maturation M and reserve R cell compartments, which finally ended in the blood pool B. For the leukaemic cells, only active and G_0 cells were considered, of which only a certain fraction were released into the blood, with no further maturation stages.

For this model, and in terms of myeloid leukaemia, the presence of a leukaemic population destabilises the homoeostatic state of the normal population, which is stable in the absence of leukaemic cells. In [53], the authors found differences between normal and leukaemic cell populations but including treatment into the model from Figure 2.4: firstly, the recovery rate was higher for normal cells, as compared to leukaemic cells from the action of cytotoxic treatment. Secondly, the S-phase duration was different for the two populations. This led the authors to the conclusion that, for patients with a "slow" growing leukaemic cell population, remission could be achieved with one or two courses of treatment, whereas for those with a "fast" growing leukaemic cell population, a similar aggressive treatment achieved remission only at the cost of great toxic effects on the normal cell population.

These mathematical models described both the processes of normal blood and myelogenous leukaemia development. It was done the same in [54], but for

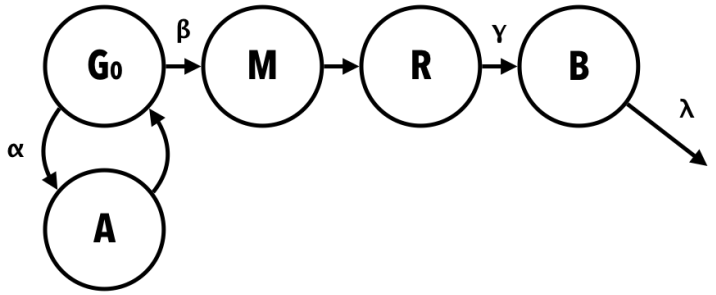


Figure 2.4: **Schematic view of the cell compartments in [51].** A, G_0, M, R and B represent different cell compartments, α, β, γ are the corresponding coupling rates between them, and λ is the irreversible blood cell loss.

CML. The authors showed how CML cells could ultimately outnumber normal cells. They used the model to study the relationship between proliferation and maturation and proposed a solution to the apparent contradiction between decreased proliferation and increased production, by assuming that a greater fraction of CML cells is produced by division rather than by maturation.

Another mathematical model [55] included details on cyclins D, E and B, a family of proteins that help to control the cell cycle. Their production has a direct influence on the transition of a cell in the G_0, G_1 and G_2 phases, respectively. Flow-cytometry data profiles for three leukaemia cell lines were analysed in this study (K-562, MEC-1, and MOLT-4, from AML, CLL and ALL patients, respectively). For the S phase, DNA replication was considered, as it is key before a cell can produce new daughter cells. The authors assumed that $G = G(C_E, t)$ was the number of cells in G_0/G_1 at time t with a cyclin E content C_E. Similarly, they denoted by $S = S(DNA, t)$ and $M = M(C_B, t)$ the number of cells in the S and G_2/M phases that had DNA content (represented as the variable DNA) and cyclin B content C_B at time t, respectively. These assumptions taken together led to the model

$$\frac{\partial G}{\partial t} + \frac{\partial \left(G \cdot \dfrac{dC_E}{dt} \right)}{\partial C_E} = -r_{G \to S}(C_E) \cdot G, \tag{2.3a}$$

$$\frac{\partial S}{\partial t} + \frac{\partial \left(S \cdot \frac{dDNA}{dt} \right)}{\partial DNA} = 0, \tag{2.3b}$$

$$\frac{\partial M}{\partial t} + \frac{\partial \left(M \cdot \frac{dC_B}{dt} \right)}{\partial C_B} = -r_{M \to G}(C_B) \cdot M, \tag{2.3c}$$

where $r_{G \to S}$, $r_{M \to G}$ are the transition fractions from G_2/M to G_0/G_1 and from G_0/G_1 to the S phase, respectively. Good agreement was found between experimental results and the model simulations. This could assist in developing clonal models of leukaemogenesis. The authors claimed that the model could help in the identification of heterogeneous leukaemia clones at diagnosis and post-treatment, and that it could have the potential to predict future outcomes in response to induction and consolidation chemotherapy as well as relapse kinetics.

2.1.3 Other data-based mathematical models of myeloid leukaemia

Myeloid leukaemia models are the most studied in the literature. In Ref. [56], for example, acute myeloid leukaemia (AML) is described using a multi-lineage multi-compartment model of the haematopoietic system and feedback via cytokines and chemokines. Analysis of the model suggested that self-renewal probabilities, mitotic rates and cytokine growth factors produced in peripheral blood determined leukocyte homoeostasis. The mitosis rate of cancer was found to be the parameter with the strongest prognostic value.

A comparison of three mathematical models that describe CML progression and aetiology was undertaken in [57]. The authors sought to identify which models could provide the best description of disease dynamics and their underlying mechanisms. The first considered the following dynamic system

$$\frac{dx_0}{dt} = a_x x_0(k - z) - b_x x_0, \tag{2.4a}$$

$$\frac{dx_1}{dt} = b_x x_0 + c_x c_1(k - z) - d_x x_1, \tag{2.4b}$$

$$\frac{dx_2}{dt} = d_x x_1 - e_x x_2, \tag{2.4c}$$

$$\frac{dy_0}{dt} = a_y y_0(k - z) - b_y y_0, \tag{2.4d}$$

$$\frac{dy_1}{dt} = b_y y_0 - e_y y_1, \tag{2.4e}$$

where x_0 were HSCs, x_1 healthy progenitors, x_2 differentiated cells, y_0 LSCs and y_1 differentiated leukaemic cells, with parameters a, b, c, d, e as the corresponding self-renewal, production and death rates. Also, k was the carrying capacity and $z = x_0 + x_1 + x_2 + y_1 + y_2$ the total number of cells. The second model, in [58], was a shorter version of the (2.11) model, to be presented in detail later. The third was [59], which allowed competition between HSC and LSCs. This latter model was based on the following ODEs:

$$\begin{aligned}
\frac{dx_0}{dt} &= \beta_x x_q + & \frac{dy_0}{dt} &= \beta_y y_q + \\
&\quad +(r_x\phi_x - d_0 - \alpha_x)x_0, & &\quad +(r_y\phi_y - d_0 - \alpha_y)y_0, \\
\frac{dx_q}{dt} &= \alpha_x x_0 - \beta_x x_q & \frac{dy_q}{dt} &= \alpha_y y_0 - \beta_y y_q \\
\frac{dx_1}{dt} &= a_x x_0 - d_1 x_1, & \frac{dy_1}{dt} &= a_y y_0 - d_1 y_1, \\
\frac{dx_2}{dt} &= b_x x_1 - d_2 x_2, & \frac{dy_2}{dt} &= b_y y_1 - d_2 y_2, \\
\frac{dx_3}{dt} &= c_x x_2 - d_3 x_3, & \frac{dy_3}{dt} &= c_y y_2 - d_3 y_3.
\end{aligned} \tag{2.5}$$

The healthy cells x_i and leukaemic cells y_i were considered at different stages $i = 0, ..., 3$ of differentiation and a compartment of quiescent cells was also added for each type, x_q and y_q. The authors found that it was not possible to choose between the models based on fits to the data of 69 patients who had experienced relapse or remission of the disease. They suggested experiments directly probing the haematopoietic stem-cell niche that could help in choosing the best model.

Finally, [60] described another model of cancer initiation for CML. The authors assumed that the clonal expansion of mutant cells is given by a logistic equation

$$\frac{dx}{da} = \frac{r-1}{\tau} x(1-x), \qquad \text{with } x(0) = \frac{1}{N}, \tag{2.6}$$

where a is the time since mutation happened, $x(a)$ the frequency of mutant clones with r relative fitness, and N the total cell population with generation time τ. q was the rate of detection and u the probability per cell division of producing a mutant cell. Letting $c = \frac{r-1}{\tau}$, and $b = N\,u\,\frac{c}{r}$, the probability of

detecting cancer before time t was given by

$$P(t) = \int_0^t b \exp(-b\,m)\,(1 - \exp(H(t,m)))\; dm, \tag{2.7}$$

for

$$H(t,m) = -q\,N \int_0^{t-m} \frac{da}{1 + (N-1)\exp(-c\,a)}, \tag{2.8}$$

with m a small probability that the first mutant arose early. Interestingly, this simple model, based only on the Philadelphia translocation, gave rise to cancer incidence curves with exponents of up to 3 as a function of age. This behaviour had been previously thought to be associated necessarily with three mutations, two of which were unknown. Thus, the model proved that CML incidence data were consistent with the hypothesis that the Philadelphia translocation alone could cause CML.

2.1.4 Other theoretical studies of myeloid leukaemias

Cancer initiation and maintenance are typically assumed to be related to cancer stem cells (CSCs) [61, 62]. Two models for cancer initiation have been derived using this assumption. The first is a genetic mutation model, where mutations determine the phenotype of the tumour. In this conceptual framework different mutations may result in different tumour morphologies, even when starting from the very same stem cell. Cells inherit the molecular alteration and regain the ability for self-renewal, which leads to a population of cancer cells. The second model assumes that different cells serve as cells of origin for the different cancer subtypes, the so-called CSCs. This model proposes that oncogenic events occur in different cells, and these produce different kinds of cancer. In this model, self-renewal potential is limited for the CSCs. Both conceptual models are shown schematically in Figure 2.5.

[63] constructed a stochastic model that considered drug resistance for CML, where the probability of treatment failure was approximated by

$$M_0 \frac{n!(L-D)L^{n-1}u^n}{(D+H-L)^n}, \tag{2.9}$$

for M_0 the initial non-mutant cells, n the quantity of drugs used, u the probability of mutation after cell division, and finally, the measurable parameters L, D and H as, respectively, the rate of growth, death and the drug-induced death rate. From the analysis of the mathematical model, the authors claimed

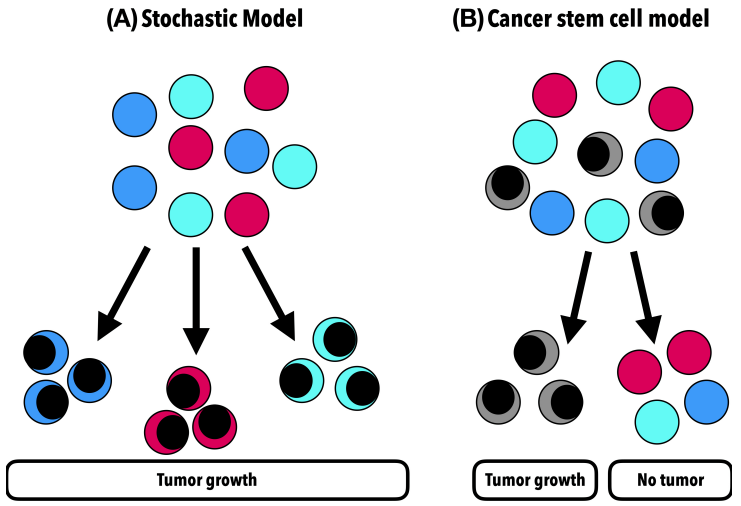


Figure 2.5: **Different representations of tumour proliferation models**. In the stochastic model, tumour cells are heterogeneous, so that genetic changes that lead to tumour cells can originate from any cell. In the CSC model, only a small subset of cancer cells has the ability to initiate new tumour growth. Figure adapted from Ref. [5].

that although drug resistance prevented successful treatment, resistance could be overcome with a combination of three targeted drugs.

Finally, several authors have built models of leukaemias using graph-theoretical methods. Graphs can be used to describe the hierarchical organisation observed in haematopoiesis, as seen in Figure 1.1. In Ref. [64] a graph-theoretical model of haematopoiesis was parametrised using publicly available RNA-Seq data in a high-dimensional space. The high-dimensional data were later reduced to $\mathbb{R}^2$ or $\mathbb{R}^3$ using reduction techniques, such as principal component analysis, diffusion maps and t-distributed stochastic neighbour embedding, and a PDE model on a graph G was constructed. $u(x,t)$ denoted the cell distribution at the differentiation continuum space location $x \in G$ and time t. Then, for every cell distribution $u_k(x,t)$ on an edge e_k, the cell density

was modelled with advection-diffusion-reaction equations

$$\frac{\partial u_k}{\partial t} = R_k u_k - \frac{\partial \left(V_k u_k\right)}{\partial x} + \frac{D_k}{2w_k}\frac{\partial}{\partial x}\left(w_k \frac{\partial u_k}{\partial x}\right), \tag{2.10}$$

for $x \in e_k = \overline{a_k b_k}$, where each edge e_k was parametrised from a_k to b_k, and the following functions were considered: $R_k = R_k(x)$ as the cell proliferation, $V_k = V_k(x)$ as the advection coefficient, and apoptosis and diffusion terms $D_k = D_k(x)$ and $w_k = w_k(x)$, which respectively describe cell fluctuation and width of a narrow domain around an edge. Using this model, the authors performed simulations consistent with the evolution of AML populations. A similar approach was used in [65], where the graphs constructed presented the essential properties of functioning bone marrow.

2.2 Mathematical description of Chronic Myeloid Leukaemia Treatments

2.2.1 Imatinib and its basic mathematical modelling

CML has been intensively studied in terms of therapy based on Imatinib. This drug is a 2-phenyl amino pyrimidine derivative that inhibits a number of tyrosine kinase (TK) enzymes. Imatinib is specific for the TK domain in ABL (the Abelson proto-oncogene), c-kit and PDGF-R (platelet-derived growth factor receptor). In chronic myelogenous leukaemia, the Philadelphia chromosome leads to a fusion protein of ABL with the breakpoint cluster region, termed BCR-ABL. Imatinib decreases the BCR-ABL activity. CML treatments have been strongly influenced by the appearance of imatinib [66], that is now the standard first-line treatment against the disease. It is a very effective drug with up to about 70% of people having a complete cytogenetic response (CCyR) within 1 year of starting imatinib. After a year, even more patients will have had a CCyR. Many of these patients also have a complete molecular response (CMR).

The capacity of the drug to impair the proliferation of leukaemic stem cells was the basic assumption behind the mathematical model of Michor and co-workers [58]. The model also included the development of resistance to

therapy and was based on the following system of differential equations:

$$\begin{aligned}
\frac{dx_0}{dt} &= (\lambda(x_0) - d_0)x_0, & \frac{dy_0}{dt} &= (r_y(1-u) - d_0)y_0, \\
\frac{dx_1}{dt} &= a_x x_0 - d_1 x_1, & \frac{dy_1}{dt} &= a_y y_0 - d_1 y_1, \\
\frac{dx_2}{dt} &= b_x x_1 - d_2 x_2, & \frac{dy_2}{dt} &= b_y y_1 - d_2 y_2, \\
\frac{dx_3}{dt} &= c_x x_2 - d_3 x_3, & \frac{dy_3}{dt} &= c_y y_2 - d_3 y_3, \\
\frac{dz_0}{dt} &= (r_z - d_0)z_0 + r_y y_0 u, \\
\frac{dz_1}{dt} &= a_z z_0 - d_1 z_1, \\
\frac{dz_2}{dt} &= b_z z_1 - d_2 z_2, \\
\frac{dz_3}{dt} &= c_z z_2 - d_3 z_3.
\end{aligned} \tag{2.11}$$

Here, x_i denotes the different populations of normal cells, y_i the imatinib-sensitive leukaemic populations and z_i the tumour clones resistant to imatinib. The indexes $i = 0, 1, 2, 3$, denote the subpopulations of stem cells, progenitors, differentiated and terminally differentiated cells in each compartment. The rate constants for each cell type (x, y, z) are given by a, b and c, and d_i are the death rates for $i = 0, 1, 2, 3$. Cell division rates are given by r_y and r_z. The parameter u is the fraction of resistant cells produced per cell division. Finally, $\lambda = \lambda(x_0)$ is a decreasing function of x_0 describing homoeostasis of normal stem cells. It models the feedback signals controlling haematopoiesis. Data from 169 CML patients were used to fit the mathematical model in [58]. The authors obtained numerical estimates for the turnover rates of leukaemic progenitors and differentiated cells and showed that imatinib dramatically reduced the rate at which these cells are being produced from leukaemic stem cells. They showed that the probability of harbouring resistance mutations increases with disease progression as a consequence of an increased leukaemic stem cell abundance, and proposed that the time to treatment failure caused by acquired resistance is given by the growth rate of the leukaemic stem cells. Their bottom line was that multiple drug therapy is especially important for patients who are diagnosed with advanced and rapidly growing disease.

A simplified version of the model (2.11) was studied in [67] by considering only the stem cell (0) and differentiated cell (1) compartments of healthy (x) and leukaemic (y) cells, i.e.

$$\frac{dx_0}{dt} = (r_x\phi - d_0)x_0, \tag{2.12a}$$
$$\frac{x_1}{dt} = a_x x_0 - d_1 x_1, \tag{2.12b}$$
$$\frac{y_0}{dt} = (r_y\varphi - d_0)y_0, \tag{2.12c}$$
$$\frac{y_1}{dt} = a_y y_0 - d_1 y_1, \tag{2.12d}$$

where $\phi = 1/\left[1 + c_x\left(x_0 + y_0\right)\right]$ and $\psi = 1/\left[1 + c_y\left(x_0 + y_0\right)\right]$ are homeostasis functions for normal and tumour stem cells respectively, and c_x, c_y are Michaelis-Menten parameters. By a combination of analysis and simulation, the authors discussed how any successful therapy would require the eradication of the pool of leukaemic stem cells; otherwise, progressive disease is very likely. Thus, successful therapeutic agents must enhance the death rate of this rare population of cells. Therapies designed to target mitosis of malignant stem cells could not eradicate the disease quickly. Nevertheless, there has been some controversy surrounding the potential effectiveness of imatinib to achieve remission [68].

In [69], the immune response targeting leukaemic cells was added to Eqs. (2.11). Using experimental data from the literature, a mathematical model was fitted in which immune response was described by delay differential equations. The authors considered that T cells targetting leukaemic cells could prevent relapse, and combine with imatinib therapy. The more simplified model in Eq. (2.12) was later used by [70] to study and numerically simulate treatment interruptions as a potential therapeutic strategy for CML patients. In many cases, strategic treatment interruptions led to the elimination of leukaemic cells in silico. The conclusion was that strategic treatment interruptions could be a feasible clinical approach to enhancing the effects of imatinib treatment for CML.

A number of extensions of the (2.11) model have been developed for CML. For example, in [71], four levels of cell differentiation were included and studied for the BCR-ABL1 gene, necessary for the pathogenesis of CML. In that study, data from 290 patients were used, 92 of them treated with dasatinib, 75 with nilotinib and 123 with imatinib. All treatments elicited similar responses. Another extension of the model was described in [72], with a focus on more theoretical aspects, including a stability analysis, and an existence proof for

positive solutions.

The global dynamics of normal and CML haematopoietic stem cells and differentiated cells were also studied in [73]. The dynamic was assumed to be governed by the following system of Lotka-Volterra equations

$$\frac{dx_0}{dt} = n\left(1 - \frac{x_0 + y_0}{K}\right)x_0 - d_0 x_0, \tag{2.13a}$$

$$\frac{dx_1}{dt} = rx_0 - (d - d_2)x_1, \tag{2.13b}$$

$$\frac{dy_0}{dt} = m\left(1 - \frac{x_0 + \alpha y_0}{K}\right)y_0 - g_0 y_0, \tag{2.13c}$$

$$\frac{dy_1}{dt} = qy_0 - (g - g_2)y_1, \tag{2.13d}$$

where $x_0(t)$ represents haematopoietic normal stem cells (HSC), $x_1(t)$ normal differentiated cells and $y_0(t)$ and $y_1(t)$ describe the same subpopulations of cancer cells. In Eqs. (2.13) n, m, r, q are division rates, d_0, d, g_0, g death rates, K the carrying capacity and $\alpha \in]0, 1[$ is a constant. The production rates for differentiated cells are given by d_2 and g_2. Several optimal control problems were solved for imatinib, whose effect on the division and mortality rates of cancer cells produces a suboptimal response. The effect of cyclic combination of two drugs in CML was studied in [74], and the modelling led to the conclusion that treatments should start with the stronger drug, and the weaker one should have cycles of longer duration.

An interaction model between naïve T cells (mature T cells from thymus), effector T cells (cells which actively respond to stimuli) and CML cancer cells was described in [75], where Latin hypercube sampling was used to estimate parameter values due to the lack of data. This is a statistical technique for generating parameters from a multidimensional distribution. In their conclusion, the authors explained that the growth rate of CML and the natural death rate were the most important parameters, suggesting that treatment for CML patients should focus on these rates. Any drug with a high cost that is included in the model could be studied in order to obtain optimal treatment, and reduce not only radiation but also financial benefits. This model was later used in [76], focusing on cancer $x = x(t)$ and effector $y = y(t)$ cell population dynamics, by considering a combined treatment with imatinib and the interferon-alpha (IFN-α) therapy. This last is a protein whose activation produces a cytogenetic

response in CML patients. The model considered the following ODEs

$$\frac{dx}{dt} = \beta_1 x(t) \ln \frac{K}{x(t)} - \gamma_1 x(t) y(t) - \omega \gamma_3 x(h(t)), \tag{2.14a}$$

$$\frac{dy}{dt} = \beta_2 \frac{x(t)}{\eta_1 + x(t)} y(t) - \gamma_2 x(t) y(t) + \tag{2.14b}$$

$$+ i n_\alpha \gamma_4 \frac{y(t)}{\eta_2 + y(t)} y(t - \tau) - \mu_y y(t),$$

where β_1, β_2 were the respective reproduction rates, K the maximal tumour population, η_1, η_2 Michaelis-Menten terms and γ_1, γ_2 the cell loss rates due to interaction. The death rate for effector cells is μ_y, while tumour death is modelled by the constants $\omega\gamma_3$ and a function $h(t)$. This function is modelled as $h(t) = t - \theta e^{-\lambda t}$, so that the influence of drugs tends to zero over time. The dose of IFN-α is modelled as $in_\alpha\gamma_4$, which increases the effector cell population with a delay τ of about 7 days. The stability analysis proposed, as well as the results obtained, were able to describe the influence of two types of the treatment. The authors claimed that the dose of IFN-α has an inhibitory effect on the number of cancer cells, but its replacement with another type of treatment should be considered in order to avoid resistance.

Finally, [77] studied optimal control problems for CML, in a model with a molecular targeted therapy such as imatinib. Naïve T cells, which are already differentiated T cells, but are precursors for more mature cells called effector cells, were also included in the model. The cancer cell population was then activated by the presence of the CML antigen. Aiming to minimise the cancer cell population and the detrimental effects of the drug, they found that a high dose level from the beginning was optimal. Also, combination therapy was better than single dosing.

2.2.2 Modelling the effect of quiescence on Imatinib treatments

Quiescence, which corresponds to the G_0 phase of the cell cycle, and its relationship to drug therapy (in this case, imatinib) is an important factor in leukaemia because quiescent cells might not be affected by therapy, as drugs target proliferative cells, and a possible relapse may occur.

Imatinib treatment was studied using Roeder model [78, 79, 80] accounting for quiescent and proliferative cell compartments. Firstly, in [78] a stochastic model of haematopoiesis was developed. On the basis of that model, another was built to describe imatinib-treated patients [79].

A more advanced model based on partial differential equations (PDEs) was studied in [80]. This model considered quiescent and cycling stem cells, denoted by A and Ω, respectively. The authors included a cell-intrinsic function $a(t)$, which determined the affinity of a cell for residing in A or Ω. With $a(t) \in [a_{\min}, a_{\max}]$, a quiescent stem cell would enter the cell cycle with probability ω and a cycling cell would become quiescent with probability α. These terms were modelled as

$$\omega(\Omega(t), a(t)) = \frac{a_{\min}}{a(t)} f_\omega(\Omega(t)), \tag{2.15a}$$

$$\alpha(A(t), a(t)) = \frac{a(t)}{a_{\max}} f_\alpha(A(t)), \tag{2.15b}$$

where the sigmoidal functions f_ω and f_α were defined by

$$f_\omega(\Omega(t)) = \frac{1}{\nu_1 + \nu_2 \exp\left(\nu_3 \frac{\Omega(t)}{N_\omega}\right)} + \nu_4, \tag{2.15c}$$

$$f_\alpha(A(t)) = \frac{1}{\mu_1 + \mu_2 \exp\left(\mu_3 \frac{A(t)}{N_\alpha}\right)} + \mu_4, \tag{2.15d}$$

for specific values of the parameters ν_j, μ_j, for $j = 1, 2$ and the scaling factors N_ω and N_α. The dynamics of the HSCs, quiescent (A) and proliferating (Ω), were governed by these equations:

$$\frac{\partial n_A}{\partial t} + v_A \cdot \frac{\partial}{\partial a} n_A = -\left(\frac{dv_A}{da} - \omega\right) \cdot n_A + \alpha \cdot n_\Omega, \tag{2.15e}$$

$$\frac{\partial n_\Omega}{\partial t} + v_\Omega \cdot \frac{\partial n_\Omega}{\partial a} = \left(-\frac{dv_\Omega}{da} + \tau - \alpha\right) \cdot n_\Omega + \omega \cdot n_A. \tag{2.15f}$$

The functions $n_A = n_A(a, t)$ and $n_\Omega = n_\Omega(a, t)$ represent the cell densities at affinity a and time t within A and Ω, respectively. Also, $v_A = v_A(a)$ and $v_\Omega = v_\Omega(a)$ were the corresponding velocities that make $v_A \cdot n_A$ and $v_\Omega \cdot n_\Omega$ the corresponding cell fluxes for each compartment. Finally, τ was a parameter which simulates average cell division depending on cell cycle duration. Eqs. (2.15e) and (2.15f) were the basis for studying leukaemia and how the imatinib treatment affect its dynamics, in a highly efficient way when it comes to huge cell populations. They considered the dynamics for every cell subpopulation in

the following system:

$$\frac{\partial n_A^{(i)}}{\partial t}+v_A^{(i)}\frac{\partial n_A^{(i)}}{\partial a}=-\left(\frac{dv_A^{(i)}}{da}-\omega^{(i)}\right)n_A^{(i)}+\alpha^{(i)}n_\Omega^{(i)}, \tag{2.15g}$$

$$\begin{aligned}\frac{\partial n_\Omega^{(1)}}{\partial t}+v_\Omega^{(1)}\frac{\partial n_\Omega^{(1)}}{\partial a}&=\omega^{(1)}n_A^{(1)}+\\&+\left(-\frac{dv_\Omega^{(1)}}{da}+\tau^{(1)}-\alpha^{(1)}\right)n_\Omega^{(1)},\end{aligned} \tag{2.15h}$$

$$\begin{aligned}\frac{\partial n_\Omega^{(2)}}{\partial t}+v_\Omega^{(2)}\frac{\partial n_\Omega^{(2)}}{\partial a}&=\omega^{(2)}n_A^{(2)}+\\&+\left(-\frac{dv_\Omega^{(2)}}{da}+\tau^{(2)}-\alpha^{(2)}-r_{\text{inh}}-r_{\text{deg}}\right)n_\Omega^{(2)},\end{aligned} \tag{2.15i}$$

$$\begin{aligned}\frac{\partial n_\Omega^{(3)}}{\partial t}+v_\Omega^{(3)}\frac{\partial n_\Omega^{(3)}}{\partial a}&=\omega^{(3)}n_A^{(3)}+r_{\text{inh}}n_\Omega^{(2)}+\\&+\left(-\frac{dv_\Omega^{(3)}}{da}+\tau^{(3)}-\alpha^{(3)}-r_{\text{deg}}\right)n_\Omega^{(3)},\end{aligned} \tag{2.15j}$$

where the super indexes i represent the different cell populations as normal cells ($i=1$), imatinib-affected leukaemic cells ($i=2$) and non-affected leukaemic cells ($i=3$). Induced cell death is denoted by a constant r_{deg}, while the constant r_{inh} denotes the proliferation inhibition on the proliferating cells $n_\Omega^{(2)}$. The model in Eq. (2.15) was proved to qualitatively and quantitatively reproduced the results of the agent-based approach for imatinib-treated patients in [79]. This was fitted to 894 peripheral blood samples, where the authors claimed that the therapeutic benefits of imatinib can, under certain circumstances, be accelerated by being combined with proliferation-stimulating treatment strategies.

[81] described an extension of the (2.15) model. This was done by considering the cycling cells Ω to be dependant, among other variables, on a counter $c(t)$, that indicates the position in the cell cycle, with a 49-hour cell cycle. An imatinib treatment was then incorporated into the model. The authors conclude that PDE formulation provided a more efficient way of simulating the dynamics of the disease. In fact, in simulations of imatinib treatment, the PDE and the discrete-time models diverged more, as in this case a continuous-time description of the disease dynamics may be more realistic than discrete-time models. This latter model was later extended [82] by including feedback from

cells and asymmetric division for stem cells and precursors. The general idea for this work was also to combine imatinib with a drug that induced cancer stem cells to cycle. Furthermore, the fact that many patients do relapse after being taken off imatinib motivates the study methods by which this therapy can be improved. [83] performed a stability analysis of the model in [81], where the authors could set differences between AML and CML in terms of transition from stable equilibrium to unstable periodic behaviour.

2.2.3 Whole body mathematical description of leukaemia and its treatment

Leukaemia treatment may affect blood flux in several tissues on the body. In order to understand the behaviour of these body parts during therapy, we set out a highly descriptive model of leukaemia, chemotherapy and blood flux throughout the entire body [84]. The inflow rate of drug j is

$$\text{inflow}_j = \frac{u_j}{duration_j}, \tag{2.16a}$$

where u_j is the drug dose over $duration_j$. This equation was then incorporated into the following equation, which models drug concentration in the blood $C_{B,j}$:

$$V_B \cdot \frac{dC_{B,j}}{dt} = \sum_{i=H,Li,M,Le,K} Q_i \cdot C_{i,j} - Q_B \cdot C_{B,j} + \\ + \text{inflow}_j. \tag{2.16b}$$

In this equation, V_B is total patient blood volume, and Q_i the blood flow in the organs i, such as heart (H), liver (Li), bone marrow (M), lean muscle (Le) and kidneys K, and so $C_{i,j}$ was the concentration of drug j in the organs i, modelled as

$$V_i \cdot \frac{dC_{i,j}}{dt} = Q_i \cdot C_{B,j} - Q_i \cdot C_{i,j} \\ - k_{k,j} \cdot C_{B,j} - k_{L,j} \cdot C_{i,j} \cdot V_{i,T}, \tag{2.16c}$$

for every organ i and drug j, where $k_{k,j}$ is the urine excretion rate, $k_{L,j}$ the elimination rate in the liver, and $V_{i,T}$ the volume of organ tissue where drug metabolism occurs. This model, along with many others, are useful in clinical terms, as it could provide guidance for optimising treatment for each patient in terms of their characteristics, as explained in Figure 2.6.

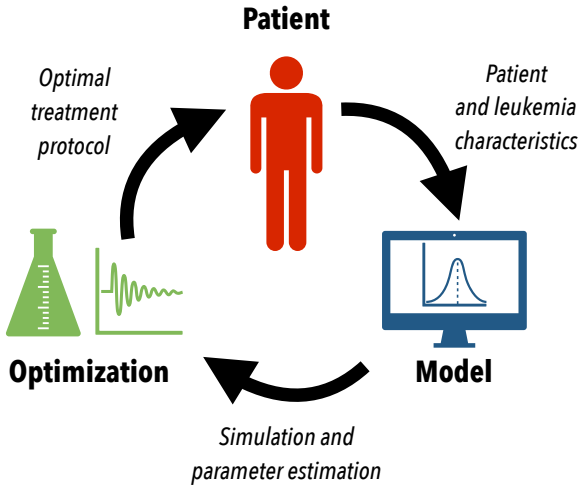


Figure 2.6: **Schematic view of the use of mathematical models to help in patient treatment**. Certain features are acquired from leukaemia, and specific ratios from each patient could be implemented in the mathematical model as parameters. After the simulation, several parameters might arise that could be advantageous for specific treatment protocol, optimised for the specific patient, who could benefit from the personalised drug. This cycle could be useful for following the disease in the patient. Figure adapted from Ref. [84].

This pharmacokinetic model was reinforced by a pharmacodynamic model, which took into account the effect of the drug. Drug concentration at the location of the tumour, which for leukaemia would be the concentration of drug in the bone marrow ($C_{M,j}$), was considered for the j effect of the drug as the function effect$_j$. It was included in the cell cycle as

$$\frac{dP_y}{dt} = k_{y-1} \cdot P_{y-1} - k_y \cdot P_y - \text{effect}_j \cdot P_y, \tag{2.16e}$$

where P_y was the cell population in phase y (G_1, S, G_2, M) and k_y the transition term from phase y to $y+1$.

Although these equations are described in a general sense, for the specific case of chemotherapy cycles of intravenous (IV) daunorubicin (DNR) and

cytarabine ($Ara - C$), typical drugs in leukaemia treatment, the reactions occurred at a subcutaneous level. That is, the drug is injected under the skin and not below muscle tissue. This drug and its subcutaneous effect have also been addressed in other studies, such as [85], fitting data from 44 AML patients during consolidation therapy to a pharmacokinetic mathematical model, obtaining optimised treatment schedules. However, the authors of [84] considered, when simulating the subcutaneous effect of the therapy, that Eq. (2.16c) could then be replaced by the following two:

$$\frac{dS}{dt} = \text{inflow} - k_a \cdot k_b \cdot S, \tag{2.16f}$$

$$V_B \cdot \frac{dC_B}{dt} = \sum_{i=H,Li,M,Le,K} -Q_i \cdot C_{i,j} - Q_B \cdot C_{B,j} + k_a \cdot k_b \cdot S, \tag{2.16g}$$

where S is the subcutaneous tissue drug delivery, k_a the absorption delay and k_b the drug bioavailability. However, the simulations performed were adapted for two acute myeloid leukaemia patients. Sensitivity analysis method was applied on the model to identify the most crucial parameters that control treatment outcome. The results clearly showed benefits from the use of optimisation as an advisory tool for treatment design.

The whole (2.16) model was a clear example of the usefulness of mathematical models for therapy planning.

2.3 Mathematical models of Acute Lymphoblastic Leukaemia treatments with cytotoxic drugs

The current standard treatment of acute lymphoblastic leukaemia involves different treatment stages: induction, consolidation, re-induction whenever needed, and maintenance [86]. The aggressiveness of treatments depends on the classification of patients into risk groups: standard, average or high (Figure 2.7).

The goal of the induction stage is to achieve a rapid reduction in tumour cell numbers. Next, the consolidation phase should ideally remove any trace of leukaemic cells in flow-cytometry or blood cell count studies. Re-induction is considered whenever leukaemic clones reappear early. The maintenance phase is administered when the first two phases are completed, and is intended to kill any possible remaining non-measurable quantities of cancer cells. Every phase

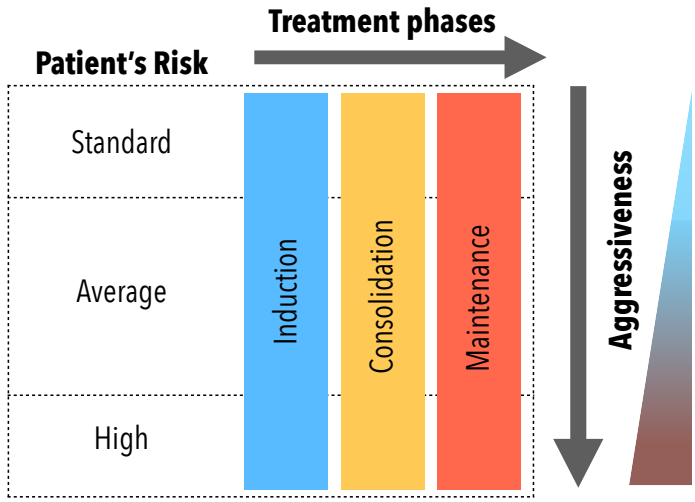


Figure 2.7: **Stages of treatment administration for ALL**. These statges depend on the patient risk group.

includes specific treatments, the doses and timings of drugs depending on the patient's risk group.

Using one mathematical model or another to describe therapy may lead to a different understanding of how treatment affects cells in terms of relapse [87]. For example, if relapse occurs and we consider a Cancer Stem Cell (CSC) model, a drug might not affect CSCs, or might only affect cells with specific mutations (in the genetic mutation model). This can be better seen in Figure 2.8.

In ALL, two drugs are used as part of these treament phases: 6-Mercaptopurine and Methotrexate. Some mathematical models of their actions are now summarised.

2.3.1 Describing the effect of mercaptopurine

Mercaptopurine (6MP) is an antimetabolite antineoplastic agent with immunosuppressant properties. It interferes with nucleic acid synthesis by inhibiting purine metabolism and is used, usually in combination with other drugs, in the treatment of or in remission maintenance programmes for leukaemia.

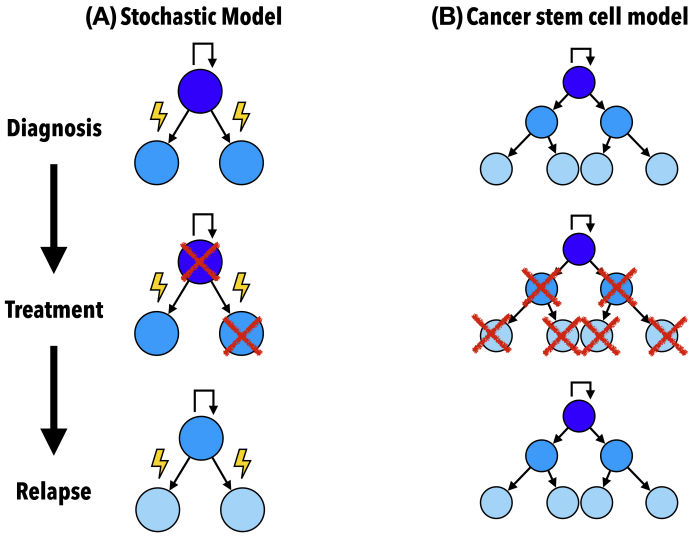


Figure 2.8: **Different representations of tumour proliferation models through the effect of therapy**. Figs. adapted from Ref. [87].

A mathematical model of the effect of 6MP in leukaemia cells was described in [88]. In this model, the number of cells in the G_0/G_1-phases was denoted by G; S in the S-phase, and M in the G_2/M-phase. The suffixed variables G_I, S_I and M_I were the equivalent variables for the thioguanine (TGN) nucleotides, which were considered as the main active metabolites. That is, the most active molecules involved in the metabolic process. Apoptotic cells A, and non-viable cells N (cells that are unable to live), were also included in the model.

The equations for the viable phases of cells are

$$\frac{dG}{dt} = -\alpha_S G + 2\beta M, \tag{2.17a}$$

$$\frac{dS}{dt} = \alpha_S G - (\alpha_M + \gamma_1) S, \tag{2.17b}$$

$$\frac{dM}{dt} = (1-f)\alpha_M S - \beta M; \tag{2.17c}$$

while those for the cells with TGN incorporated are

$$\frac{dG_I}{dt} = f\alpha_M S + \alpha_{MI} S_I - \beta_I M_I, \tag{2.17d}$$

$$\frac{dS_I}{dt} = -\alpha_{SI} G_I + 2\beta_I M_I, \tag{2.17e}$$

$$\frac{dM_I}{dt} = \alpha_{SI} G_I - (\gamma_{MP} + \alpha_{MI}) S_I. \tag{2.17f}$$

Finally, the apoptotic and non-viable phases are modelled as

$$\frac{dA}{dt} = \gamma_1 S + \gamma_{MP} S_I - \gamma_2 A, \tag{2.17g}$$

$$\frac{dN}{dt} = \gamma_2 A - \gamma_3 N. \tag{2.17h}$$

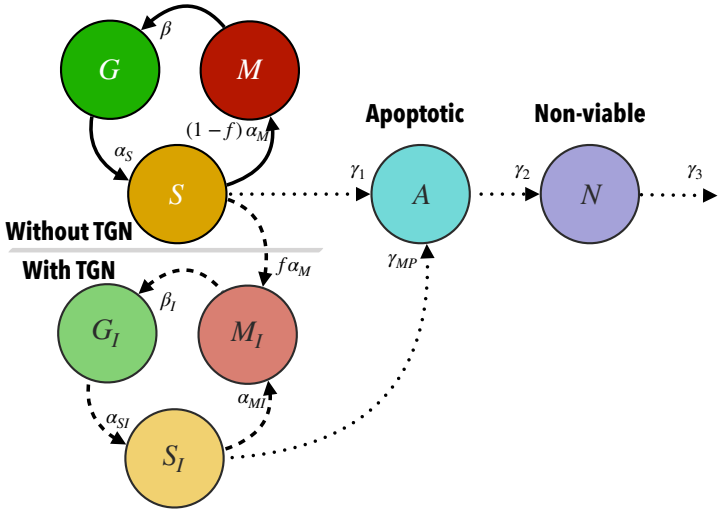


Figure 2.9: **Diagram of the model** (2.17). The different phases of the cell cycle are represented for cells with and without TGN incorporated into their DNA, before they reach the apoptotic and finally the non-viable state.

The dynamics of the model are summarised in Figure 2.9. The model parameters describe the transition between phases, except for $f \in [0, 1]$,

which measures the fraction of cells continuing the cell cycle after TGNs were incorporated into the cell DNA. To estimate these parameters, the model was fitted to data for different cell lines treated with MP. The mathematical model provided a quantitative assessment to compare the cell cycle effects of MP in cell lines with varying degrees of MP resistance.

In a different study [89], semi-mechanistic mathematical models were also designed and validated for MP metabolism, by studying red blood cell mean corpuscular volume (MCV) dynamics, a biomarker of treatment effectiveness and leukopenia, a major side effect related to very low percentages of leukocytes. The model was validated with real patient data obtained from literature and a local institution. Models were individualised for each patient using nonlinear model-predictive control. The authors claimed that their approach could be implemented with routinely measured complete blood counts (CBC) and a few additional metabolite measurements. This would allow model-based individualised treatment, as opposed to a standard dose for all, and to prescribe an optimal dose for a desired outcome with minimum side-effects.

2.3.2 Mathematics of methotrexate treatments

Methotrexate (MTX) is an antimetabolite of the antifolate type. It is thought to affect cancer by inhibiting dihydrofolate reductase, an enzyme that participates in the tetrahydrofolate synthesis. This leads to an inhibitory effect on the synthesis of DNA, RNA, thymidylates, and proteins.

A first mathematical model of MTX effect in ALL was constructed in [90]. The authors based their approach on the fact that within cells, MTX is metabolised to more active methotrexate polyglutamates (MTXPG), and these polyglutamates are subsequently cleaved in lysosomes by glutamyl hydrolase (GGH). GGH acts as either an endopeptidase or an exopeptidase. To better define the in-vivo functions of GGH in human leukaemia cells, GGH activity was characterised with different MTXPG substrates in human T- and B-lineage leukaemia cell lines and primary cultures. Parameters estimated from fitting a series of hypothetical mathematical models to the data revealed that the experimental data were best fitted by a model where GGH simultaneously cleaved multiple glutamyl residues, with the highest activity on cleaving the outermost or two outermost residues from a polyglutamate chain. The model also revealed that GGH has a higher affinity for longer chain polyglutamates.

Further research led to the development of an improved model in [91]:

$$\frac{dMTX}{dt} = -(k_e + k_{12})MTX + k_{21}MTX_p, \tag{2.18a}$$

$$\frac{dMTX_p}{dt} = k_{12}MTX - k_{21}MTX_p, \tag{2.18b}$$

$$\begin{aligned}\frac{dMTXPG_1}{dt} =& \frac{V_{\max-in}MTX/V}{K_{m-in} + MTX/V} + k_pMTX/V - \\ &- k_{eff}MTXPG_1 + k_{GGH}MTXPG_{2-7} - \\ &- \frac{V_{\max-FPGS}MTXPG_1}{K_{m-FPGS} + MTXPG_1},\end{aligned} \tag{2.18c}$$

$$\begin{aligned}\frac{dMTXPG_{2-7}}{dt} =& \frac{V_{\max-FPGS}MTXPG_1}{K_{m-FPGS} + MTXPG_1} - \\ &- k_{GGH}MTXPG_{2-7}.\end{aligned} \tag{2.18d}$$

This latter model simulated the concentration of MTXPG_i, where the subscripts denoted the number of glutamates attached to each MTX molecule. This provided new insights into the intracellular disposition of MTX in leukaemic cells and how it affects treatment efficacy. The variables MTX and MTX_p denoted the central and peripheral compartments of MTX. The parameters described: an elimination of plasma (k_e); transition between peripheral and central compartments of MTX (k_{12}, k_{21}); systemic volume (V); influx of MTX into the leukaemic blasts ($V_{\max-in}$, K_{m-in}); first order influx and efflux (k_p and k_{eff}, respectively); FPGS activity ($V_{\max-FPGS}$, K_{m-FPGS}) and γ-glutamyl hydrolase activity (k_{GGH}). Data from 791 plasma samples from 194 patients were used to validate the model. The study of the mathematical equations revealed that GGH activity had a higher affinity for longer chain polyglutamates and FPGS activity was higher in B-lineage ALL in comparison to T-lineage ALL.

Finally, [92] constructed a model involving a combination of several drugs, for chemotherapy-induced leukopenia in paediatric ALL patients. The model accounted for the action of both 6-MP and MTX and their cytotoxic metabolites 6-TGNc and MTXPGs during maintenance therapy. The equations were built on the basis of the previously discussed models [89, 91]. The model predicted WBC counts for the available patient data surprisingly well, given the large variation of individual response patterns in the clinical data. The mathematical model and algorithmic procedure proposed could be used to guide personalised clinical decision support in childhood ALL maintenance therapy. Another model based on Refs. [89, 91] gave rise to a compartmental model in [93], including pharmacokinetics and a myelosuppression model for ALL, considering both 6-MP and MTX. The model was cross-validated with data from 116 patients, and simulations of different treatment protocols were performed to exploit the

optimal effect of maintenance therapy on survival.

2.4 Modelling immune response and immunotherapy in leukaemias

2.4.1 Immune response mathematical models

Immunotherapy is a type of therapy that stimulates cells within the immune system in order to help the body fight against cancer or infections. Interactions between cells are key in understanding processes such as, for example, proliferation or resource competition between cells. The immune system is one way in which the body may influence external agents and a greater understanding of it could be useful in fighting leukaemia.

An extension of the model already described, (2.11), was introduced in [94], where the CML populations were distributed as stem cells (y_0), progenitor (y_1) and mature leukaemic cells (y_2). In this study, the concentration of immune cells was also included and denoted as z. The authors designed a mathematical model integrating CML and an autologous immune response to the patients' data by considering the following system

$$\frac{dy_0}{dt} = b_1 y_1 - a_0 y_0 - \frac{\mu y_0 z}{1+\varepsilon y_3^2}, \tag{2.19a}$$

$$\frac{dy_1}{dt} = a_0 y_0 - b_1 y_1 + r y_1 \left(1 - \frac{y_1}{K}\right) - d_1 y_1 - \frac{\mu y_1 z}{1+\varepsilon y_3^2}, \tag{2.19b}$$

$$\frac{dy_2}{dt} = a_1 y_1 - d_2 y_2 - \frac{\mu y_2 z}{1+\varepsilon y_3^2}, \tag{2.19c}$$

$$\frac{dy_3}{dt} = a_2 y_2 - d_3 y_3 - \frac{\mu y_3 z}{1+\varepsilon y_3^2}, \tag{2.19d}$$

$$\frac{dz}{dt} = s_z - d_z z + \frac{\alpha y_3 z}{1+\varepsilon y_3^2}, \tag{2.19e}$$

where a_0, b_1 represents transition terms; d_z and d_i, for each cell type $i = 1, 2, 3$, denotes cell death; and a logistic growth for progenitor cells y_1 was included, with a reproduction rate r. The immune system action rate μ was included in the mass action term "$\mu\, y_i\, z$" in the last term of the leukaemic population equations from Eq. (2.19a) to Eq. (2.19d). The proliferation of the immune system pool included a constant factor s_z and was activated by mature leukaemia cells with the term "$\alpha\, y_3\, z$" in Eq. (2.19e). These latter terms included an inhibition of

the immune cells expansion, as they were divided by "$1 + \epsilon y_3^2$", where ϵ was the strength of the immunosuppression. This model included data from patients treated with imatinib, and their BCR-ABL transcripts, related to leukaemia diagnosis. The authors considered that variations in BCR–ABL transcripts during imatinib therapy may represent a signature of the patient's individual autologous immune response. The use of immunotherapy was then considered to be a useful complement to the usual treatment, playing a significant role in eliminating the residual leukaemic burden.

A general mathematical model for tumour immune resistance and drug therapy was proposed in [95]. By including tumour cells, immune cells, host cells and drug interaction, an optimal control problem was constructed. This would provide a basis for the study of leukaemia immune cell interaction, shedding some light on the modelling for B leukaemia. For B-cells, fundamental in both acute and chronic lymphocytic leukaemia diagnosis, a more extensive model was presented in [96], including four different cell populations in the peripheral blood of humans: B cells, able to bind to antigens which will initiate antibody responses; NK cells, critical to the immune system; cytotoxic T cells, able to kill cancer cells; and helper T cells, which may help other immune cells by releasing T cell cytokines. This model was considered a tool that may shed light on factors affecting the course of disease progression in patients, and focused on sensitivity analysis for parameters and bifurcation analysis. Based on [95], an immunotherapy approach was considered in [97] by developing a model focused on B and T lymphocytes and their relation with a chemotherapeutic agent. The ODE system for this model was the following

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{k}\right) - c_1 NI - \frac{\mu NQ}{a+Q}, \tag{2.20a}$$

$$\frac{dI}{dt} = s(t) + s_0 - dI + \frac{\rho NI}{\gamma + N} - c_2 NI - \frac{\delta IQ}{b+Q}, \tag{2.20b}$$

$$\frac{dQ}{dt} = q(t) - \lambda Q, \tag{2.20c}$$

where $N = N(t)$ represented the neoplastic B lymphocytes, $I = I(t)$ the healthy T lymphocytes (this is, the immune cells), and $Q = Q(t)$ the amount of a chemotherapeutic agent in the bloodstream. In Eq. (2.20a) N follows a logistic growth with a proliferation rate r, and dies due to both interaction with immune cells at a rate c_1 and with the chemotherapeutic agent at a rate μ. Immune cells in Eq. (2.20b) have a constant source s_0 and die naturally at a constant rate d and also due to interaction with cancer cells at a rate c_2, and with drugs

at a rate δ. However, there is a production rate ρ of immune cells stimulated by cancer cells. Both N and I have Michaelis-Menten terms with rates a, γ and b. For the case of the chemotherapeutic agent Q in Eq. (2.20c), λ is considered as the washout rate of a given cycle-nonspecific chemotherapeutic drug with $\lambda = \ln(2)/t_{\frac{1}{2}}$, where $t_{\frac{1}{2}}$ is the drug elimination half-life. Finally, the functions $s(t)$ and $q(t)$ are source terms, which can be considered to be constants. These parameters were all taken from the literature and claimed to simulate CLL behaviour. This model reinforces the option of combining treatments such as chemo- and immunotherapy, where the first may decrease cells to a point where immune cells may act.

A model for AML was considered in [98] by including the role of leukaemic blast cells (L), mature regulatory T cells (T_{reg}) and mature effector T cells (T_{eff}), this last also including cytotoxic T lymphocytes and Natural Killers. The aim of including such cells was to create an activated immune cell infusion with selective T_{reg} depletion. This was done by converting the intracellular interaction into a model, as the following system:

$$\frac{d[L]}{dt} = a_L \left(\frac{k_1^p}{k_1^p + [T_{\text{eff}}]^p} \right) - d_L[L], \tag{2.21a}$$

$$\frac{d\,[T_{\text{eff}}]}{dt} = a_{T_{\text{eff}}} \left(\frac{k_2^p}{k_2^p + [r_{\text{reg}}]^p} \right) - d_{T_{\text{eff}}}\,[T_{\text{eff}}]\,, \tag{2.21b}$$

$$\frac{d\,[T_{\text{reg}}]}{dt} = a_{T_{\text{reg}}} \left(\frac{[L]^p}{k_3^p + [L]^p} \right) - d_{T_{\text{reg}}}\,[T_{\text{reg}}]\,, \tag{2.21c}$$

where a_L, $a_{T_{\text{eff}}}$, $a_{T_{\text{reg}}}$ represented influx rates, and d_L, $d_{T_{\text{eff}}}$, $d_{T_{\text{reg}}}$ the decay rates. Intercellular interactions were modelled as Hill functions with threshold constants (k_1, k_2, k_3) with strength p. Two existing steady states were found for this model in [98], corresponding to leukaemia diagnosis or relapse, and to complete remission. The authors considered that the model explained the influence of the duration of complete remission on the survival of patients with AML after allogeneic stem cell transplantation. In [99], simulations were run for this model by performing Monte Carlo simulation of trajectories in the phase plane, and generated relapse-free survival curves, which were then compared with clinical data. This provided valuable information for the future design of immunotherapy in AML.

2.4.2 Including interleukins in mathematical models

Interleukins (ILs) are a group of cytokines first seen to be expressed by white blood cells (leukocytes). The immune system depends on interleukins as these signals between cells are useful for acting against several pathogens.

The interaction between the actively responding effector cells $E = E(t)$, tumour cells ($T = T(t)$) and the concentration of the cytokine IL-2 ($I_L = I_L(t)$) was the basis for the latter study, influenced by [100]. The reason behind the modelling of this cytokine is due to the fact that IL-2 might boost the immune system to fight tumours. This was described via the following system:

$$\frac{dE}{dt} = cT - \mu_2 E + \frac{p_1\, E\, I_L}{g_1 + I_L} + s_1, \tag{2.22a}$$

$$\frac{dT}{dt} = r_2(T) - \frac{a\, E\, T}{g_2 + T}, \tag{2.22b}$$

$$\frac{dI_L}{dt} = \frac{p_2\, E\, T}{g_3 + T} - \mu_3 I_L + s_2. \tag{2.22c}$$

In this model, c was antigenicity or ability to provoke an immune response, $\frac{1}{\mu_2}$ was the average natural lifespan, a the loss of tumour cells by interaction, μ_3 the degraded rate of IL-2, and s_1, s_2 were treatment terms. The fraction terms were of the Michaelis-Mentis form, to indicate saturation effects. The function $r_2(T)$ could be described as a constant for linear growth, or with limiting-growth as logistic or Gompertz terms. With this model, the authors concluded that with only IL-2 treatment, the immune system might not be enough to clear tumours. These and other models were reviewed in [101] in terms of equilibrium points, considering T lymphocytes and their interaction with other cells, and it was found that there are two stable equilibrium points, one where there is no tumour, and the other where there is a large one.

Interaction between cells via interleukins was also studied in [102], as IL-21 is being developed as an immunotherapeutic cancer drug. Its effect has been studied in relation to Natural Killer (NK) cells, and $CD8^+$ T-cells, which have the ability to make cytokines, with the model

$$\frac{du}{dt} = \text{input} - \mu_1 u, \tag{2.23a}$$

$$\frac{dx}{dt} = r_1 x \left(1 - \frac{x}{h_1(u)}\right), \tag{2.23b}$$

$$\frac{dy}{dt} = r_2 y \left(1 - \frac{y}{h_2(m)}\right), \tag{2.23c}$$

$$\frac{dm}{dt} = au - \mu_2 m, \tag{2.23d}$$

$$\frac{dp}{dt} = \frac{b_1 u}{b_2 + u} - \mu_3 p, \tag{2.23e}$$

$$\frac{dn}{dt} = g(n) - k_1 pxn - k_2 pym, \tag{2.23f}$$

where Eq. (2.23a) represented the concentration of IL-21, Eq. (2.23b) the concentration of NK in the spleen, Eq. (2.23c) the antitumour $CD8^+$ T-cells in the lymph, Eq. (2.23d) a facilitating T-cell memory factor useful for expressing the recognition of foreign invaders for memory T-cells, Eq. (2.23e) a cytotoxic protein affecting tumour lysis, and finally tumour mass at any time was represented by Eq. (2.23f). The functions involved were defined in the monotonic decreasing function

$$h_1(u) = \frac{p_1 u + p_2}{u + q_1}, \tag{2.23g}$$

the function of the memory factor m

$$h_2(m) = h_2(0) + \frac{\sigma m}{1 + \frac{m}{D}}, \tag{2.23h}$$

and $g(n)$ the dynamics of tumour cell number, which is constructed separately for each tumour type according to the observed growth curves. Parameters were estimated in terms of certain values from the literature, so that simulations were run to show IL-21 as a promising antitumour therapeutic. For more immunotherapeutic approaches towards cancer modelling, we highlight the work in [103], where some general aspects of cancer were also reviewed, including diffusion, angiogenesis and invasion.

Finally, for the case of immune response to leukaemia, other studies have been undertaken, though not specially in the form of an ODE or PDE system. Some numerical simulations were run in [104] by proposing an integro-differential equation model. This study proposed a new possibility for defining the activation states for cancer, cytotoxic T and T helper cells. Using these definitions, the authors suggested that it would be easier to organise experiments suitable for measuring cell states. They also claimed that cell-mediated immunity is one of the most crucial components of antitumour immunity. Immune T-cells were studied in [105] in terms of a stochastic model from which was derived a Fokker-Planck equation. Stability analysis and behaviour of the solutions of the model led to the conclusion that more accurate simulations of cancer genesis

and treatment were needed. Lastly, in [106], cytotoxic T cells were dynamically and structurally analysed in terms of a Boolean network model for T cell large granular lymphocyte leukaemia. Nineteen potential therapeutic targets were found, and these were versatile enough to be applicable to a wide variety of signals and regulatory networks related to diseases.

2.4.3 Novel therapies for leukaemia models: CAR-T cells

Immunotherapy based on chimeric antigen receptor T (CAR-T) cells has been especially successful in patients who did not respond to the usual types of chemotherapy. This technique is based on the patient's own T-cells, which are extracted from them, genetically modified and reinfused. This modification allows T-cells to kill tumour cells in a more effective way than the usual chemotherapies.

We have designed a general model for CAR-T cells in [107] considering several cell compartments. Firstly, for T-cell leukaemia, the number of CAR-T cells was denoted by C, leukaemic T cells by L, and normal T-cells by T. The dynamics of the model were as follows

$$\frac{dC}{dt} = \rho_C(T + L + C)C - \frac{1}{\tau_C}C - \alpha C^2 + \rho_I C, \tag{2.24a}$$

$$\frac{dL}{dt} = \rho_L L - \alpha LC, \tag{2.24b}$$

$$\frac{dT}{dt} = g(T, L, C) - \alpha TC. \tag{2.24c}$$

The parameter ρ_L represents leukaemic proliferation rate, while ρ_C represents stimulation of CAR-T cell mitosis after encounters with target cells; τ_C is the finite lifespan of CAR-T cells; the parameter α represents death due to encounters with CAR-T cells; parameter ρ_I is the external cytokine signal strength used for division of CAR-T cells; finally, the function $g(T, L, C)$ denotes the rate of production of normal T cells, assumed to contribute only at a minimal residual level. The stability analysis of the cell dynamics leads to several conclusions: firstly, CAR-T cells allow for control of T-cell leukaemia in the presence of fratricide; secondly, the initial number of CAR-T cells injected, as well as re-injections, does not affect the outcome of therapy, while higher mitotic stimulation rates do; lastly, tumour proliferation rates have an impact on relapse time. A second, similar model was constructed for B cells, in [108], where CAR-T and now leukaemic B cells where again denoted as C and L, but

the inclusion of mature healthy B cells B, CD19$^-$ B cells P, and CD19$^+$ cells I was considered. The initial autonomous system of differential equations was

$$\frac{dC}{dt} = \rho_C(L+B)C + \rho_\beta IC - \frac{1}{\tau_C}C, \tag{2.25a}$$

$$\frac{dL}{dt} = \rho_L L - \alpha LC, \tag{2.25b}$$

$$\frac{dB}{dt} = \frac{1}{\tau_I}I - \alpha BC - \frac{1}{\tau_B}B, \tag{2.25c}$$

$$\frac{dP}{dt} = \rho_P\left(2a_P s(t) - 1\right)P - \frac{1}{\tau_P}P, \tag{2.25d}$$

$$\frac{dI}{dt} = \rho_I\left(2a_I s(t) - 1\right)I - \frac{1}{\tau_I}I + \frac{1}{\tau_P}P - \alpha\beta IC, \tag{2.25e}$$

where parameters ρ_C, τ_C, ρ_L and α were the same as the considered in the previous model. Parameter $\rho_B = \beta\rho_C$, where $0 < \beta < 1$, accounts for the fact that represented B cells are located mostly in the bone marrow and encounters with CAR-T cells will be less frequent. Parameters ρ_P and ρ_I represent growth rates for P and I cells, while τ_I, τ_B and τ_P represent the finite lifespan of I, B and P cells respectively. A signalling function $s(t) = 1/[1 + k_s(P + I)]$, with $k_s > 0$ was constructed as in [32], also including the asymmetric division rates a_P and a_I for P and I. This general model is reduced, in order to understand the dynamics of the expansion of CAR-T cells and their effect on the healthy B and leukaemic cells, neglecting the contribution of the haematopoietic compartments. Parameters are estimated from the literature and the main conclusion obtained is that not only does CAR-T cell persistence depend on T-cell mean lifetime, but also that reinjection may allow the severity of relapse to be controlled. The dynamics of the model from Eq. (2.25) are summarised in Figure 2.10.

A general model taken from the literature and applied to CAR-T cells is set out in [109]. The authors denote $s = s(t)$ as the population of susceptible blood cells, $i = i(t)$ as the population of infected blood cells, $c = c(t)$ as the population of leukaemic cells (abnormal cells), and $w = w(t)$ as the population of white blood cells or immune cells. The dynamics are modelled as

$$\frac{ds}{dt} = A - a_0 s - \beta sc, \tag{2.26a}$$

$$\frac{di}{dt} = \beta sc - \beta_0 i - \beta_1 ci, \tag{2.26b}$$

$$\frac{dc}{dt} = k - k_0 c - k_1 cw, \tag{2.26c}$$

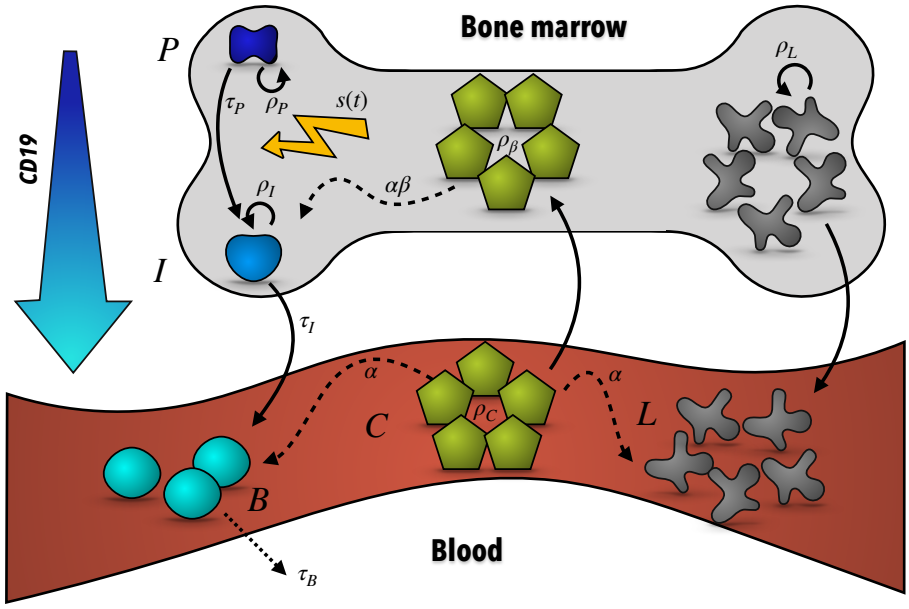


Figure 2.10: **Illustration of the dynamics in model** (2.25). B cells (in blue) develop in the bone marrow, arising from progenitor CD19$^-$ cells (P), then turning, with rate τ_P, into CD19$^+$ cells I and reaching, with rate τ_I a mature stage of healthy B cells B, finally dying after a time τ_B. During this process, a signalling effect $s(t)$ affects the proliferation rates of the early stages ρ_P and ρ_I. Leukaemic cells L develop in the bone marrow with rate ρ_L, invading this tissue as well as the blood compartment. CAR-T cells C attack mature B cells and leukaemic cells with rate α, also inducing growth, with rate ρ_C. In the bone marrow, they also attack CD19$^+$ cells I, with a lower rate $\alpha\beta$. This interaction induces growth with rate ρ_β. Solid lines represent cell growth and change between compartments. The dotted line represents the natural death of the healthy B cells. Dashed lines represent cell death due to CAR-T cell interaction.

$$\frac{dw}{dt} = B + bc - b_0 w - b_1 wc, \tag{2.26d}$$

where a_0, β_0, k_0, and b_0 are the natural death rate of susceptible blood cells, infected cells, cancer cells, and immune cells, respectively; for susceptible cells, A is the recruitment rate and β is the loss rate of susceptible blood cells due to infection; β_1 is the decay rate parameter of infected cells; k is the constant recruitment rate of cancer cells, while k_1 and b_1 are the loss rates of cancer and immune cells due to interaction; finally, parameter B is considered as the external re-infusion rate of immune cells (CAR-T). This model was studied in terms of stability, and it was observed that the external re-infusion of immune cells by adoptive T-cell therapy reduces the concentration of cancer cells and infected cells in the blood.

With the success of T-cell-engaging immunotherapeutic agents, there has been growing interest in the so-called cytokine release syndrome (CRS), as it represents one of the most frequent serious adverse effects of these therapies. CRS is a systemic inflammatory response that can be caused by a variety of factors, such as infections and certain drugs. A more specific model that included the action of cytokines was studied by considering Tisagenlecleucel, a personalised cellular therapy of CAR-T cells for B-cell ALLs, associated with a high remission rate. It was modelled in [110] by considering the interaction of a CAR-T cell population $c_T = c_T(t)$ with B-cell leukaemic population $l = l(t)$, as well as with healthy B cells $h = h(t)$, both marked with CD19, a characteristic of B lymphocytes. Other circulating lymphocytes were denoted as $c = c(t)$, while the number of cytokines, key to understanding inflammatory processes, was generally considered as $s = s(t)$. The dynamics of the model were as follows:

$$\frac{dc_T}{dt} = d_1 c_T - d_2 c_T - \alpha_1 c_T l - \beta_1 c_T h, \tag{2.27a}$$

$$\frac{di}{dt} = kl - \alpha_2 c_T l, \tag{2.27b}$$

$$\frac{dh}{dt} = ah(1 - bh) - d_3 h - \beta_2 c_T h, \tag{2.27c}$$

$$\frac{dc}{dt} = \lambda - \sigma c + \alpha_3 \frac{c_T c}{\beta_3 + c_T}, \tag{2.27d}$$

$$\frac{ds}{dt} = \alpha_4 - \beta_4 s + d_4 \left(\frac{c_T}{c_T + m} \right), \tag{2.27e}$$

where Eq. (2.27a) represented the dynamics of CAR-T cells with growth rate d_1 and natural death rate d_2, while α_1 and β_1 were cell death given by interaction with leukaemic and healthy cells, respectively. Eq. (2.27b) includes a growth

rate of leukaemic cells k and a cell death α_2 by interaction with c_T. Eq. (2.27c) described a logistic growth of healthy cells with rates a and b, as well as a natural death rate d_3 and death β_2 due to interaction with c_T. Circulating lymphocyte dynamics were considered in Eq. (2.27d) to have a constant input λ, death rate σ and growth dependant on c_T, attenuated via a Hill function with constants α_3 and β_3. Finally, for Eq. (2.27e), cytokines were secreted at a maximum rate α_4 and altered by a negative feedback mechanism corresponding to the term $-\beta_4 s$. Furthermore, the stimulation of CAR-T cells increased the levels of cytokines with rate d_4 and a constant m from the correspondent Hill function. Optimal control theory was applied for this model, controlling the injection of CAR-T cells and cytokines, to finally minimise the level of cancer cells and to keep healthy cells above a desired level.

Effector T cells are a group of cells including several T-cell types that actively respond to a stimulus. Following an infection, memory T cells are antigen-specific T cells that remain in the long term. This distinction is considered to help understand the dynamics of CAR-T cells in several models. For instance, a general description of Tisagenlecleucel was performed in [111], where data from 91 paediatric and young adult B-ALL patients were used for the analysis. The model describes the expansion of CAR-T cells up to a time $T_{\max}$, and then two phases: a first contraction phase, with rapid decline; and a second persistence phase, declining more gradually. This was represented by a dynamic system considering effector E and memory CAR-T cells M, as

$$\frac{dE}{dt} = \rho \cdot F(t) \cdot E, \quad \text{for } T \leq T_{\max}, \tag{2.28a}$$

$$\frac{dE}{dt} = -\alpha \cdot E, \quad \text{for } T > T_{\max}, \tag{2.28b}$$

$$\frac{dM}{dt} = k \cdot E - \beta \cdot M, \quad \text{for } T > T_{\max}, \tag{2.28c}$$

and $M = 0$, for $T \leq T_{\max}$. After $T_{\max}$, effector cells rapidly decline at a rate α and convert to memory cells at a rate k, which decline at a rate β. However, before $T_{\max}$, only effector cells grow at a rate ρ and proportionally to a function $F(t)$ which simulates the inclusion with step-wise functions of the co-medication of corticosteroids and tocilizumab (anti–IL-6 receptor antibody). This simple model was able to show the long-term persistence used in CAR-T therapies.

The authors in [112] also considered a division between tumour T, effector CAR-T cells C_T and memory CAR-T cells C_M in the following model

$$\frac{dT}{dt} = T f(T) - d_T\left(T, C_T\right), \tag{2.29a}$$

$$\frac{dC_T}{dt} = p_{C_T}(C_T) - a_{C_T}(C_T) + \tag{2.29b}$$
$$+ p_{C_T}(T, C_M) - d_{C_T}(T, C_T),$$
$$\frac{dC_M}{dt} = p_{C_M}(C_M) - d_{C_M}(T, C_M) - a_{C_M}(C_M), \tag{2.29c}$$

where $f(T)$ is the density dependence growth of tumour cells, and respectively for effector and memory CAR-T cells, we have the following: $p_{C_T}(C_T)$ and $p_{C_M}(C_M)$ as cell production functions, $d_{C_M}(T, C_M)$ and $d_{C_T}(T, C_T)$ as cell inhibition functions, and $a_{C_M}(C_M)$ and $a_{C_T}(C_T)$ as natural death functions. For this model, most functions were considered to be linear, except for the tumour growth function, considered to be logistic growth. Simulations were run for mice data found in the literature, showing different outcomes depending on tumour burden or initial therapy dose. The authors considered that a high CAR-T cell inhibition from tumour leads to tumour escape and absence of CAR-T cell memory. The same CAR-T cell division was considered in the model from [113], not only showing a distinction between effector and memory, but also between the cytotoxic ($CD8^+$) and helper ($CD4^+$) cells. Again, parameter values were not obtained from actual data, but from simulated clinical data. Their results suggest the hypothesis that initial tumour burden is a stronger predictor of toxicity than the initial dose of CAR-T cells. Also, the authors considered an inflammatory immune response regulated via a Hill function to maintain a realistic bound on the activation rate of T cells. This function gave rise to tumour-burden-correlated toxicity, while the correlation of CAR-T cell dose alone and toxicity was poor.

The pharmacological model in [114] considered both the influence of CAR-T cells in inflammatory responses with cytokines (such as interleukins IL_6, IL_{10} or interferon IFN_γ), as well as the distinction between CAR-T cells into effector and memory cells. This was also done in order to understand toxicity related to cytokine release syndrome. In the model, the variable B represents CLL tumour B cells in peripheral blood (PB). CAR-T cells in PB are divided into effector E_{PB} and memory M_{PB} cells. This division is also performed for the CAR-T cells in the tissue compartments (E_T and M_T). The complete mathematical model is shown in Figure 2.11, and reads

$$\frac{dB_{PB}}{dt} = r_B B_{PB} - d_B B_P - K_{BC} E_{PB} B_{PB}, \tag{2.30a}$$
$$\frac{dIL_6}{dt} = \rho_{\text{endo IL}_6} + \rho_{\max \text{IL}_6} B_{PB} E_P - d_{\text{IL}_6} IL_6, \tag{2.30b}$$

$$\frac{dIL_{10}}{dt} = \rho_{\text{endo IL}_{10}} + \rho_{\max \text{IL}_{10}} B_P E_P - d_{\text{IL}_{10}} \text{IL}_{10}, \tag{2.30c}$$

$$\frac{dIFN_\gamma}{dt} = \rho_{\text{endo IFN}_\gamma} - d_{\text{IFN}_\gamma} IFN_\gamma + \tag{2.30d}$$
$$+ \rho_{\max \text{IFN}_\gamma} B_{PB} E_{PB} \left(a_G + (1 - a_G) \frac{b_G}{IL_{10} + b_G} \right),$$

$$\frac{dE_{PB}}{dt} = D_{\text{inj}} + r_E E_{PB} B_{PB} - d_E E_{PB} - k_{in} E_{PB} + \tag{2.30e}$$
$$+ k_{\text{out}} E_T - a_E E_{PB} (1 - f(B_{PB})) + a_M M_{PB} f(B),$$

$$\frac{dE_T}{dt} = k_{\text{in}} E_{PB} - k_{\text{out}} E_T, \tag{2.30f}$$

$$\frac{dM_{PB}}{dt} = r_M M_{PB} - d_M M_{PB} - k_{\text{in}} M_{PB} + k_{\text{out}} M_T + \tag{2.30g}$$
$$+ a_E M_{PB} (1 - f(B_{PB})) - a_M M_{PB} f(B_{PB}),$$

$$\frac{dM_{\text{T}}}{dt} = k_{\text{in}} M_{PB} - k_{\text{out}} M_T. \tag{2.30h}$$

In this model, parameters r_B, r_E and r_M represent growth rates, while d_B, d_E and d_M are death rate constants, respectively for B, E_{PB} and M_{PB} cells. Parameter K_{BC} is the is the effector CAR-T-mediated B-cell CLL degradation rate constant in peripheral blood. For the inflammatory immune responses we have, respectively for IL_6, IL_{10} and IFN_γ the following constants: $\rho_{\text{endo IL}_6}, \rho_{\text{endo IL}_{10}}$ and $\rho_{\text{endo IFN}_\gamma}$ as endogenous synthesis rates; parameters $\rho_{\max \text{IL}_6}, \rho_{\max \text{IL}_{10}}$ and $\rho_{\max \text{IFN}_\gamma}$ as production rates; and finally, $d_{\text{IL}_6}, d_{\text{IL}_{10}}$ and d_{IFN_γ} are the natural death rates by the activated CAR-T cells. Constants a_G and b_G are the inhibitory parameters of IL_{10} on IFN_γ production. PB and tissue compartments are distributed via rate constants k_{in} and k_{out} after intravenous infusion. Peripheral blood effector memory CAR-T cells are activated via activation rates a_E and a_M. Finally, function $f(B_{PB})$ is chosen as a Hill function such that $f(B_{PB}) = \frac{B_{PB}}{B_{PB}+h}$, with h the half-saturation constant of the tumour. This model was adjusted to data from 3 patients obtained from the literature. Its main conclusion is that toxic inflammatory response is correlated to disease burden, i.e. the number of tumour cells in bone marrow, and not with CAR-T cells doses, contrary to what is observed with most cancer chemotherapies. Other models have also considered these hypotheses, such as the discretised model in [115] for CAR-T cells. In this study, a logistic equation of growth was considered to explain the interaction between CAR-T cells and malignant tumour cells. The binding affinity of the CAR-T cell construct

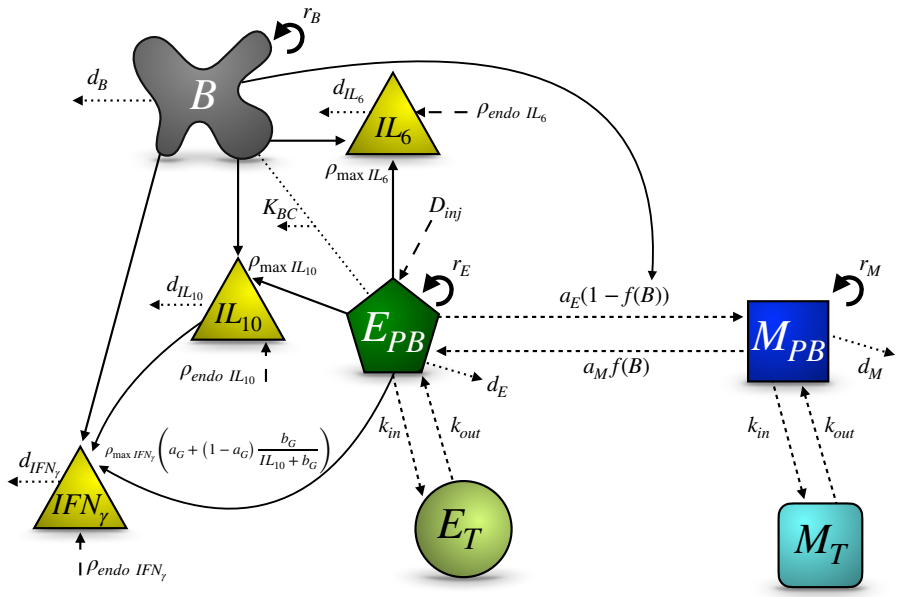


Figure 2.11: **Illustration of the dynamics in model** (2.30). The grey irregular shape represents tumour B-cells in CLL. Yellow triangles represent the inflammatory cytokines $IL_6 = IL_6(t)$, $IL_{10} = IL_{10}(t)$ and interferon $IFN_\gamma = IFN_\gamma(t)$. The green circle and pentagon represent respectively effector CAR-T cells from the PB and from the tissue. Squared, blue shapes represent memory CAR-T cells, also from PB and tissue. Solid lines represent promotion of cell production, while dotted lines represent cell loss due to natural death or due to encounters between cells. Short-dashed lines represent exchange between cell compartments, and finally long-dashed lines represent constant production in the cell compartments.

(the so-called single-chain variable fragment) and the antigenic epitope (the molecule binding to the antibody) on the malignant target was considered a critical parameter for all T-cell subtypes modelled. Both studies show the need for CAR-T cell doses to account for tumour burden, which would require a relatively low number of infused CAR-T cells to achieve the desired target.

2.5 Theoretical studies of leukaemia treatment models

In previous sections we have described leukaemia growth and response to therapy models that are careful to account for experimental facts or available data. There have been also many studies of models that focus their attention more on methodological mathematical aspects, and provide insight of a more fundamental type. For instance, some of them do not specify which type of leukaemia or treatment they describe.

For instance, some optimal control problems for general leukaemia treatment models have been discussed in the literature. In [116] the authors describe the dynamics of a healthy cell population $N(t)$, a leukaemic population $L(t)$ and a drug $h(t)$ governed by the equations

$$\frac{dL(t)}{dt} = r_l L(t) \ln\left(\frac{L_a}{L(t)}\right) - \gamma_l L(t) - f_l(h)L(t), \tag{2.31a}$$

$$\frac{dN(t)}{dt} = r_n N(t) \ln\left(\frac{N_a}{N(t)}\right) - \gamma_n N(t) - \frac{cN(t)L(t)}{1+L(t)} - f_n(h)N(t), \tag{2.31b}$$

$$\frac{dh(t)}{dt} = -\gamma_h h(t) + u(t), \tag{2.31c}$$

for $L(0) = L_0, N(0) = N_0, h(0) = 0$ and γ_h the drug dissipation rate. The effect of the drug was described differently for diseased and healthy cells by the therapy functions $f_l(h)$ and $f_n(h)$, respectively. Here, L_a and N_a were the maximum number of diseased and healthy cells respectively, and γ_l and γ_n were respectively the death rates for the two kinds of cells. Interaction between these subsets was expressed by the parameter c. Finally, the control function $u(t)$ is the quantity of drug given to the patient. The authors solved the optimal control problem using the Pontryagin maximum principle. Later research provided additional results along these lines in [117], by using a non-Gompertz interaction term and several phase constraints. Analysis of the switching points was performed, as well as several simulations. Some optimal therapy protocols are shown by introducing a 'shifting-variable', which avoids the violation of the normal cell constraint. Other studies have considered the combined effect of Haematopoietic Inducing Agents (HIA) and Chemotherapeutic Agents (CTA) on stem cells, with the goal of minimising leukopenia [118]. Proliferating (P)

and non-proliferating cells (N) were included in the model:

$$\frac{dP}{dt} = -\gamma P + \beta(N)N - \exp(-\gamma t)\beta(N_\tau)N_\tau + \\ + \beta_{HIA}(P)N - \beta_C(P)N, \quad \tau < t, \tag{2.32a}$$

$$\frac{dN}{dt} = -[\beta(N)N + \delta N] + 2\exp(-\gamma t)\beta(N_\tau)N_\tau - \\ - \beta_{HIA}(P)N + \beta_C(P)N, \quad \tau < t, \tag{2.32b}$$

for $t < \tau$, where τ was the time for a cell to complete one cycle of proliferation, γ the apoptosis rate, and δ the random cell loss. The expression N_τ stood for $N(t-\tau)$, introducing a time delay into the equation, and

$$\beta(N) = \beta_0 \frac{\theta^n}{\theta^n + N^n}; \tag{2.32c}$$

$$\beta_{HIA}(P) = \beta_{0,HIA} \frac{\theta_1^m}{\theta_1^m + P^m} g_{HIA}(t); \tag{2.32d}$$

$$\beta_C(P) = \beta_{0,C} \frac{P^w}{\theta_2^w + P^w} g_C(t); \tag{2.32e}$$

were Hill functions measuring the rate of cell re-entry into proliferation, the effect of HIA, and the effect of CTA on stem cells, respectively. Also,

$$g_{HIA}(t) = \begin{cases} 1, & 0 < t \leq \tau_1, \\ \exp(-s_1(t-\tau_1)), & t > \tau_1, \end{cases} \tag{2.32f}$$

simulated the time decay of HIA. Finally, CTA time decay was modelled by

$$g_C(t) = \begin{cases} 1, & 0 < t \leq \tau_2, \\ \exp(-s_2(t-\tau_2)), & \tau_2 < t \leq \tau_3, \\ \exp(-s_3(t-\tau_3)), & t > \tau_3. \end{cases} \tag{2.32g}$$

Using this set of equations the authors found that HIA administration increases the nadir observed in the proliferative cell line compared with when CTA treatment alone is administered. This is significant in preventing patients undergoing chemotherapy treatment from experiencing secondary effects. Furthermore, the steady state value of the proliferating cells was found to be significantly lower in silico after CTA treatment. The model and accompanying analysis give rise to an interesting question: Is concurrent administration of an HIA during chemotherapy a prudent approach for reducing toxicity during

chemotherapy? There is substantial clinical evidence to suggest that HIAs could be useful in cases of anemia. They argued that prophylactic benefits of HIAs use together with chemotherapeutic agents at the onset of treatment, although rational, should be balanced with the treatment cost and the risk that HIAs will cause adverse side effects such as venous thromboembolism and tumour progression.

2.6 Conclusion

Mathematical models have proved to be a essential asset in biomedicine. Haematological diseases are well suited to mathematical modelling, not only with differential equations, but also with stochastic models or other techniques. Therefore, there is a huge amount of data to combine with the mathematical models already in the current literature. Even so, these models may not be sufficient to characterise specific disease behaviours in leukaemia diagnosis: one could take, for example, acute lymphoblastic leukaemia dynamics as a particularly undeveloped issue, as studies of chronic myeloid leukaemia appear to us to have attracted more attention. This is probably because myeloid malignancies are most common in adults.

Despite the importance of the models presented, the only way to integrate them into clinical practice successfully is to through collaboration between mathematicians, biomedical scientists and clinicians. This can lead to new questions and conclusions for both mathematical models and biological problems. The development of such a useful weapon against cancer should be unified, so that the models can be useful for the actual observation and treatment of disease in patients, beyond the theoretical framework. Mathematical models require refinement in terms of being included in hospital protocols, as a diagnostic or prognostic tool and this can only be achieved by cooperation between the mathematical and medical world.

The review exposed in this Chapter can be found in Ref. [119].

CHAPTER 3

A mathematical model of B lymphopoiesis

Blood is a tissue under continuous regeneration, and its renovation is one of the most studied developmental processes in biology [120]. It is initiated by haematopoietic stem cells and develops through a multi-step differentiation cascade [7], resulting in the generation of red blood cells, platelets and cells of the immune system. Figure 3.1(A) shows the standard representation of haematopoiesis as a tree. At the top, stem cells with the potential for self-renewal give rise to respective lineage progenitors. These cells become progressively more specialised as they move towards the bottom of the tree. There are two major cell lineages: the myeloid line and the lymphoid line. The latter generates lymphocytes, involved in adaptive immune response, which is responsible for 'targeted' reactions to infections.

In this work, we will focus on the description of B lymphopoiesis, i.e. the development and maturation of B cells. These cells have a range of roles, being mainly associated with the secretion of antibodies, the elements in charge of the neutralisation of foreign invaders [121]. Figure 3.1(B) shows a schematic representation of the route from common lymphoid progenitor to immature B cells, which eventually exit the bone marrow to complete maturation elsewhere. Alterations in the generation of B cells are related to diseases like autoimmune reactions, immunodeficiencies or lymphoproliferative disorders like lymphomas or leukaemias [122]. The latter have especial incidence in children and constitute around one third of all childhood cancer cases [123].

Haematopoiesis is a perfect example of self-renewal and stemness in tissues [6]. Mathematical descriptions have been performed using multi-compartmental,

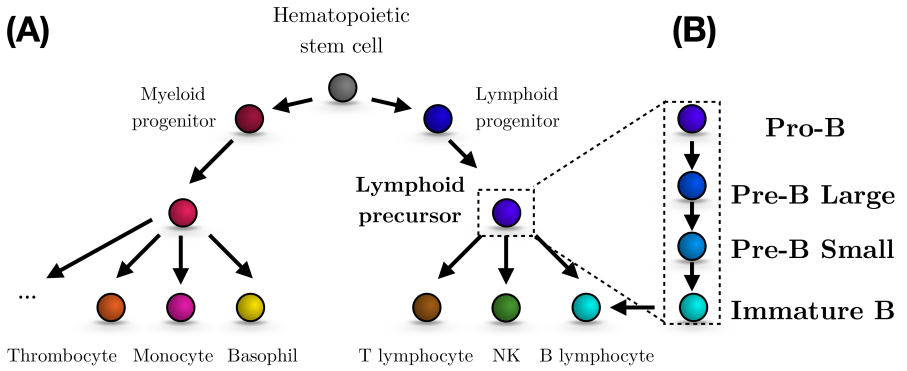


Figure 3.1: **Representation of haematopoiesis and B-cell lineage in study**. **(A)** Schematic representation of haematopoiesis. **(B)** B-cell lineage starts from a common lymphoid progenitor and then progresses towards immature B cells, which eventually leave the bone marrow. Hallmarks of this process are the acquisition of the immunophenotypic cell surface markers CD19 for the whole B line and CD10 for immature B cells.

continuously structured models [28, 38, 124]. In those models, each cell type in Figure 3.1(A) can be thought of as a cell compartment that receives input from the upper elements and sends its output to the lower compartments [125, 126, 127, 128, 129, 130, 131]. Some models have focused explicitly on the B-cell line [132, 133]. Unlike full haematopoiesis, this process is sequential, simplifying its mathematical conceptualisation. The compartmental models described above become nonlinear when the interactions between the different cell stages are included. Indeed, this process requires some kind of negative regulatory feedback in order to ensure steady production [134, 135]. This feature is common in the modelling of biological systems since they normally consist of a considerable number of interacting components [136].

In the case of B cells, their development depends on the joint action of a number of factors that support or inhibit B-cell growth and differentiation [137]. A clear description of the participants at each stage of development is lacking [122], which is in part due to the difficulties in recreating *in-vivo* conditions in experimental designs [138]. Mathematical models can help in elucidating which processes are more influenced by a given type of signal, as has been done for a number of lineages and scenarios [35, 139, 140, 141, 142], and a better

comprehension of these interactions can be useful in understanding progression to malignancies [143].

In this context, a complete picture of B lymphopoiesis would be a useful complement to those modelling scenarios that consider an input of B cells or that represent B-cell-related phenomena [108, 133, 144, 145, 146]. Precise knowledge of the dynamical behaviour of each stage of differentiation stage can also be of help in clinical situations where disorders are especially linked to the characteristics of the cell of origin. In leukaemias, for instance, the phenotype of the tumour cells is an important diagnostic criterion [147]. Our aim in this paper is thus to develop a model of B lymphopoiesis in the bone marrow. We will construct and investigate the properties and dynamical behaviour of a series of models. This will be complemented with data from the literature, which mainly comes from *in-vitro* assays and immune reconstitution studies, and with clinical data from haematological patients.

This chapter is structured as follows: In Sec. 3.1 we explain basic haematological models, going from a general model to a reduced family of models more suitable for lymphopoiesis. In Sec. 3.2 we perform a mathematical analysis of these models, including positivity, boundedness and stability. In Sec. 3.4 we carry out numerical simulations, taking into account clinical data and information from the literature. In Sec. 3.5 we discuss these results and examine the potential of these models to describe the process, concluding with the kind of research opportunities that this analysis paves the way for.

3.1 Mathematical models and methods

Our aim here is to describe B-cell development taking into account what the data can tell us about the structure of the population of this haematopoietic line. An illustration of the representation of cell development can be found in [32, 36], where the authors proposed n maturation stages for a cell population $u_n = u_n(t)$, with $t \in \mathbb{R}$ representing time. In this case, u_1 would represent stem cell population, u_n a mature specialised cell and u_i $(i = 2, \ldots, n-1)$ intermediate stages.

The model was studied in terms of proliferation rates $p_i = p_i(t)$ and the so-called self renewal fraction $a_i = a_i(t)$, for each maturation stage $i = 1, ..., n$. The latter is considered to be the probability of a cell remaining in the same cell compartment after mitosis. The authors assumed that cells at stage i enter mitosis with a rate p_i, resulting in a total number of $2p_iu_i$ after mitosis. Then, with probability a_i, they remain in the same compartment, whereas with

probability $1 - a_i$ they go on to the next maturation stage. Therefore, each cell compartment has an output of $p_i u_i$ and an input $2p_i a_i u_i$ from their own compartment. Consequently, the previous compartment (except at the first stage) provides the input corresponding to the number of cells that go on to the next maturation stage after mitosis: $2p_{i-1}(1 - a_{i-1})u_{i-1}$. Lastly, mature cells die at a rate d, and they are not considered to enter mitosis. If we consider $n = 3$ stages for stem cells (u_1), intermediate cells (u_2) and specialised cells (u_3), the result is the following system of equations:

$$\frac{du_1}{dt} = p_1(2a_1 - 1)u_1, \tag{3.1a}$$

$$\frac{du_2}{dt} = p_2(2a_2 - 1)u_2 + 2p_1(1 - a_1)u_1, \tag{3.1b}$$

$$\frac{du_3}{dt} = 2p_2(1 - a_2)u_2 - du_3. \tag{3.1c}$$

B cells are far from the haematopoietic stem cells since they are already committed to the B line, thus losing part of their potential for self-renewal. We then choose to specify cell behaviour in each compartment in terms of proliferation and maturation, i.e. progression to the next stage, and to restrict the system to three different cell compartments, considering the most common immunophenotypical characterisation used in clinical practice [18]. The first compartment would also receive input from previous lymphoid progenitors. However, this early compartment is smaller and thus a constant source term contribution would be less significant [148]. Furthermore, this input is also regulated, which would require adding an equation from the previous compartment, and similarly for even earlier compartments. Our aim was to restrict the analysis to the $CD19^+$ fraction of the B-cell line, as suggested by the data (see Section 3.3).

Thus we will consider three compartments accounting for the different maturation stages: early B cells ($C_1 = C_1(t)$), intermediate B cells ($C_2 = C_2(t)$), and finally late B cells ($C_3 = C_3(t)$), where $t \in \mathbb{R}$ represents time. A compartmental model can then be written as

$$\frac{dC_1}{dt} = \rho_1 C_1 - \alpha_1 C_1, \tag{3.2a}$$

$$\frac{dC_2}{dt} = \rho_2 C_2 + \alpha_1 C_1 - \alpha_2 C_2, \tag{3.2b}$$

$$\frac{dC_3}{dt} = \alpha_2 C_2 - \alpha_3 C_3. \tag{3.2c}$$

Note that this formulation is equivalent to the model from Eq. (3.1) with $u_i = C_i$ for $i = 1, 2, 3$ and parameters $\rho_i = p_i$, $\alpha_i = 2p_i(1 - a_i)$ for $i = 1, 2$, and $\alpha_3 = d$.

Early B cells, described here by Eq. (3.2a), have a proliferation rate ρ_1 and a transition rate into intermediate B cells of α_1. Analogously, the proliferation rate for intermediate B cells and transition rate into the late compartment are ρ_2 and α_2, as described in Eq. (3.2b). In this equation, a fraction of $\alpha_1 C_1$ cells comes from the early B compartment. This also happens with the $\alpha_2 C_2$ cells that change their phenotypes from the intermediate B-cell into the late B-cell compartment. Late B cells in Eq. (3.2c) are not considered to enter mitosis and they go into the blood flow with a blood transition rate α_3.

This compartmental model needs to be complemented with a regulatory system involving cell feedback signalling $s(t)$. Different types of signalling and the importance of the regulation of self-renewal in homeostatic (steady) state have been previously studied in the literature [35, 129, 134, 135, 139, 140, 141]. In our case, there are different ways of specifying which cells participate in signalling and through which processes. In this Chapter we consider two different hypotheses. The first is that signals are produced either by late cells (model 1) or by all cells (model 2). The second is that signals can alternatively affect either the proliferation rate (model A) or the transition rate of the model (model B). Therefore, we will consider four possibilities regarding feedback signalling. This is summarised in Figure 3.2.

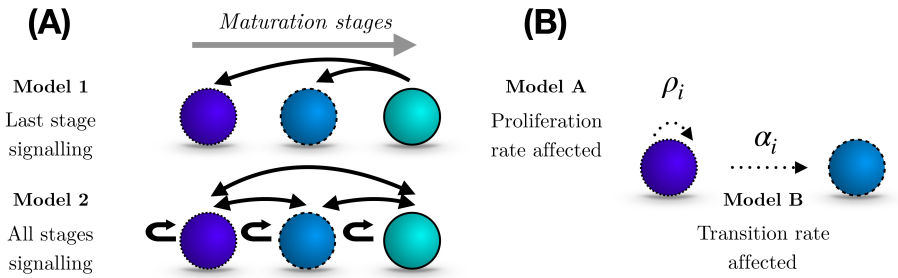


Figure 3.2: **Representation of the different feedback signalling possibilities**. **(A)** The signal may be produced by the most mature cells and then affect the previous compartments (model 1); or it may be produced by all cells, influencing the whole population (model 2). **(B)** Signalling can alternatively affect either the proliferation rate (model A) or the transition rate (model B).

As stated in the introduction, the precise number and role of interacting elements is unclear. The basic immune system chemical messengers are the cytokines, proteins that control cell production. They can be generated by microenvironmental elements but also from cells themselves [149]. For B cells, a number of cytokines have been shown to be relevant: IL-7 for proliferation, differentiation (transition) and survival [150], and SCF and FLT3LG [151] for proliferation of early stages [152]. Evidence in this regard comes mainly from *in-vitro* and murine models and differences with humans can be significant [138]. Due to this uncertainty and complexity we decided to follow an implicit formulation for the signalling as in [133], where instead of including each contributor explicitly we consider the systemic action of each, and gather them together in a single function $s(t)$.

With respect to the form of this signalling function $s(t)$, we follow the development set out in [32]. First, let us consider a maximal signal ρ_S, which is self-limited with rate α_S. Consider then a number of cells $N = N(t)$ that regulates the production of this signal in such a way that it decreases when there is a large number. Thus, a general signalling $S = S(t)$ can be modelled as

$$\frac{dS}{dt} = \rho_S - \alpha_S S - \beta SN, \tag{3.3}$$

where β is the inhibitory influence of cells N. By making a change of variables $s(t) = \alpha_S/\rho_S \cdot S(t)$ and $k = \beta/\alpha_S$, we obtain a differential equation whose quasi-steady state is $s(t) = 1/(1 + kN(t))$. A rigorous analysis of this quasi-steady state approximation can be found in [153]. Signal concentration depends on the number of cells of type N. Following the explanation in Figure 3.2 we will consider the case when only the late cells participate in signalling, i.e.

$$s_1(t) = \frac{1}{1 + kC_3}, \tag{3.4}$$

or the case where all cells participate

$$s_2(t) = \frac{1}{1 + k\sum_{i=1}^{3} C_i}. \tag{3.5}$$

Note that $s(t)$ is a decreasing function of the cell numbers C_i so it has an inhibitory role. The parameter k measures the strength of the inhibitory feedback.

To sum up, we will consider the models:

$$\frac{dC_1}{dt} = s_\rho \rho_1 C_1 - s_\alpha \alpha_1 C_1, \tag{3.6a}$$

$$\frac{dC_2}{dt} = s_\rho \rho_2 C_2 + s_\alpha \alpha_1 C_1 - s_\alpha \alpha_2 C_2, \tag{3.6b}$$

$$\frac{dC_3}{dt} = s_\alpha \alpha_2 C_2 - s_\alpha \alpha_3 C_3. \tag{3.6c}$$

Feedback signalling affects either proliferation (signal $s_\rho = s_\rho(t)$) or transition rates (signal $s_\alpha = s_\alpha(t)$). The form of the signals s_ρ, s_α depends on the signalling source, either $s_1(t)$ or $s_2(t)$. This yields four possible models, summarised in Table 3.1.

Signalling	1. Late cells	2. All Cells
A. Affecting proliferation	Model A1 ($s_\rho = s_1, s_\alpha = 1$)	Model A2 ($s_\rho = s_2, s_\alpha = 1$)
B. Affecting transition	Model B1 ($s_\rho = 1, s_\alpha = s_1$)	Model B2 ($s_\rho = 1, s_\alpha = s_2$)

Table 3.1: Mathematical models considered depending on signalling.

3.2 Theoretical results

3.2.1 Existence, boundedness and positivity of solutions

Models A can be written as

$$\frac{dC_1}{dt} = \frac{\rho_1 C_1}{1 + kN} - \alpha_1 C_1, \tag{3.7a}$$

$$\frac{dC_2}{dt} = \frac{\rho_2 C_2}{1 + kN} + \alpha_1 C_1 - \alpha_2 C_2, \tag{3.7b}$$

$$\frac{dC_3}{dt} = \alpha_2 C_2 - \alpha_3 C_3, \tag{3.7c}$$

and models B have the form

$$\frac{dC_1}{dt} = \rho_1 C_1 - \frac{\alpha_1 C_1}{1 + kN}, \tag{3.8a}$$

$$\frac{dC_2}{dt} = \rho_2 C_2 + \frac{\alpha_1 C_1}{1 + kN} - \frac{\alpha_2 C_2}{1 + kN}, \tag{3.8b}$$

$$\frac{dC_3}{dt} = \frac{\alpha_2 C_2}{1 + kN} - \frac{\alpha_3 C_3}{1 + kN}. \tag{3.8c}$$

Feedback signalling can depend on cells from the late stage with $N = C_3$ (models A1 and B1 with signal $s_1(t)$ as in Eq. (3.4)) or on all cells with $N = \sum_{i=1}^{3} C_i$ (models A2 and B2 with signal $s_2(t)$ as in Eq. (3.5)).

Theorem 3.2.1. *Let us consider the following set*

$$Q = \{(C_1, C_2, C_3) \in \mathbb{R}^3 : C_1, C_2, C_3 > 0\}, \tag{3.9}$$

and the initial values in Q

$$C_1(t_0) = C_1^0, \quad C_2(t_0) = C_2^0, \quad C_3(t_0) = C_3^0, \tag{3.10}$$

Then, the initial value problem for either Eqs. (3.7) *or Eqs.* (3.8) *has a unique local-in-time solution for each* $t \in [t_0 - \epsilon, t_0 + \epsilon]$*, for some value* $\epsilon > 0$*.*

Proof. The existence of a solution for systems from Eq. (3.7) and Eq. (3.8) is guaranteed for each $(C_1, C_2, C_3) \in Q$ by continuity of the functions

$$f_1^\rho = f_1^\rho(C_1, C_2, C_3) = \frac{\rho_1 C_1}{1 + kN} - \alpha_1 C_1, \tag{3.11a}$$

$$f_2^\rho = f_2^\rho(C_1, C_2, C_3) = \frac{\rho_2 C_2}{1 + kN} + \alpha_1 C_1 - \alpha_2 C_2, \tag{3.11b}$$

$$f_3^\rho = f_3^\rho(C_1, C_2, C_3) = \alpha_2 C_2 - \alpha_3 C_3, \tag{3.11c}$$

and

$$f_1^\alpha = f_1^\alpha(C_1, C_2, C_3) = \rho_1 C_1 - \frac{\alpha_1 C_1}{1 + kN}, \tag{3.12a}$$

$$f_2^\alpha = f_2^\alpha(C_1, C_2, C_3) = \rho_2 C_2 + \frac{\alpha_1 C_1}{1 + kN} - \frac{\alpha_2 C_2}{1 + kN}, \tag{3.12b}$$

$$f_3^\alpha = f_3^\alpha(C_1, C_2, C_3) = \frac{\alpha_2 C_2}{1 + kN} - \frac{\alpha_3 C_3}{1 + kN}, \tag{3.12c}$$

where again $N = C_3$ or $N = \sum_{i=1}^{3} C_i$. Boundedness of the respective partial derivatives of f_i^α and f_i^ρ for $i = 1, 2, 3$ proves that they satisfy the Lipschitz conditions, and therefore the solutions of systems from Eq. (3.7) and Eq. (3.8) with initial values as in Eq. (3.10) are unique by the Picard-Lindelöf theorem. ■

Henceforth we will consider that all parameters ρ_i, α_i and initial conditions C_i^0 are positive for $i = 1, 2, 3$.

Theorem 3.2.2. *The solutions of Eqs.* (3.7) *and Eqs.* (3.8) *with* $(C_1^0, C_2^0, C_3^0) \in Q$ *are positive.*

Proof. Let us consider the functions f_i^ρ, f_i^α with $i = 1, 2, 3$ from Eqs. (3.11) and Eqs. (3.12), respectively. As $f_i^\rho, f_i^\alpha > -\alpha_i C_i$, we know that

$$\frac{dC_i}{dt} > -\alpha_i C_i. \tag{3.13}$$

By integrating both sides of the equation from t_0 to t we obtain

$$C_i(t) > C_i^0 \exp(-\alpha_i t) > 0. \tag{3.14}$$

And therefore all solutions $C_i(t)$, $i = 1, 2, 3$ are positive over their domain of definition. ■

Theorem 3.2.3. *The solutions* $C_1(t), C_2(t), C_3(t)$ *of Eqs.* (3.7) *with* $(C_1^0, C_2^0, C_3^0) \in Q$ *are bounded.*

Proof. In model from Eq. (3.7) if we consider C_1 to be unbounded, then

$$\lim_{C_1 \to \infty} \frac{dC_2}{dt} = \infty, \tag{3.15}$$

which would imply that C_2 would also be unbounded, and analogously for the case of C_3. We could then write

$$\lim_{C_i \to \infty} \frac{C_i}{1 + kC_3} = \lim_{C_i \to \infty} \frac{C_i}{1 + k\sum_{j=1}^{3} C_j} = \frac{1}{k} \quad \text{for } i = 1, 2, 3. \tag{3.16}$$

Let us now consider the functions f_i^ρ from Eq. (3.11). From Eq. (3.16) we get

$$f_1^\rho < \frac{\rho_1}{k} - \alpha_1 C_1, \tag{3.17}$$

and then for all t

$$C_1 < Ae^{-\alpha_1 t} + \frac{\rho_1}{k\alpha_1}, \quad A \in \mathbb{R}, \tag{3.18}$$

which yields that $C_1(t)$ is bounded. Considering $C_1(t) < M_1 \in \mathbb{R}$, for all t, then

$$f_2^\rho < \frac{\rho_2}{k} + \alpha_1 M_1 - \alpha_2 C_2, \tag{3.19}$$

and integrating as above implies C_2 is bounded. Considering $C_2 < M_2 \in \mathbb{R}$,

$$f_3^\rho < \alpha_2 M_2 - \alpha_3 C_3, \tag{3.20}$$

and thus C_3 is also bounded. ■

For the models B, ruled by Eqs. (3.8), we can sum the three equations to obtain

$$\frac{dC_T}{dt} = \rho_1 C_1 + \rho_2 C_2 - \frac{\alpha_3 C_3}{1+kN}, \tag{3.21}$$

where $C_T = C_1 + C_2 + C_3$. It is clear that for C_1 and C_2 sufficiently large the negative term makes only a small contribution and thus solutions are not bounded. This implies that models in which signalling affects only transition rates are not appropriate for representing biological processes of this kind.

3.2.2 Steady States and stability conditions

Biological processes in homeostasis are stable and robust. In addition to being mathematically well posed, we need to ensure that the models have a positive stable equilibrium in which the three populations coexist. In this Section we study the existence of such states and their local stability. We focus on models A (Eq. (3.7)), since models B (Eq. (3.8)) do not lead to biologically relevant dynamics. An expanded analysis of the stability conditions for certain steady states of models A can be found in Appendix A.1. Further analysis of models B, showing that they have only unstable non-trivial positive equilibria, is presented in Appendix A.2.

Model A1. Let us consider last stage signalling $s_\rho = s_1(t)$ as in Eq. (3.4) affecting the proliferation term, i.e. we study Eq. (3.7) with $N = C_3$. The three steady states for this model are

$$P_1^{A1} = (0,0,0), \tag{3.22a}$$

$$P_2^{A1} = \left(0, \frac{\alpha_3(\rho_2-\alpha_2)}{k\alpha_2^2}, \frac{\rho_2-\alpha_2}{k\alpha_2}\right), \tag{3.22b}$$

$$P_3^{A1} = \left(\frac{\alpha_3(\rho_1-\alpha_1)(\alpha_2\rho_1-\alpha_1\rho_2)}{k\alpha_1^2\alpha_2\rho_1}, \frac{\alpha_3(\rho_1-\alpha_1)}{k\alpha_1\alpha_2}, \frac{\rho_1-\alpha_1}{k\alpha_1}\right). \tag{3.22c}$$

The Jacobian matrix of the system at any point (C_1, C_2, C_3) is

$$J_{A1}(C_1,C_2,C_3) = \begin{pmatrix} \dfrac{\rho_1}{C_3k+1} - \alpha_1 & 0 & -\dfrac{C_1k\rho_1}{(C_3k+1)^2} \\ \alpha_1 & \dfrac{\rho_2}{C_3k+1} - \alpha_2 & -\dfrac{C_2k\rho_2}{(C_3k+1)^2} \\ 0 & \alpha_2 & -\alpha_3 \end{pmatrix}. \tag{3.23}$$

Substituting $P_1^{A1} = (0, 0, 0)$ in Eq. (3.23) we get the eigenvalues

$$\lambda_{1,1}^{A1} = -\alpha_3, \tag{3.24a}$$
$$\lambda_{1,2}^{A1} = \rho_1 - \alpha_1, \tag{3.24b}$$
$$\lambda_{1,3}^{A1} = \rho_2 - \alpha_2. \tag{3.24c}$$

This is the trivial equilibrium that would be unstable in normal homeostatic processes. Instability conditions are

$$\rho_1 > \alpha_1, \text{or } \rho_2 > \alpha_2. \tag{3.25}$$

As we consider $P_2^{A1} > 0$, it must be $\rho_2 > \alpha_2$, and therefore P_1^{A1} is unstable.

For P_2^{A1}, we obtain

$$\lambda_{2,1}^{A1} = \frac{\alpha_2\rho_1}{\rho_2} - \alpha_1, \tag{3.26a}$$
$$\lambda_{2,2}^{A1} = -\frac{\alpha_3}{2} - \frac{\sqrt{\alpha_3(4\alpha_2^2 - 4\alpha_2\rho_2 + \alpha_3\rho_2)}}{2\sqrt{\rho_2}}, \tag{3.26b}$$
$$\lambda_{2,3}^{A1} = -\frac{\alpha_3}{2} + \frac{\sqrt{\alpha_3(4\alpha_2^2 - 4\alpha_2\rho_2 + \alpha_3\rho_2)}}{2\sqrt{\rho_2}}. \tag{3.26c}$$

This equilibrium point corresponds to a situation where the less differentiated compartment disappears and the system is maintained only by the proliferation of the second, leading to mature cells. As before, this is not a biologically feasible situation, thus this equilibrium must be unstable. Since $\mathcal{R}(\lambda_{2,2}) < 0$ and $\mathcal{R}(\lambda_{2,3}) < 0$, then $\frac{\alpha_2\rho_1}{\rho_2} - \alpha_1 > 0$, which means that for P_2^{A1} to be unstable we must have

$$\alpha_2\rho_1 - \alpha_1\rho_2 > 0. \tag{3.27}$$

From the positivity of the non-trivial equilibrium point P_3^{A1} we require

$$\rho_1 > \alpha_1, \tag{3.28a}$$
$$\frac{\rho_1}{\rho_2} > \frac{\alpha_1}{\alpha_2}. \tag{3.28b}$$

Conditions (3.28b) and (3.27) are identical which means that the existence of this positive equilibrium implies the instability of P_2^{A1}. Stability conditions for P_3^{A1} are lengthy and can be found in Appendix A.1.

Model A2. Let us know focus on the model given by Eqs. (3.7) with $N = \sum_{i=1}^{3} C_i$. The equilibria are

$$P_1^{A2} = (0, 0, 0), \tag{3.29a}$$

$$P_2^{A2} = \left(0, \frac{\alpha_3(\rho_2 - \alpha_2)}{k\alpha_2(\alpha_2 + \alpha_3)}, \frac{\rho_2 - \alpha_2}{k(\alpha_2 + \alpha_3)}\right), \tag{3.29b}$$

$$P_3^{A2} = \left(\frac{\alpha_3(\rho_1 - \alpha_1)(\alpha_2\rho_1 - \alpha_1\rho_2)}{\alpha_1 k\beta}, \frac{\alpha_3\rho_1(\rho_1 - \alpha_1)}{k\beta}, \frac{\alpha_2\rho_1(\rho_1 - \alpha_1)}{k\beta}\right), \tag{3.29c}$$

where

$$\beta = (\alpha_2\alpha_3\rho_1 + \alpha_1(\alpha_2\rho_1 + \alpha_3(\rho_1 - \rho_2))). \tag{3.30}$$

The Jacobian matrix is

$$J_{A2} = \begin{pmatrix} -\alpha_1 + (s^{-1} - kC_1)s^2\rho_1 & -C_1ks^2\rho_1 & -C_1ks^2\rho_1 \\ \alpha_1 - C_2ks^2\rho_2 & -\alpha_2 + (s^{-1} - kC_2)s^2\rho_2 & -C_2ks^2\rho_2 \\ 0 & \alpha_2 & -\alpha_3 \end{pmatrix}. \tag{3.31}$$

Substituting P_1^{A2} in Eq. (3.31) we obtain

$$\lambda_{1,1}^{A2} = -\alpha_3, \tag{3.32a}$$

$$\lambda_{1,2}^{A2} = \rho_1 - \alpha_1, \tag{3.32b}$$

$$\lambda_{1,3}^{A2} = \rho_2 - \alpha_2. \tag{3.32c}$$

As before, for this equilibrium to be unstable we should have either $\rho_1 > \alpha_1$ or/and $\rho_2 > \alpha_2$. From the positivity of P_2^{A2}, it must be the case that $\rho_2 > \alpha_2$.

For P_2^{A2}, we obtain the eigenvalues

$$\lambda_{2,1}^{A2} = \frac{\alpha_2\rho_1}{\rho_2} - \alpha_1, \tag{3.33a}$$

$$\lambda_{2,2}^{A2} = \frac{\alpha_2^2\alpha_3 - 2\alpha_2\alpha_3\rho_2 - \alpha_3^2\rho_2 + h(\alpha_2, \alpha_3, \rho_2)}{2\rho_2(\alpha_2 + \alpha_3)}, \tag{3.33b}$$

$$\lambda_{2,3}^{A2} = \frac{\alpha_2^2\alpha_3 - 2\alpha_2\alpha_3\rho_2 - \alpha_3^2\rho_2 - h(\alpha_2, \alpha_3, \rho_2)}{2\rho_2(\alpha_2 + \alpha_3)}. \tag{3.33c}$$

where $h = h(\alpha_2, \alpha_3, \rho_2)$ such that

$$h = \sqrt{\alpha_3}\sqrt{2\alpha_2^2\alpha_3(\alpha_3 - 2\rho_2)\rho_2 + 4\alpha_2^3(\alpha_3 - \rho_2)\rho_2 + \alpha_3^3\rho_2^2 \ + \alpha_2^4(\alpha_3 + 4\rho_2)}. \tag{3.34}$$

As in the previous model, the existence of P_3^{A2} as a positive equilibrium influences the stability of P_2^{A2}. For the positivity of P_3^{A2} we have a first set of conditions

$$\beta > 0, \tag{3.35a}$$
$$\rho_1 > \alpha_1, \tag{3.35b}$$
$$\alpha_2\rho_1 > \alpha_1\rho_2; \tag{3.35c}$$

or a second set of conditions

$$\beta < 0, \tag{3.36a}$$
$$\rho_1 < \alpha_1, \tag{3.36b}$$
$$\alpha_2\rho_1 < \alpha_1\rho_2. \tag{3.36c}$$

Let us first consider that Eq. (3.35) holds. Then, in Eq. (3.33), the eigenvalue $\lambda_{2,1}^{A2}$ is positive and P_2^{A2} would be unstable. On the other hand, if Eq. (3.36) holds, then in Eq. (3.33) the eigenvalue $\lambda_{2,1}^{A2} < 0$. Then P_2^{A2} would be stable whenever

$$|\mathcal{R}(h)| > \alpha_2^2\alpha_3 - 2\alpha_2\alpha_3\rho_2 - \alpha_3^2\rho_2 \tag{3.37}$$

for h as defined in Eq. (3.34).

In this case, for the stability of P_2^{A2}, the difference with model A1 is the existence of a denominator β (Eq. (3.30)). The stability conditions for P_3^{A2} are not shown here for reasons of space, but they are presented in Appendix A.1. Stability conditions related to model B, as well as a summary of all the stability conditions depending on both models, can be found in Appendices A.2 and A.3, respectively.

3.3 Data

3.3.1 Patients

Bone marrow samples from six individuals of paediatric age (1 to 13 years) were used to estimate cell subset proportions. Four patients diagnosed with Idiopathic Thrombocytopenic Purpura (1 from Jerez Hospital and 3 from Niño Jesús Hospital) and two patients with neutropenia (Jerez Hospital). Due to the difficulty in obtaining healthy bone marrow samples from patients of paediatric age, we selected the above as surrogate examples of normal B-cell development. Bone marrow samples were extracted from these patients in order to check

for more severe disorders, but they were later diagnosed with B cell-unrelated diseases. Sample inspection further ensured lack of B-line affection. Instances of this can be found in the literature [154, 155].

3.3.2 Flow Cytometry

Bone marrow samples were analysed by flow cytometry. This technique measures the expression of the immunophenotypic markers that characterise each maturation stage. Marker expression for both hospitals' data was acquired on a FACSCanto II flow cytometer following the manufacturer's specifications Becton Dickinson (BD) for sample preparation. Samples were stained using an 8-colour panel consisting of the following six fluorochrome-conjugated antibodies provided by BD: CD38 FITC/ CD10 PE/ CD34 PerCP-Cy5-5/ CD19 PE-Cy7/ CD20 APC/ CD45 V-450. This panel allows for the identification of B cell subpopulations [147]. Forward (FSC) and side scatter (SSC) were also measured.

Samples were preprocessed removing debris, doublets and marginal events as is routinely done in clinical and computational flow cytometry [23]. $CD19^+$ cells were then gated in order to select B lymphocytes [156]. Since the model consists of three B cell populations, we performed k-means clustering on each sample, including all B cell markers, with 3 predefined clusters. The algorithm was initialised randomly and 100 random sets were generated, selecting the one with lower within-cluster variation. Following standard immunophenotyping of B cells [18], clusters were classified into early ($CD45^-/CD10^+$), intermediate ($CD45^+/CD10^+$) or late ($CD45^+/CD10^-$) B cells. Proportions were then computed with respect to the total B lymphocyte count ($CD19^+$ population), which correlates with experimental data [148]. Figure 3.3 shows the three stages of the process. All computations were carried out in RStudio using packages flowCore [157] and flowPeaks [158].

3.4 Numerical results

3.4.1 Parameter estimation

In the models presented above there are two parameters related to proliferation ρ_i $(i = 1, 2)$, three related to compartmental transitions α_i $(i = 1, 2, 3)$, and another related to the strength of signalling, the inhibition constant k.

A direct measure of the proliferation rates for the specific subsets considered in this Chapter is lacking, but we can provide an estimation based on qualitative biological information. Normal B-cell development can be compared to data

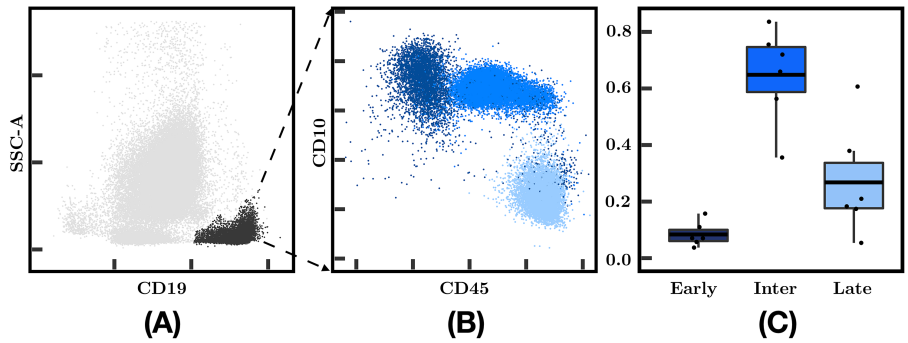


Figure 3.3: **CD19$^+$ B cell subpopulation gating in flow cytometry data**. **(A)** B Lymphocyte selection in grey (CD19$^+$ population). Y-axis represents cellular complexity, which is low for lymphocytes, and X-axis represents B Lymphocyte surface antigen CD19. **(B)** Within B population, three clusters were found corresponding to three maturation stages: early, intermediate and late populations. As B cells mature, they gain expression of CD45 antigen and lose expression of CD10 antigen [18]. **(C)** Boxplots with proportions of each maturation stage from 6 patients displaying mean and 1st and 3rd quartile. Mean values with standard deviations: $0.083 \pm 0.009, 0.648 \pm 0.145, 0.268 \pm 0.192$.

from autologous bone marrow transplantation. This type of transplantation is more likely to reproduce developmental ontogeny, especially if there is no prior or coadjuvant anti-B-cell therapy [159]. In this case B-cell progenitors can be detected in bone marrow as early as 1 month after transplantation [160, 161], and in blood after some delay [162]. *In-vitro* studies with mice show that proliferation rates are of the order of magnitude of days [149, 163, 164]. Human lymphoid cultures suggest doubling times of 1 day [165]. With respect to the relative values, current schemes for B-cell maturation indicate that upon CD19 acquisition there is sustained proliferation that decreases as the cell matures [149, 166, 167]. Late B cells already express B-cell receptor and a negative selection process occurs prior to this [137, 168]. Based on this we can consider that the net production rate of intermediate cells is lower than early cells ($\rho_1 > \rho_2$).

In order to obtain values for transition rates we make use of steady state expressions given by (3.22c) or (3.29c). These values can be compared to the flow cytometry data of normal bone marrow (see Section 3.3). However, in our

simulations we are measuring absolute cell counts while flow cytometry is only able to measure relative cell proportions. The positive non-trivial equilibrium (C_1^*, C_2^*, C_3^*) given by (3.22c) or (3.29c) satisfies the relationships

$$\frac{C_2^*}{C_3^*} = \frac{\alpha_3}{\alpha_2}, \tag{3.38a}$$

$$\frac{C_1^*}{C_2^*} = \frac{\alpha_2\rho_1 - \alpha_1\rho_2}{\alpha_1\rho_1}. \tag{3.38b}$$

These quantities can be obtained from the analysis of the relative abundance of the three different populations. Thus, for each blood transition rate α_3 we can obtain values of α_1, α_2 that agree with steady-state data and at the same time belong in the positive stability region of parameter space. An example of the correlation found between transition rates is shown in Figure 3.4. The order of magnitude of α_3 can be estimated as follows. The term $\alpha_3 C_3^*$ gives the total number of B cells per hour sent to blood by the bone marrow in homeostatic circumstances. In mice, bone marrow produces around 0.1% of the steady state population per hour [169]. In humans, the total steady-state B-lymphocyte population can be obtained from lymphocyte proportions and bone-marrow volumes from the literature and we estimate it to be 10^{10} cells [170, 171, 172, 173, 174]. This yields a B-cell production of 10^7 cells per hour, and given that $C_3^* \sim 10^9$ we estimate α_3 to be of the order of magnitude of 10^{-2} h^{-1}.

The last parameter to be determined is the inhibition constant k. Inspection of steady states shows that, given transition and proliferation rates of the order of magnitude of days, k is of the order of magnitude of the inverse of the total steady-state population $k \propto (C_1^* + C_2^* + C_3^*)^{-1}$. The precise values for k and α_3 are selected so that we recover steady-state values and reconstitution times compatible with the literature cited above. Ranges of parameters in agreement with positivity, stability and steady-state conditions are shown in Table 3.2.

With respect to the initial state, the absolute number of mononucleated transplanted cells (MNC) is in the range of 10^9 cells [162]. From these cells, only a 1% are B early cells (10^7 cells) [148, 175]. These cells travel through blood into the bone marrow but only 10% of cells eventually reach the bone marrow [176]. Therefore, we will consider for the numerical simulations of autologous transplantation an initial absolute number of cells of $C_1(0) = 10^6$, $C_2(0) = 0$, $C_3(0) = 0$. The influence of this initial value on the dynamics of the system is described in Sec. 3.4.3.

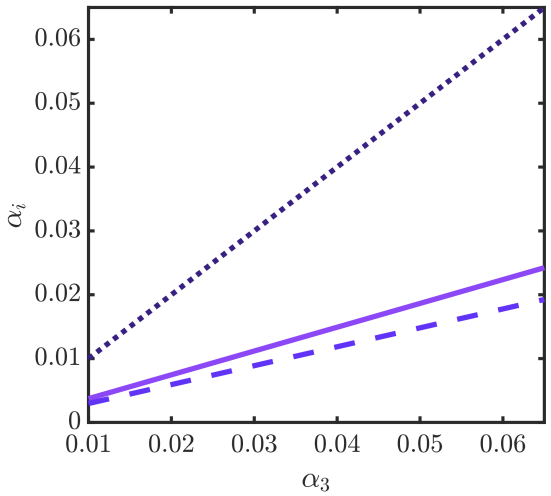


Figure 3.4: **Dependence of the transition rates over the range of stability**. Lines represent α_1 (light purple solid line), α_2 (purple dashed line), α_3 (dark purple dotted line). Across this range α_j satisfies the relationship $\alpha_3 > \alpha_1 > \alpha_2$. Values computed from Eq. (3.38) with $\rho_1 = 0.0289$ h^{-1}, $\rho_2 = 0.0193$ h^{-1} and steady-state cell proportions $C_1^*/C_2^* = 0.0986$ and $C_2^*/C_3^* = 3.227$.

3.4.2 Global feedback signalling results in a smoother transition to steady states

Typical results of simulations of models A1 and A2 (Eq. (3.7)) are shown in Figure 3.5. Recall that in model A1 (Figure 3.5(A)) the signal depends only on cells in the final compartment while in model A2 (Figure 3.5(B)) all cells are involved. Both models exhibit qualitatively similar behaviour. Early cells appear first, reaching a peak in population numbers slightly before day 30. Intermediate cells follow, reaching the respective peak with days of delay and with larger cell numbers. Late cells appear last and stabilise in between. The system settles into the steady state from day 80 onwards. This behaviour agrees with the conditions expressed above for the parameter values. In particular, the proportion of population from each stage is coherent with clinical data (see Section 3.3) and experimental data [148]. For both models we have 8.99% early cells, 70.21% intermediate cells and 20.80% late cells.

Param.	Meaning	Value	Units
ρ_1	Early B proliferation rate	$\ln(2)/24$	hours^{-1}
ρ_2	Intermediate B proliferation rate	$\ln(2)/36$	hours^{-1}
α_1	Transition rate: early to intermediate	(0.004, 0.025)	hours^{-1}
α_2	Transition rate: intermediate to late	(0.003, 0.02)	hours^{-1}
α_3	Transition rate: late to blood	(0.01, 0.065)	hours^{-1}
k	Inhibition constant	$(10^{-11}, 10^{-10})$	cells^{-1}

Table 3.2: Parameter values for Eqs. (3.7).

There are two main differences between the two models. The first relates to the magnitude of the steady state, which is larger when only late cells participate in signalling (model A1, Figure 3.5(A)) for the same parameter values. This can be observed in Figure 3.6(A), where total cell numbers for both models are shown. Note that peak lymphocyte count, i.e. the largest cell number, occurs at day 30, when intermediate cells are maximal. The second difference relates to the early behaviour of the reconstitution. In global signalling simulation (model A2, Figure 3.5(B)) there is a much less pronounced peak than when only late cells perform the signalling, presenting a smoother transition to the equilibrium state.

Figures 3.5(C) and (D) show the evolution of the percentage of each maturation stage. It is interesting to relate this to absolute counts since, as explained in Sec. 3.4.1, flow cytometry data only captures relative cell proportions. We observe very close behaviour between the models. The percentage of early cells quickly decreases as more mature stages appear and steady-state proportions are reached from day 80 onwards. Note that even though intermediate and mature cells have a peak in absolute cell count, this peak is not represented in terms of percentage. Also, this figure shows that a decrease in cell percentage does not necessarily means a decrease in absolute cell count, something to take into account when interpreting longitudinal flow cytometry data.

3.4.3 Time to peak decreases exponentially with initial value

In Sec. 3.4.1 we described the rationale for the choice of the initial value of early cells. Despite this, we sought to determine how the scale of this data impacted reconstitution times. In Figure 3.6(B) we show the time to peak cell count for a range of initial values for early cells. There exists a decreasing

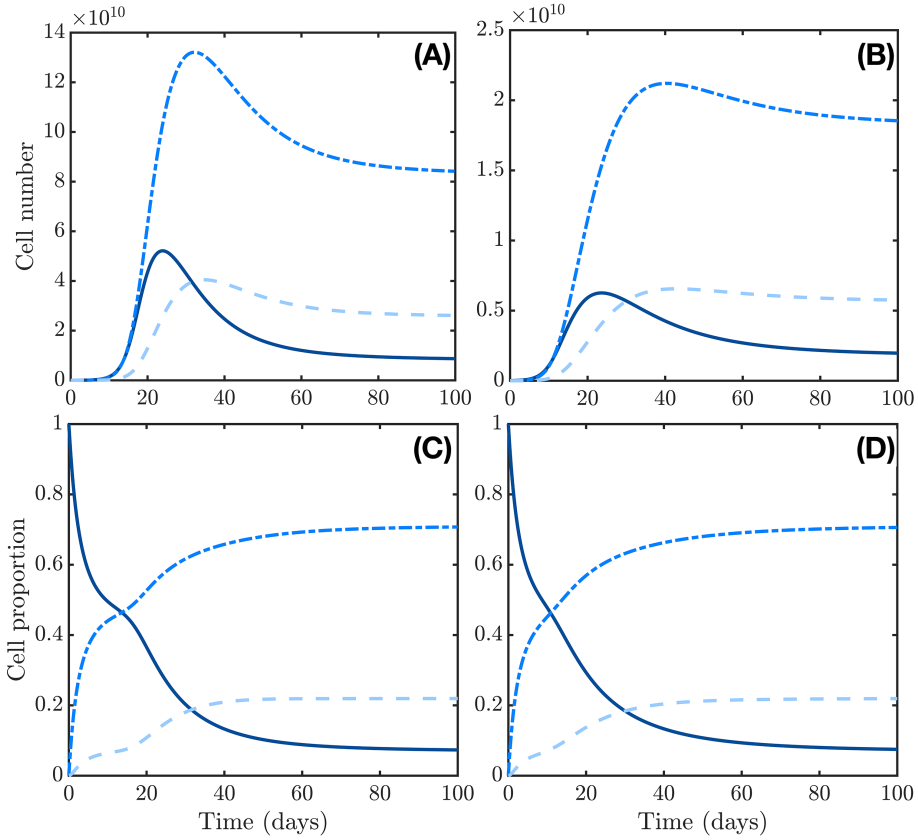


Figure 3.5: **Comparison of numerical solutions and relative proportions of models A1 and A2 from Eqs.** (3.7)**.** Comparison of numerical solutions of models A1 **(A)** and A2 **(B)** (Eq. (3.7)) and relative proportions **(C)** and **(D)**, respectively, during the first 100 days. Curves represent early cells C_1 (dark blue solid line), intermediate cells C_2 (blue dashed-dotted line) and late cells C_3 (light blue dashed line). Both simulations have initial data $C_1(0) = 10^6$, $C_2(0) = 0$, $C_3(0) = 0$ cells and parameter values $\rho_1 = 0.0289\ \mathrm{h}^{-1}$, $\rho_2 = 0.0193\ \mathrm{h}^{-1}$, $\alpha_1 = 0.008\ \mathrm{h}^{-1}$, $\alpha_2 = 0.006\ \mathrm{h}^{-1}$, $\alpha_3 = 0.02\ \mathrm{h}^{-1}$ and $k = 10^{-10}$.

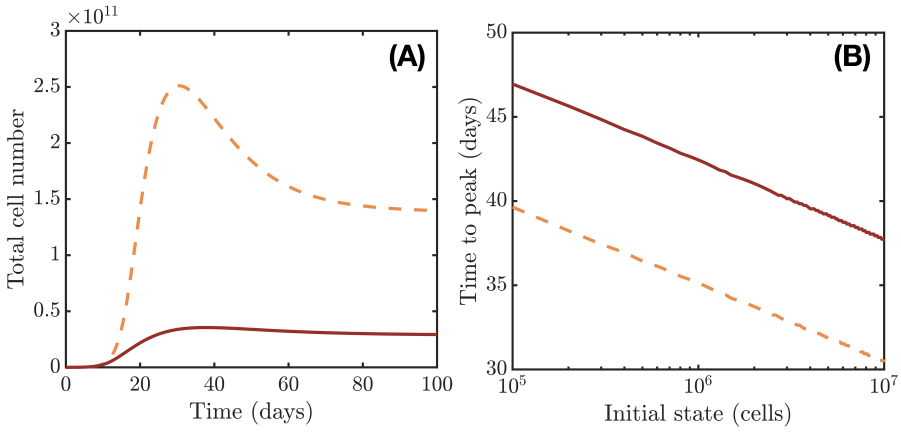


Figure 3.6: **Comparison of numerical simulations of model A1 (orange solid line) and A2 (red dashed line) from Eqs.** (3.7) for: **(A)** total number of cells $C_1 + C_2 + C_3$, with initial cell numbers $C_1(0) = 10^6$, $C_2(0) = 0$, $C_3(0) = 0$; **(B)** time to peak lymphocyte count as a function of initial early B cell number. Both simulations have parameter values $\rho_1 = 0.0289$ h^{-1}, $\rho_2 = 0.0193$ h^{-1}, $\alpha_1 = 0.008$ h^{-1}, $\alpha_2 = 0.006$ h^{-1}, $\alpha_3 = 0.02$ h^{-1} and $k = 10^{-10}$.

exponential relationship between the two magnitudes, although the delay is not significant when considering that the literature on reconstitution after autologous transplantation describes reconstitution times in the range of 1-2 months [160]. Multiplying initial cell numbers by 10 results in a displacement in time of 5 days.

3.4.4 Blood transition rate influences time to bone marrow reconstitution

Clinical data suggests that homeostatic bone marrow displays relatively constant subset proportions (see Section 3.3). This, together with the analysis of the expressions of the steady states allowed us to derive a connection between the three transition rates in the model (Eq. 3.38). For the ranges of parameters considered, we observed that $\alpha_3 > \alpha_1 > \alpha_2$ (see Figure 3.4), which suggests that the second compartment being more numerous could be due not to a higher

proliferation rate but rather to a slower maturation time. This calls for the analysis of the influence of transition rates on the dynamics of the system.

In order to do this we focused on variations of α_3, the rate at which late cells exit bone marrow and enter the blood flow. We select a range of variation that lies in the positive stability region and observe the qualitative differences in the immune reconstitution. Results are shown in Figure 3.7(A).

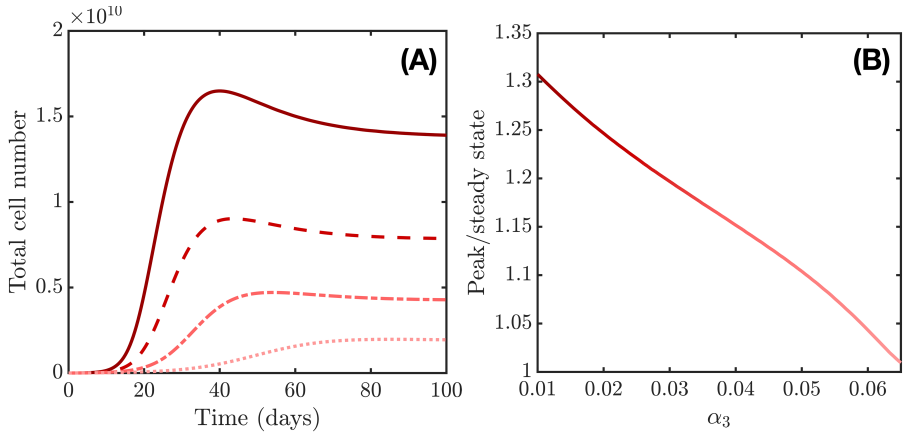


Figure 3.7: **Numerical simulations for variable blood transition rate**. **(A)** Total cell number $C_1 + C_2 + C_3$ for $t \in [0, 100]$ days. Blood transition rate values: $\alpha_3 = 0.03\ \mathrm{h}^{-1}$ (solid line), $\alpha_3 = 0.04\ \mathrm{h}^{-1}$ (dashed line), $\alpha_3 = 0.05\ \mathrm{h}^{-1}$ (dashed-dotted line), $\alpha_3 = 0.06\ \mathrm{h}^{-1}$ (dotted line). Line colour goes from dark red to pink as blood transition rate increases. **(B)** Peak to steady state value ratio for the same range of blood transition rates. Both simulations belong to model A2 (Eq. (3.7)) with initial state $C_1(0) = 10^6$, $C_2(0) = 0$, $C_3(0) = 0$ cells and parameter values $\rho_1 = 0.0289\ \mathrm{h}^{-1}$, $\rho_2 = 0.0193\ \mathrm{h}^{-1}$ and $k = 10^{-10}$. Parameters α_2 and α_1 vary with α_3 according to Eq. (3.38).

We observe that an increasing blood transition rate means that the system reaches a lower number of total cells. Indeed, Eq. (3.22) shows that population levels depend on transition rates. Note that the peak during reconstitution also decreases. In order to quantify this reduction, we show, in Figure 3.7(B), the proportion of the height of the peak with respect to the final steady state, for the same range of values for α_3. For lower values of blood transition rate, the transitory population of lymphocytes can be 1.3 times the population in

homeostatic conditions. This proportion decreases if cells increase the rate at which they join blood flow. Another consequence coming from these numerical simulations is that the lower the transition rates, the longer it takes for the system to stabilise.

3.4.5 Inhibition constant has no qualitative impact on the dynamical process

Eq. (3.22) shows that steady state values depend on transition rates (see also Figure 3.7) but also on the inhibition constant k. To quantify the impact of the parameter k on the dynamics of the system, we computed the total steady-state cell number as well as the proportion of peak height to steady state in a range of α_3 (and thus α_1 and α_2) and k values. The results are shown in Figure 3.8 for model A2. With respect to the absolute final cell count, only low levels of both k and α_3 result in a much higher number of cells. We highlighted an area for which the total cell amount C_T is in the range $(10^{10}, 10^{11})$, as estimated in Sec. 3.4.1 from reference values. For the proportion of peak height to steady state, there is little variation in the direction of k, so signalling intensity has little influence on the dynamics. As shown in Figure 3.7, high peak values belong in the low blood transition rate area. We conclude that transition rates primarily cause the overshoot during reconstitution, while k is mainly responsible for the existence of stability regimes and the size of the final states.

3.5 Discussion and conclusions

Haematopoiesis is one of the most widely studied biological developmental processes [120]. Interesting questions arise related to the processes of cell lineage specification [7, 177], the role of stem cells [6, 178] and the way cells communicate to regulate and ensure steady production [179, 180]. This is also true for B lymphopoiesis, the branch of haematopoiesis pertaining to B-cell formation. Specific unknowns in B-cell biology are the origins of some developmental stages, the role of senescence or the array of cytokines that regulates this process [181]. Studies of human B lymphopoiesis are encouraged, given that most of our knowledge about this line comes from mouse models [122]. Answering these questions and obtaining a precise description of the dynamics of B-cell maturation are fundamental for research avenues. Some examples are the characterisation of haematological malignancies, the reconstitution after

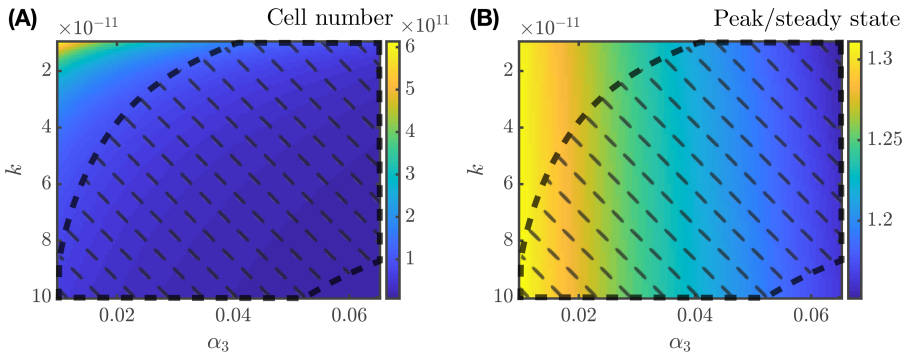


Figure 3.8: **Influence of inhibition constant k and blood transition rate α_3 parameters in model A2** (Eq. (3.7) with all cell signalling) for **(A)** total steady state cell number, **(B)** peak vs steady states ratio. The dashed highlighted area represents the range in which total cell number $C_T \in (10^{10}, 10^{11})$. Each coordinate is the result of a numerical simulation of model A2 with initial state cell numbers $C_1(0) = 10^6$, $C_2(0) = 0$, $C_3(0) = 0$ and parameter values $\rho_1 = 0.0289$ h^{-1} and $\rho_2 = 0.0193$ h^{-1}. Parameters α_1 and α_2 vary with $\alpha_3 \in (0.01$ h^{-1}, 0.03 h$^{-1})$ as explained in Eq. (3.38).

bone marrow transplant or chemotherapy or the generation of new lymphocytes from human embryonic stem cells [181].

Mathematical models have the potential to integrate biological hypotheses and clinical data in order to provide an abstract representation of biological processes [126]. This representation can be useful not only for understanding the dynamical properties of the system, but also for testing and elucidating more inaccessible phenomena. Our goal in this Chapter was to establish a biologically sensible mathematical characterization of the B lymphopoiesis. Spurred by previous models of haematopoietic processes [32, 129], we designed four models each with three differentiation stages. We added an implicit and systemic consideration of cell feedback signalling resulting in four nonlinear models. We first analysed these models from a theoretical perspective, addressing existence, positivity, boundedness and local stability. We collected data from the literature and clinical data from haematological patients and then used numerical simulations in order to understand the role and influence of each parameter.

We learn from theoretical analysis that a stable, homeostatic situation cannot arise solely by regulating the transition rate, i.e. the process of differentiation or maturation to the next compartment. We have focused from there onward on feedback regulation of cell proliferation. Inspection of the steady states allowed us to use flow cytometry data to establish a relationship between the three transition rates, analysing their influence by manipulating them one by one. The relationship suggests that, for the specific cell proportions in the B line, cells transition faster from early to intermediate than from intermediate to late. Intermediate cells could then be more numerous not because they proliferate more, but because of their slower transition rates downstream towards more mature cell types. Also, numerical simulations show that cell proportion is independent of which cells perform the signalling.

In this sense, we observed that signalling coming from all stages results in a smoother reconstitution of the B cell line. Indeed, when all populations participate in signalling, their influence occurs earlier and proliferation decreases faster with time than when only late cells do. In this case we observed a peak that we understood as a consequence of the delay in the reaction of the system to overpopulation. The amplitude of this peak is also correlated with lower transition rates, which implies lower steady states values. This is biologically understandable: the late compartment saturates due to excessive input from previous compartments, delaying access to stability and thus maintaining cell production in upstream compartments. It is important to remark here that subset percentage, a common metric in follow-up samples in a clinical context, can be misleading when dynamics of expansion are at play. For example, the overshoot during early reconstitution is not observed in terms of relative proportions. Finally, we noted that the strength of the signalling has no impact on the dynamics. The feedback loop could then be understood as a mechanism for the existence of stable output, while dynamical characteristics (time to reconstitution or early peak) are more dependent on intrinsic cellular processes.

The idea of this study was to determine which conditions are sufficient, from a mathematical perspective, to represent the kind of biological data that is currently available for B-line development. While we obtain a behaviour that fits with the time scales of the *in-vivo* process [148, 159, 160, 161, 162, 165], our study has a series of limitations. Firstly, the choice of three compartments could be refined or expanded following a more detailed characterisation of the cells. Multidimensional flow cytometry data shows that surface markers, those that specify to which stage a given cell belongs, vary continuously. A mathematical model where these markers vary continuously might be able to capture this variation. Secondly, we described signalling as a systemic

phenomenon. While this was enough to recapitulate known B-cell behaviour, a more detailed description including two or more types of signalling is desirable. Lastly, the model would benefit from longitudinal data coming from immune reconstitution of the B-cell line. In this regard, flow cytometry analyses of both peripheral blood and bone marrow in routine follow-up would allow for a more precise parametrisation and enable the hypotheses presented above to be contrasted.

To conclude, we have constructed and studied several non-linear compartmental models describing B cell lymphocyte reconstitution. These simple models describe the process of B-cell generation as portrayed by bone marrow data, and we consider it a first step in a deeper exploration of the phenomena associated with B-cell development. We verified mathematical and biological consistence, opening the door to interesting mathematical research like the existence of bifurcations or the conditions for global stability, something that finds immediate application in cases of immune reconstitution. Studies of this kind can function as a source of hypothesis generation in biomedical research, for example when contrasting mouse versus human dynamics. Ultimately, we aim to extend the methodology to situations of stability disruption and abnormal growth like B-cell disorders and other haematological diseases.

The methods, results and conclusions related to this Chapter can be found in Ref. [182].

PART II

Analyses of acute lymphoblastic leukaemia high-dimensional data

CHAPTER 4

Discriminant analysis of relapse biomarkers

Acute Lymphoblastic Leukaemia (ALL) is the most common childhood cancer, accounting for 40% of all paediatric neoplasias [183]. This disease is characterised by the abnormal growth of immature lympchocytes in the bone marrow (BM). ALLs are classified as B- or T-ALL depending on the lineage of the cells of origin of the malignancy [184]. The former comprises the majority of cases in children and has better prognosis than the later. The clinical manifestations of B-ALL are the result of the invasion of the bone marrow, having more than 25% of blasts or immature lymphocytes, which leads to a shortage of healthy haematological cells. Progresses in diagnosis, risk assessment and therapeutics have increased survival rates to around 80% [185]. The prognosis of relapsing patients is substantially worse. The early identification of relapsing patients is of high clinical interest since it could allow to use immunotherapy or bone marrow transplant where appropriate, as a first line treatment [186].

Flow cytometry is widely used in the diagnosis and follow up of B-ALL [187]. This technique allows the measurement of the surface expression levels of selected proteins for individual cells. Since each cell development state is characterised by a set of these markers, flow cytometry allows to classify the different cell populations within the BM in comparison with the normal BM and assist with the disease diagnosis [19]. Typical diagnostic flow cytometry studies interrogate between 10^5 and 10^6 cells. The outcome of the analysis is a dataset with surface protein expressions, complexity and size at the single-cell level. Flow cytometers used in clinical contexts can detect between 4 and 18 markers [188]. The size of the flow-cytometry datasets is increasing as technology progresses, and at some point the manual analysis carried out nowadays by

cytometrists will no longer be feasible [189].

Combinations of single cell cytometric data and bioinformatics pipelines have been recently used for biomarker discovery in lymphoma [190], renal cell carcinoma [191], melanoma [192] or lung adenocarcinoma [193]. In childhood B-cell ALL, there are studies on automated follow up [194] and relapse prediction, using either clinical characteristics [195] or mass cytometry data [196]. Having biomarkers of response to current chemotherapeutical protocols on diagnosis is of high interest in order to consider alternative therapeutic options, such as bone marrow transplants or CAR-T cells.

The aim of this study was to find a collection of surface proteins showing significant differences in expression levels between relapsed and non-relapsed patients on diagnosis. To do we took advantage of the high-dimensionality of flow cytometry data and a multicentre database of patients. To do so we performed several tasks. First, data had to be preprocessed to solve problems like missing data values or data imbalance. Next, we had to find a subset of relevant features to be used for classification, what we addressed using Fisher's linear discriminant analysis. Finally, a mathematical model using those features classifying patients on diagnosis was developed and validated.

4.1 Materials and Methods

4.1.1 Patients

A retrospective study was designed in accordance with the Declaration of Helsinki, and the protocol was approved by the institutional review board (IRB) of the two participating local institutions (LLAMAT Project, 2018).

Inclusion criteria for the study were ALL diagnosis between February 2009 and October 2017, age over 1 year and less than 19 years, and availability of bone marrow flow cytometry data. A total of 105 patients satisfied the inclusion criteria. Exclusion criteria were availability of Flow Cytometry files FCS below 3.0, patients without a minimum of 15 immunophenotypic (IPT) markers in common with others in the dataset, and insufficient follow-up for non-relapsing patients, i.e. patients without relapse but with less than three years after no refractory values for minimal residual disease were found. Finally, 54 patients diagnosed in two of the local institutions were retained for further analysis. Dataset 1 included 28 non-relapsed and 8 relapsed patients, while dataset 2 included 13 non-relapsed and 5 relapsed patients.

4.1.2 Flow cytometer machines and antibodies

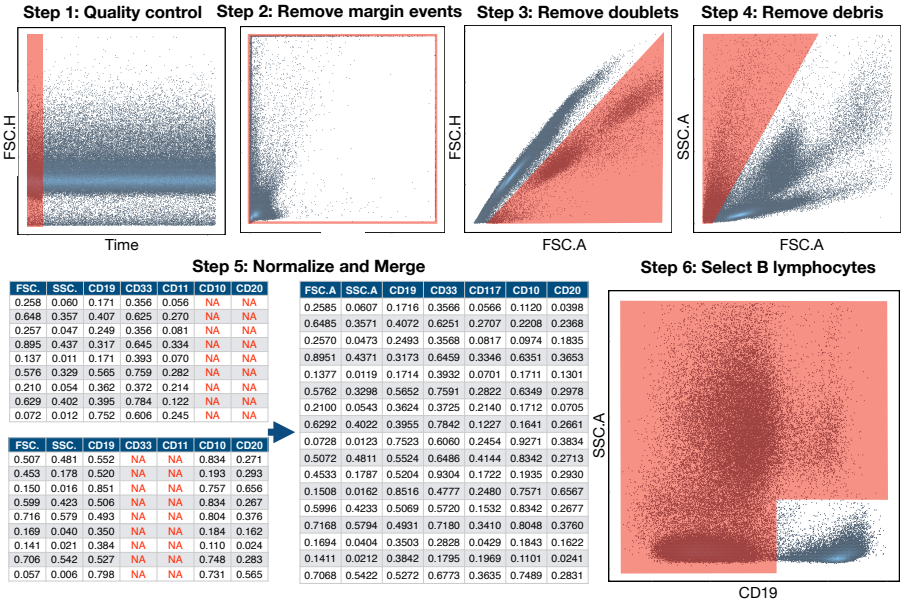
Marker expression was obtained on FACSCanto II flow cytometers, in accordance with the manufacturer's specifications for sample preparation. Final samples were stained using an 8-colour panel with six fluorochrome-conjugated antibodies.

FCS 3.0 files contained information on forward scatter (FSC) (interpreted as size), side scatter (SSC) light (interpreted as complexity) and monoclonal antibodies used routinely in diagnosis. The markers used included B-cell (CD19, CD10, CD20, CD22, CD24, IgM, CD66c, CD79a, kappa, lambda, etc.) and T-cell-related (CD7, cyCD3) IPT markers, markers related to the myeloid lineage (CD9, CD13, CD33, CD123), and some general ones (CD15, CD34, CD38, CD45, CD58, CD71, HLA-DR).

4.1.3 Preprocessing

Files were first imported into FlowJo (Becton Dickinson, 10.6.1) and FACSDiva (Becton Dickinson, 8.0.1) and inspected manually. Quality control was performed and margin events, debris, dead cells and doublets were removed, as shown in Figure 4.1 steps 1-4. Files were then further processed in R (3.6.0) and RStudio (1.2.1335). This software, in conjunction with Bioconductor (3.11) provides packages and methods for analysis of flow cytometry data. Tubes were compensated by means of the spillover matrix included in each file and then transformed with the Logicle transformation [197] included in the flowCore package (2.0.1) [157] with parameters $w = 0.75$, $t = 262144$, $m = 4.5$ and $a = 0$. Our next step was to bring into a single file the information contained in each of the patient's tubes. Since each tube contains marker intensity for different markers and cells, the full set of 20 markers was not available for any of the cells, as shown in Figure 4.1 step 5. This posed a problem of missing data imputation, that is addressed in different ways in the context of flow cytometry [198, 199, 200, 201]. We followed the methodology described in [198], which consist of nearest-neighbour imputation using the common or backbone markers in all aliquots. The result of this process was a set of 38 files, one per patient, containing complete information of the 20 IPT markers selected for the analysis. After this step, 10^5 events were randomly sampled from each file in order to have the same number of cells for each patient.

Since data of multicenter retrospective studies can be affected by batch effects and technical variations across time and centre, we performed a normalisation based on a modified min-max transformation. This transformation brings all



FSC.	SSC.	CD19	CD33	CD11	CD10	CD20
0.258	0.060	0.171	0.356	0.056	NA	NA
0.648	0.357	0.407	0.625	0.270	NA	NA
0.257	0.047	0.249	0.356	0.081	NA	NA
0.895	0.437	0.317	0.645	0.334	NA	NA
0.137	0.011	0.171	0.393	0.070	NA	NA
0.576	0.329	0.565	0.759	0.282	NA	NA
0.210	0.054	0.362	0.372	0.214	NA	NA
0.629	0.402	0.395	0.784	0.122	NA	NA
0.072	0.012	0.752	0.606	0.245	NA	NA

FSC.	SSC.	CD19	CD33	CD11	CD10	CD20
0.507	0.481	0.552	NA	NA	0.834	0.271
0.453	0.178	0.520	NA	NA	0.193	0.293
0.150	0.016	0.851	NA	NA	0.757	0.656
0.599	0.423	0.506	NA	NA	0.834	0.267
0.716	0.579	0.493	NA	NA	0.804	0.376
0.169	0.040	0.350	NA	NA	0.184	0.162
0.141	0.021	0.384	NA	NA	0.110	0.024
0.706	0.542	0.527	NA	NA	0.748	0.283
0.057	0.006	0.798	NA	NA	0.731	0.565

FSC.A	SSC.A	CD19	CD33	CD117	CD10	CD20
0.2585	0.0607	0.1716	0.3566	0.0566	0.1120	0.0398
0.6485	0.3571	0.4072	0.6251	0.2707	0.2208	0.2368
0.2570	0.0473	0.2493	0.3568	0.0817	0.0974	0.1835
0.8951	0.4371	0.3173	0.6459	0.3346	0.6351	0.3653
0.1377	0.0119	0.1714	0.3932	0.0701	0.1711	0.1301
0.5762	0.3298	0.5652	0.7591	0.2822	0.6349	0.2978
0.2100	0.0543	0.3624	0.3725	0.2140	0.1712	0.0705
0.6292	0.4022	0.3955	0.7842	0.1227	0.1641	0.2661
0.0728	0.0123	0.7523	0.6060	0.2454	0.9271	0.3834
0.5072	0.4811	0.5524	0.6486	0.4144	0.8342	0.2713
0.4533	0.1787	0.5204	0.9304	0.1722	0.1935	0.2930
0.1508	0.0162	0.8516	0.4777	0.2480	0.7571	0.6567
0.5996	0.4233	0.5069	0.5720	0.1532	0.8342	0.2677
0.7168	0.5794	0.4931	0.7180	0.3410	0.8048	0.3760
0.1694	0.0404	0.3503	0.2828	0.0429	0.1843	0.1622
0.1411	0.0212	0.3842	0.1795	0.1969	0.1101	0.0241
0.7068	0.5422	0.5272	0.6773	0.3635	0.7489	0.2831

Figure 4.1: **Preprocessing pipeline**. Preprocessing was carried out in six steps. The first four were performed in FlowJo and consisted in the removal of abnormal acquisitions (quality control), margin events, doublets and debris. Files were then imported into R in step 5, and for each patient all tubes or aliquots were merged into a single file by means of nearest-neighbour imputation. Finally, in step 6, the CD19$^+$ population (B cells) was selected for further analysis.

data points to the range $[0, 1]$ but it is sensitive to outliers. Instead of selecting the maximum and minimum values, we chose quantiles 0.05 and 0.95 and applied the transformation:

$$x' = \frac{x - x_{q0.05}}{x_{q0.95} - x_{q0.05}}, \tag{4.1}$$

where $x_{q0.05}$ is the 5$^{\text{th}}$ percentile and $x_{q0.95}$ is the 95$^{\text{th}}$ percentile. Finally, we used the common B-cell antigen CD19 to select the B-cell subpopulation, as shown in Figure 4.1 step 6. Files were imported in MATLAB (Mathworks, R_2020a) via the *fca_readfcs* function [202].

4.1.4 Fisher's linear discriminant for relapse prediction

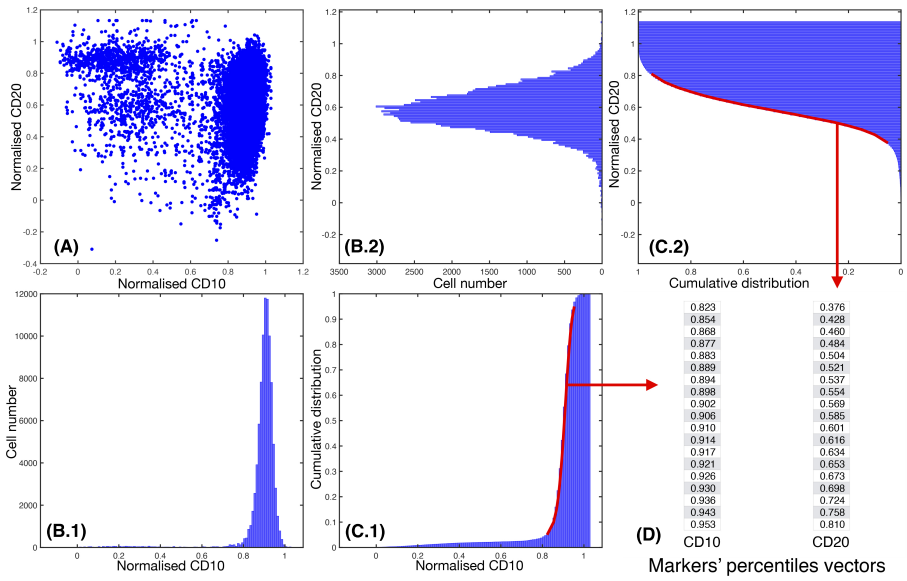


Figure 4.2: **Percentile vector construction**. **(A)** Scatter plot of a patient i for two normalised parameters, k_1=CD10 and k_2=CD20. **(B.1)** and **(B.2)**: Histograms cell count of, respectively, k_1=CD10 and k_2=CD20. **(C.1)** and **(C.2)**: Cumulative distribution of markers k_1=CD10 and k_2=CD20, respectively. In red, percentiles curve from 5th to 95th percentile. **(D)** Each percentile curve for each patient i and marker k results in a vector P_{ij}, where j represents each percentile chosen.

Let us consider then $P_{ij} \in \mathbb{R}^p$ as vectors obtained for each patient i and each common feature j, for $i = 1, ..., n$ patients and $j = 1, ..., m$ IPT markers. Thus, for each patient i, this results in a matrix $M \in \mathbb{R}^{m \times p}$ of the p percentiles from all IPT markers m, as shown in Figure 4.2. Let us define the general Fisher's Ratio (FR) Matrix $FR \in \mathbb{R}^{m \times p}$ [203], where

$$FR_{jk} = \frac{(\mu_{R_{jk}} - \mu_{N_{jk}})^2}{\sigma^2_{R_{jk}} + \sigma^2_{N_{jk}}}, \tag{4.2}$$

for each IPT marker j in each percentile k, for $j = 1, .., m$ and $k = 1, ..., p$. In

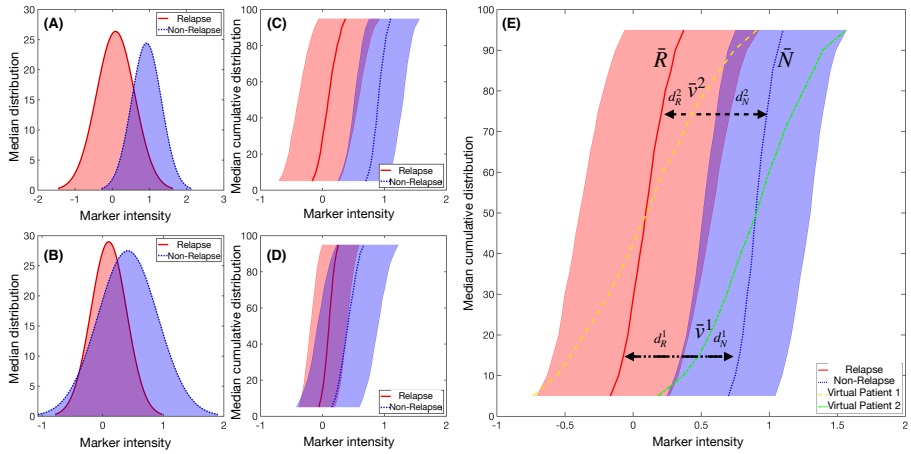


Figure 4.3: **Example of synthetic IPT markers distributions**. Mean distribution of a marker with respectively **(A)** high and **(B)** low Fisher's Ratio, with **(C)** and **(D)** their respective cumulative distribution of the median±the standard deviation values. **(E)** Median cumulative distribution of the two sets of patients for a marker with high Fisher's Ratio. In solid red line, median cumulative distribution of relapsed patients $\bar{R}$ and in blue dotted line for the non-relapse ones. In yellow dashed line and green dashed dotted line the median cumulative distribution for the marker $\bar{v}^i$ was represented for two different virtual patients i. The distances to each set median, d_R^i and d_N^i, are represented with black headed arrows, with dashed lines for Patient 1 and dashed dotted lines for Patient 2. In this example, Patient 1 would be considered as a relapsed patient, while Patient 2 would belong in the non-relapsed set.

this case, $\mu_{R_{jk}}$ and $\mu_{N_{jk}}$ are the median percentiles k for the IPT marker's distribution j in each class of patients. R stands for relapsing patients, while N refers to non-relapsing ones. Parameters $\sigma_{R_{jk}}$ and $\sigma_{N_{jk}}$ are the standard deviation measures within the classes.

To construct a classifier, we can select the highest FR_{jk} for $j = 1, ..., m$ and $k = 1, ..., p$, thus obtaining percentiles from several IPT markers with lowest deviation and highest difference in median between each subset. Thus we would obtain a general discriminant classifier of a maximum of $m^* \leq m$ markers and a maximum associated discriminant percentiles $p^* \leq p$.

In order to classify the patients, let us consider a certain IPT marker j and

percentiles k for each class of patients. We can associate it then to an specific central measure $\mu_{R_{jk}}$ or $\mu_{N_{jk}}$ and dispersion measure $\sigma_{R_{jk}}$ or $\sigma_{N_{jk}}$. Thus we set two control points

$$\begin{aligned} \bar{R}_{jk} &= \frac{\mu_{R_{jk}}}{\sigma_{R_{jk}}}, \\ \bar{N}_{jk} &= \frac{\mu_{N_{jk}}}{\sigma_{N_{jk}}}, \end{aligned} \tag{4.3}$$

of the central measure of the IPT marker for each class of patients, normalised by the dispersion measure. If we consider now a patient not assigned to any set and $\bar{v}_{jk}$ as the vector containing the percentiles k of a IPT marker j, we can compute a control point

$$\bar{P}_{jk} = \frac{\bar{v}_{jk}}{\left(\frac{\sigma_{R_{jk}}+\sigma_{N_{jk}}}{2}\right)}. \tag{4.4}$$

This point is normalised by the mean of both dispersion measures, as we consider $\bar{P}$ as a non-assigned patient control point. Now, we can use a distance function $d : \mathbb{R}^k \times \mathbb{R}^k \to [0, \infty)$ to measure the separation between the new patient $\bar{P}_{jk}$ and the control points $\bar{R}_{jk}$ and $\bar{R}_{jk}$ (Figure 4.3).

For each IPT marker, we construct a probability measure for each IPT marker and percentile $\mathcal{P}$ ias

$$\begin{aligned} \mathcal{P}\left(\bar{P}_{jk} \in R\right) &= \frac{d(\bar{P}_{jk}, \bar{R}_{jk})}{d(\bar{P}_{jk}, \bar{R}_{jk}) + d(\bar{P}_{jk}, \bar{N}_{jk})}, \\ \mathcal{P}\left(\bar{P}_{jk} \in N\right) &= \frac{d(\bar{P}_{jk}, \bar{N}_{jk})}{d(\bar{P}_{jk}, \bar{R}_{jk}) + d(\bar{P}_{jk}, \bar{N}_{jk})}. \end{aligned} \tag{4.5}$$

The mean of the probability measures for all the IPT markers selected for each patient may allow us to classify the patient in the relapsing or non-relapsing classes.

4.1.5 Validation and feature relevance

To validate the classification algorithm, both K-fold, and leave-one-out cross-validation techniques were applied. The resulting performance of each model was obtained by averaging over 20 evaluations each K-Fold, and 20 for Leave-One-Out cross-validation (LOOCV). Each technique was repeated in each evaluation to fully cover each data set. For both cases, a minimum of one patient of each set was always in the training set.

For each validation technique, we constructed a classifier with the most significant IPT markers j according to the Fisher Ratio ($FR_{jk} > 0.5$). We

computed in each simulation a receiver operating characteristic (ROC) curve and its associated Area Under Curve (AUC). Accuracy was obtained as the number of correctly classified samples divided by the total number of classified samples. Along with these magnitudes we computed from each confusion matrices the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

To choose the IPT markers with better prognostic value, we performed a Monte-Carlo based train-test split of the whole set of patients. We ran 100 simulations where each class of patients was divided into a 75% training and 25% test sets. We constructed a classifier for each splitting as described and then evaluated its accuracy. Once an accuracy threshold was fixed, we computed the frequency of every marker for the set of classifiers that were above that threshold. This calculation was performed for different values of the accuracy threshold.

Finally, we resorted to another method in order to compare the results. We performed 100 Random Forest classifications with 50 trees each and a 75-25 split of patients and then recorded the out-of-bag error and the permutation feature relevance.

4.2 Results

4.2.1 CD38 distribution differed significantly between relapsing and non-relapsing patients

We examined for Datasets 1 and 2 and for both of them combined those IPT markers with higher FR. Results are shown in Figure 4.4 and 4.5, as well as the median cumulative distribution of the arising markers.

For Dataset 1, CD38 FR was high in almost all percentiles, with $FR_{jk} > 0.3$, as seen in Figure 4.4 (A.1). IPT marker CD123 had high FR for the highest percentiles, with $FR_{jk} > 0.3$ for $k \in (50, 95)$. For Dataset 2, the differences between FR were significantly higher, with $FR_{jk} > 3.5$ in percentiles $j \in (20, 95)$ for IPT marker CD38, and mean $FR_{jk} > 2.5$ for IPT marker CD66c, as shown in Figure 4.4 (B.1.). For the combination of both datasets, only CD38 achieved a high FR with mean $FR_{jk} > 0.9$ in all percentiles, as shown in Figure 4.5 (C.1.).

Immunophenotypical markers CD38 and CD123, for dataset 1, and markers CD38 and CD66c, for dataset 2, predicted relapse after cross-validation. K-fold and Leave-One-Out cross-validations were run in

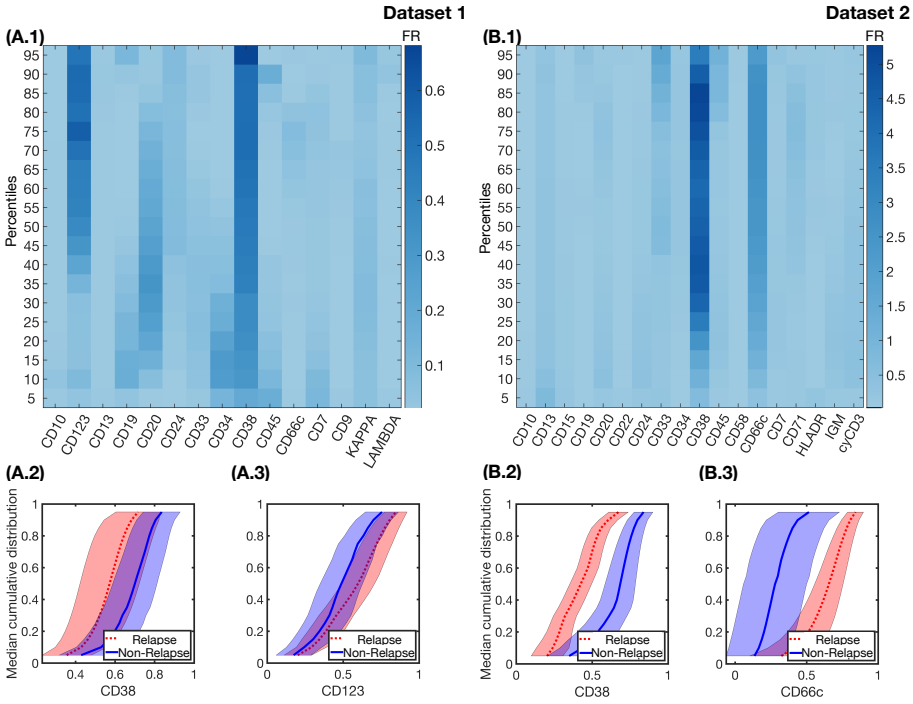


Figure 4.4: **Fisher's Ratio Matrices for Dataset 1 and 2 separatedly and median cumulative distributions of IPT markers with highest ratio**. Fisher's Ratio Matrices for Dataset 1 **(A.1)**, and 2 **(B.1)**. The common parameters within each dataset are represented in the x-axis, while in the y-axis we represent the percentiles of the median cumulative distribution. The intensity of the Fisher's Ratio for each percentile and markers are represented for each dataset in a colorbar for each chart. Median cumulative distributions and standard deviation bands of the IPT markers with higher FR, for relapsed (red, dotted lines) and non-relapsed (blue, solid lines) patients are represented in the following charts: for Dataset 1, CD38 **(A.2)** and CD123 **(A.3)**; for Dataset 2, CD38 **(B.2)** and CD66c **(B.3)**.

both directions: first, to know the most common IPT markers used in the training set for each simulation. This resulted in differences between median distribution differences of relapsed and non-relapsed patients again for the

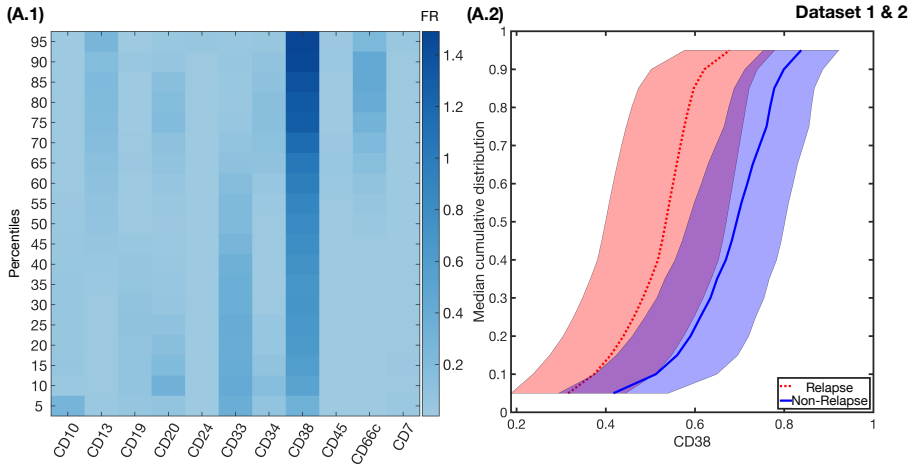


Figure 4.5: **Fisher's Ratio Matrices for Dataset 1 and 2 combined and median cumulative distributions of IPT marker with highest ratio. (A.1)** Fisher's Ratio Matrices Datasets 1 and 2 combined. The common parameters within each dataset are represented in the x-axis, while in the y-axis we represent the percentiles of the median cumulative distribution. The intensity of the Fisher's Ratio for each percentile and markers are represented for each dataset in a colorbar for each chart. **(A.2)** Median cumulative distributions and standard deviation bands of the IPT markers with higher FR, in this case CD38 for relapsed (red, dotted lines) and non-relapsed (blue, solid lines) patients are represented for datasets combined, CD38.

same sets of markers: CD38 and CD123 for Dataset 1 (Figures 4.4 (A.2.) and (A.3.)), CD38 and CD66c for Dataset 2 (Figures 4.4 (B.2.) and (B.3.)), and finally, only CD38 for the combination of both datasets (Figure 4.5 (C.2.)). Secondly, cross-validation techniques were repeated considering only common IPT markers. The results are shown in Table 4.1. The maximal number of folds was determined by the number of relapsing patients (8 for Dataset 1, and 5 for Dataset 2).

	Method	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
Set 1	LOOCV	0.75±0.04	0.74±0.05	0.76±0.05	0.76±0.04	0.75±0.04	0.76±0.02
	2-Fold	0.59±0.1	0.63±0.14	0.43±0.2	0.81±0.04	0.24±0.12	0.56±0.1
	4-Fold	0.62±0.07	0.63±0.1	0.58±0.12	0.85±0.03	0.3±0.06	0.65±0.06
	6-Fold	0.64±0.05	0.66±0.05	0.58±0.13	0.85±0.04	0.31±0.06	0.67±0.06
	8-Fold	0.7±0.04	0.7±0.04	0.71±0.06	0.9±0.02	0.39±0.04	0.72±0.03
Set 2	LOOCV	0.66±0.06	0.95±0.05	0.37±0.1	0.6±0.04	0.88±0.1	0.89±0.05
	2-Fold	0.72±0.07	0.95±0.06	0.13±0.22	0.74±0.05	0.42±0.41	0.68±0.16
	4-Fold.	0.78±0.04	0.95±0.05	0.34±0.15	0.79±0.03	0.81±0.2	0.86±0.06
Sets 1&2	LOOCV	0.69±0.05	0.62±0.09	0.75±0.09	0.72±0.07	0.67±0.05	0.78±0.04
	2-Fold	0.64±0.13	0.6±0.17	0.75±0.12	0.87±0.09	0.38±0.08	0.73±0.11
	4-Fold	0.69±0.01	0.67±0.02	0.77±0.01	0.91±0.01	0.41±0.01	0.77±0.04
	6-Fold	0.7±0.02	0.68±0.02	0.77±0.01	0.91±0.01	0.42±0.02	0.79±0.02
	8-Fold	0.7±0.01	0.68±0.02	0.77±0.01	0.91±0.01	0.42±0.02	0.79±0.02
	10-Fold	0.7±0.01	0.68±0.02	0.77±0.01	0.91±0.01	0.42±0.01	0.8±0.02
	12-Fold	0.69±0.01	0.67±0.02	0.77±0.01	0.91±0 .01	0.41±0.01	0.79±0.01

Table 4.1: Validated predictive performance of best classifiers CD38 and CD123 (Dataset 1), CD38 and CD66c (Dataset 2) and CD38 (Datasets 1& 2). PPV: Positive Predictive Value. NPV: Negative Predictive value. AUC: Area under curve.

4.2.2 Train-test splitting revealed other markers with potential predictive value

We tested the accuracy of the variables by splitting Dataset 1 and 2 combined into a training and test set with ratio 75:25. After 100 simulations, the frequency of the IPT markers used in the classifiers is shown in Figure 4.6(A). IPT marker CD38 arose again as the marker used in all classifiers, while CD33 was used on almost 70% of them. Having obtained the accuracy for the 100 classifiers, we count the number of IPT markers whose prediction accuracy is above a certain threshold, as shown in Figure 4.6(B). IPT markers CD13, CD24, CD33, CD38, CD45 and CD66c are those with an accuracy higher than 0.5%.

4.2.3 Random-Forest analysis matched the results from the constructed classifiers

Random Forest analysis after 100 simulations was considered, resulting in IPT markers CD33, CD38 and CD66c as those only with positive Out-of-bag Feature importance. However, after repeating the simulations only considering these markers, Out-Of Bag Classification Error was not significantly lower in comparison to the analysis with the whole set of IPT markers (mean out-of-bag error of 0.28 versus 0.31, respectively). Nevertheless, feature

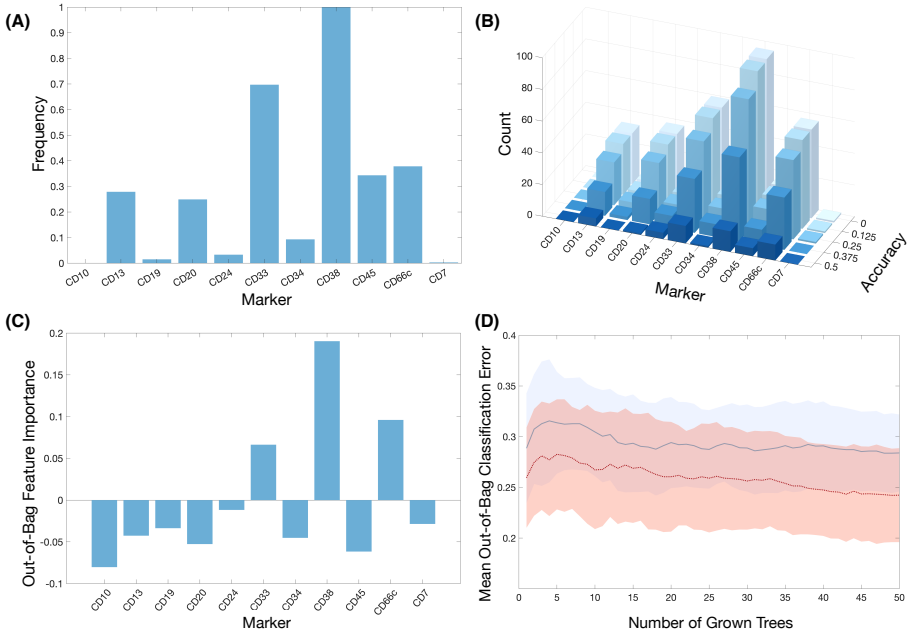


Figure 4.6: **Results for both datasets combined from train-test split and Random Forest analysis. (A)** Frequency of the markers in all classifiers after 100 simulations of train-test splitting. **(B)** Histograms of the number of markers after establishing a threshold for the accuracy. **(C)** Out-of-bag feature importance of the markers after 100 simulations of Random Forest. **(D)** Mean and standard deviation bands of the Out-of-bag Classification Error in Random Forest analysis for the whole set of markers (blue, solid line) and for the set of markers with positive feature importance CD33, CD38 and CD66c (red, dotted line).

importance coincided with those markers with highest frequency in our previous classification.

4.3 Discussion and conclusions

The unprecedented amount and complexity of clinical data that is available nowadays has resulted in the proliferation of bioinformatics pipelines and

artificial intelligence algorithms to extract information from data. In flow cytometry, the routine analysis that is carried out by visualizing histograms and bidimensional plots is falling behind technical progress in the field [24]. Machine learning algorithms have the potential to speed up, automatise and reduce bias in conventional analyses, but also to complement the work done by the human operator [23]. Recent examples in haematology include leukocyte recognition, prediction of refractory Hodgkin lymphoma, minimal residual disease detection in acute myeloid leukaemia, risk stratification in multiple myeloma or predicting resistance in myelodisplastic syndrome [204]. In childhood B-cell ALL, machine learning has taken advantage of clinical data in order to predict either diagnosis [205] or relapse [195], with the work of Good et al including proteomics data for the latter purpose [196]. Reiter et al proposed a way to automatise Minimal residual disease follow-up [194].

Leaving aside accuracy and prediction reliability concerns, which we can expect to be solved or dampened as the scientific production continues, there are a number of issues that still hamper the integration of AI and the respective clinical context. As happens in general with the relationship between mathematics and medicine, researchers in both ends often speak a different language [20]. Many AI algorithms behave as a "black boxes", providing an outcome directly from raw data and hindering a mechanistic interpretation of the underlying phenomena. For clinical use, it is highly desiderable that the features uncovered by these algorithms canbe interpretable and actionable. As Radakovich et. al. puts it, "Algorithms can only be as clinically meaningful as the outcomes that they are designed to predict" [206].

In this Chapter, we designed an intuitive algorithm allowing to identify on diagnosis patients with potential of relapse versus those with no risk of relapse in B-cell childhood ALL. We used flow cytometry data obtained at diagnosis from two local institutions and based the analysis on two concepts that are already employed in this context; the intensity and range of surface markers expression and the frequency of cells within that range. We took those two factors into account by assigning each patient and marker its percentile curve and then used the Fisher's ratio to look for meaningful differences between both groups of patients. That approach allowed us to construct a classifier based on this measure in order to assess the significance of the previously obtained differences. Given the small sample size, we used cross-validation routines to assess the validity of the Fisher's ratio-based measure. Despite the exploratory nature of the study, we were able to find some common trends in the data.

Firstly, we observed that Fisher's ratio displays differences in expression levels between relapsing and non relapsing patients. This was specially significant

for the second dataset. Given that both datasets were pre-processed identically, the difference in the significance of the measure could be due to either sample size or different acquisition routines in either hospital. We expect to have a clearer understanding of this after increasing the number of patients in each dataset or the number of datasets as a whole. K-fold cross validation showed that, when restricting the analysis to the most important features according to the previously calculated Fisher's ratio, the algorithm was able to separate better between relapsing and non relapsing patients, always using only data available on diagnosis.

Measurements of performance yielded good values for this biomarker as measured by Accuracy and AUC. However, although specificity was high, we obtained a low negative predictive value, i.e. the algorithm underperformed when detecting relapses. This could be due to the intrinsic unbalance in the datasets, with only 25% of relapsed patients. The relevant information, nonetheless, was the agreement in the extraction of the most important features. This was later confirmed by the Monte-Carlo based and Random Forest feature importance computation. Both approaches agreed in this selection, specially when being more restrictive with the classification accuracy in the first one.

The most consistent result, in the different analyses and for both local institutions, was the association between a lower expression of CD38 marker and relapse. CD38 is a surface receptor present in a broad variety of immune cells. It is considered to be a cell activation marker and operates both as a receptor and an enzyme [207]. In the B cell compartment, both bone marrow precursors and terminally differentiated cells express CD38 [208]. In the context of haematological disease, high CD38 levels have been associated with worse prognosis in chronic lymphocytic leukaemia [209]. Previous studies have suggested that CD38 is a suitable therapeutic target in both AML and ALL [210, 211]. There has been some controversy concerning the existence of a CD34+/CD38- population of leukaemia initiating stem cells [212, 213, 214, 215]. In B-ALL, the accumulated evidence indicates that lower levels of CD38 could be associated with worse outcome in terms of survival [216, 217, 218, 219]. Our results aligned with this evidence, suggesting that a higher frequency of low CD38 expressing B cells could be an early indicator of relapse risk.

Other markers that were found to be relevant in this study were CD33 and CD66c. These two markers are normally expressed in cells of the myeloid lineage, and they have been linked to paediatric B-ALL in the context of myeloid antigen expressing B-cell malignancies. This refers to the fact that some malignant B cells can express markers from the myeloid line. CD66c is the most frequently observed aberrant myeloid antigen in B-cell ALL. Upon studying the correlation

of the expression of this antigen with known prognostic factors, previous studies have found that CD66c is associated with BCR/ABL translocation, which has been shown to confer the worst prognosis [220, 221, 222]. Here, we found that patients relapsing were more prone to overexpression of this marker on diagnosis. With respect to CD33, there has been some controversy with respect to its prognostic value, but evidence suggests that the presence of high expressing CD33+ cells identifies patients with worse prognosis [223], contrary to the differences exhibited by percentile curves here.

Finally, the immunophenotypical marker CD123 was also highlighted by Fisher's ratio but only in dataset 1. Its importance could not be further assessed since it was not available in dataset 2. This marker was first described as a marker of acute myeloid leukaemia stem cells. It was later shown to be uniformly expressed in B-ALL blasts, being proposed for the detection of minimal residual disease [224, 225] and identified recently as a potential target for immunotherapies [226, 227]. Interestingly, high expression of CD123 correlated with hyperdiploid karyotype, an indicator of favourable prognosis in childhood B-cell ALL [228]. In our cohort we found a high proportion of CD123 expressing cells in relapsing patients.

While CD38 differences were present through the whole range of expression of the marker, that was not the case for CD33 and CD123. For those markers differences were observed only in the low expression region for the former and in the high expression region for the latter. The fact that there is less evidence for their prognostic value suggests that the method presented here leads to significant results if there is a constant difference in expression levels between both sets of patients. This is indeed a limitation of the study; by representing the expression as a percentile curve, we may miss information that can be clinically relevant and that refers not to the frequency of cells or intensity of expression, but to the presence or absence of a given subpopulation. In this regard, we already mentioned that a subpopulation of CD34+/CD38- cells could be associated with leukaemia initiating cells, and the same could happen for a restricted subpopulation of CD34+/CD38-/CD123+, this one agreeing with the results presented.

Another limitation of our analysis is the nature of the data, a recurrent concern in artificial intelligence in haematology [206]. Apart from having only 54 patients, the set of relapsing patients represented only 25% of the whole dataset and that unbalance could introduce biases in the analysis. In the future, as we increase the size of our dataset, it would be better to increase the number of patients to carry out a 50/50 analysis. Further, there is the issue of data variability, given that it was collected retrospectively, belonging to

patients from different years and hospitals. This highlights the importance of the preprocessing routine, which is also amenable to improvements in order to ensure the comparability of the samples. These weaknesses provide future lines of work. While in the process of recruiting more patients and hospitals, efforts will we directed towards the automation of the preprocessing workflow and towards the combination with more complex analysis like dimensionality reduction, network analysis and clustering. Finally, this work could be complemented with the inclusion of other clinical data like cytogenetics and molecular biology information, also relevant in the prognostic assessment of haematological diseases.

Notwithstanding these limitations, this works adds to the growing field of artificial intelligence in haematology and specifically in B-cell childhood acute lymphoblastic leukaemia. We attempted to delineate differences in marker expression between patients who relapse from the disease and those that respond to treatment, obtaining results that are directly interpretable from the clinical point of view. The main result would be the underexpression of surface marker CD38 in patients experiencing relapse after the first-line chemotherapy treatment. This is very important knowledge since it could aid in therapy personalisation by considering alternative therapies as upfront therapies for patients with high risk of relapse.

The methods, results and conclusions related to this Chapter can be found in Ref. [229].

CHAPTER 5

Topological data analysis and relapse prediction of flow cytometry data

Cancer is a heterogeneous disease at the genetic and phenotypic levels. Genetic heterogeneity depends on the diversity of genetic alterations in tumour cells, while phenotypic heterogeneity refers to the ways in which cells with a similar genetic background can exhibit differences in, for example, morphology, differentiation status, biomarker expression, or likelihood of therapeutic response. At the clinical level, heterogeneity manifests in differences in treatment responses, the emergence of resistance and the likelihood of relapse of histologically similar tumours [230].

The availability of quantitative data from multiple sources, including multi-omics, high resolution images and flow cytometry may enable this complexity to be understood. Major challenges of personalised medicine in oncology that can be assisted by data analysis include patient phenotyping, biomarker discovery and prediction of response and relapse [231, 232].

Topological data analysis (TDA) approaches the analysis of datasets using methods from topology. The datasets available in oncology are usually high-dimensional, noisy and incomplete, which represents a significant challenge for conventional data-analysis methods. Advantages of TDA are that it is not sensitive to the metrics chosen and that it provides dimensionality reduction and robustness to noise [233]. Persistence homology (PH) is a powerful tool from TDA [234] allowing the construction of topological invariants, which can be interpreted as characteristics of shapes, in high-dimensional data. The

data space is simplified by using structures, the so-called simplicial complexes. The features analysed include connected components in dimension 0, loops in dimension 1 and holes in higher dimensions. Topological features in a nested sequence of simplicial complexes, the so-called filtrations are represented by intervals in a barcode. These diagrams allow to understand the lifetime, or *persistence*, of a topological feature. This enables PH to describe large noisy datasets. PH has a myriad of emerging applications (see e.g.[235, 236, 237, 238, 239, 240, 241] and references therein). While there have been fewer applications in cancer, it has been used to construct new biomarkers [242, 243, 244] and to classify tumours according to histological architecture [245] or genetic alterations [246].

In this Chapter we combine methods from TDA and machine learning (ML) to predict relapse in cancer. As a specific example we focus on Acute Lymphoblastic Leukaemia (ALL), which is the most frequent type of paediatric cancer [247]. Leukaemias are cancers in which lymphocytes grow abnormally in the bone marrow, blood or lymphatic system. B-lymphocyte ALL is the most common subtype in paediatric patients [248] and is characterised by rapid growth of early lymphocytes in the bone marrow. Although current chemotherapy regimes have improved survival rates, still more than 20% of patients relapse. It is important, on diagnosis, to identify patients with at high risk of relapse so that instead of first-line chemotherapy regimes, they receive alternative treatments such as CAR-T cells, more intensive chemotherapeutic regimes, or undergo bone-marrow transplants. Beyond this example, estimating the likelihood of relapse before first-line treatments is of the utmost importance in Oncology.

Currently, initial risk assessments are based on a combination of biomolecular, cell morphology and genetic analyses [249]. A quantitative method used routinely for diagnosis in haematological malignancies is flow cytometry [250]. It provides large amounts of high-dimensional data, which is currently used in a limited way, as routine analysis are performed mostly manually and analysing visually 2-dimensional projections of the data [23].

The focus here is on anticipating the risk of relapse in paediatric ALL patients, but the methodology we developed in this Chapter could be readily extended to other haematological malignancies, including leukaemias or lymphomas where similar diagnostic and therapeutic options are available [250].

5.1 Materials and Methods

5.1.1 Patients

A retrospective study was designed and approved by the institutional review board (IRB) of the four participating local institutions. Inclusion criteria for the study were ALL diagnosis between February 2009 and October 2017, age over 1 year and less than 19 years, and availability of bone marrow flow cytometry data. A total of 105 patients satisfied the inclusion criteria. Exclusion criteria for these patients were availability of Flow Cytometry files FCS below 3.0, patients without a minimum of 15 IPT markers in common with others in the dataset, and insufficient follow-up for non-relapsing patients, i.e. patients without relapse but with less than three years after no refractory values for minimal residual disease were found. Finally, 54 patients diagnosed in two of the local institutions were retained for further analysis. Dataset 1 included 28 non-relapsed patients and 8 relapsed, while dataset 2 included 13 non-relapsed patients and 5 relapsed.

5.1.2 Flow cytometers and antibodies

Marker expression was obtained on FACSCanto II flow cytometers, in accordance with the manufacturer's specifications for sample preparation. Final samples were stained using an 8-colour panel with six fluorochrome-conjugated antibodies.

5.1.3 Flow cytometry markers and data preprocessing

FCS 3.0 files contained the following information: forward scatter (FSC) (interpreted as size), and side scatter (SSC) light (interpreted as complexity), as well as several fluorochromed antibodies acting against cell proteins, as immunohistochemical markers. The diagnosis included several B-cell-related (CD19, CD10, CD20, CD22, CD24, IgM, CD66c, CD79a, kappa, lambda, etc.) and T-cell-related IPT markers (CD7, cyCD3), as well as more general ones (CD15, CD34, CD38, CD45, CD58, CD71, HLA-DR), and IPT markers related to the myeloid lineage (CD9, CD13, CD33, CD123).

The flow cytometry files were preprocessed prior to analysis. Certain events were removed, such as events at the margins (measurements that match the maximum or minimum values for any parameter), doublets (cells that are accidentally analysed together as a single event), or debris and dead cells [23],

as shown in Figure 5.1(A). Since gating was performed manually, geometric markers were ommited from the persistence analysis, to prevent unconscious bias. Next, we randomly selected 10^5 points from the whole data cloud, to establish a common minimum number of points, and thus have the same number of events for all patients. Tubes were compensated by means of the spillover matrix included in each file, and then transformed with the Logicle transformation [197]. Nearest-neighbour imputation, as in [198], was used to merge all tubes or aliquots into a single file for each patient. We chose the 0.05 and 0.95 quantiles for each marker x and used the transformation $x' = (x - x_{q0.05})/(x_{q0.95} - x_{q0.05})$ to avoid outliers, where $x_{q0.05}$ is the 5^{th} percentile and $x_{q0.95}$ is the 95^{th} percentile. We identified the B-lymphocyte cloud, by selecting $CD19^+$ cells, as this is the IPT marker associated with early B cells (Figure 5.1(B)). Finally, 10^4 landmarks were selected from the lymphocyte cloud by applying the maxmin algorithm [251] to each group of markers considered (see Figure 5.1(C)).

Persistence analysis on pairwise combinations of all IPT markers, was performed in dimensions 0 and 1. Because of computational constraints, the number of landmarks was reduced to 10^3 for studies performed in dimension 2.

5.1.4 TDA and Persistence Images

Given M points $x_i \in \mathbb{R}^n$ as in Figure 5.2(A), and a distance function d, we can place a ball of radius r at the centre of each point. For a given point x_i, we identify all points closer than r by $b_r(x_i) = \{y \in \mathbb{R}^n : d(x_i, y) \leq r\}$ (see e.g. Figure 5.2(B)). The set of points contained in the union of all these balls $\cup_{i=1}^{M} b_r(x_i)$ depends on the radius r. This parameter allows the construction of simplicial complexes along a nested sequence of them, the so-called filtration. A simplicial complex is a set of structures able to capture the number of "holes" in a certain dimension. A homology group over this filtration can provide us information about how close or connected points are for dimension 0, and how loops are created in higher dimensions. These are the so-called Betti numbers β_i for each dimension i. Topological features, in this case, one "hole" of dimension i, that appear at a filtration "time" r_{birth} and disappear at a time r_{death} can be considered to have a "lifetime" $\rho_r = r_{\text{death}} - r_{\text{birth}}$, the so-called persistence. Persistence can be represented by barcodes, where intervals for each filtration time are represented, as observed in Figure 5.2(C) and (D) for dimension 0 and dimension 1, respectively.

In our study, we focused on persistence images, which rely on persistence barcodes. These images can be considered as real-valued matrices that can be

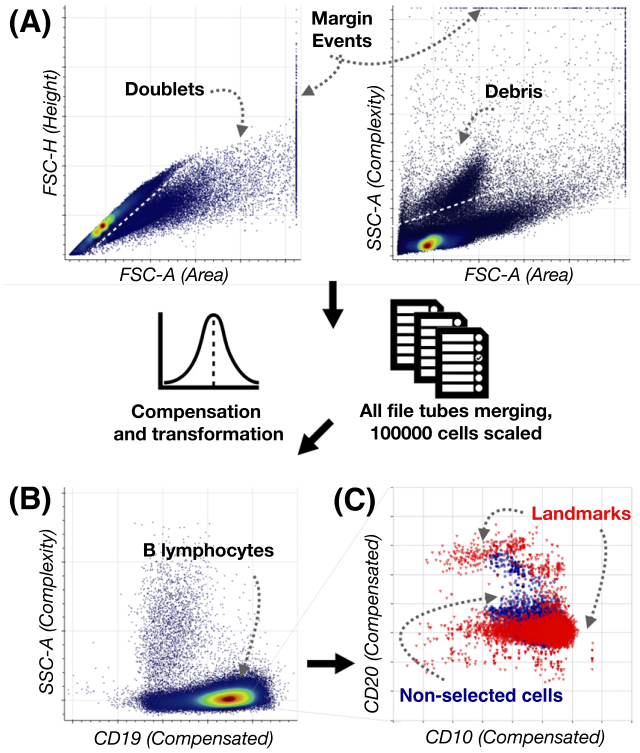


Figure 5.1: **Pipeline of the flow cytometry data preprocessing.** **(A)** Manual gating process for the selection of B lymphocytes. First, doublets, i.e., two or more cells considered as single events, are removed from the panels FSC-A and FSC-H, below the white dashed line. Secondly, debris is removed from the panel FSC-A and SSC-A, omitting highly complex cells over the white dashed line. In both panels, events at the margins (located at the extreme of the axes) are removed. Next, single files are compensated and transformed, and finally merged into one file per patient. A number of 10^5 cells is taken randomly and scaled to avoid outliers. **(B)** B lymphocytes are selected as $CD19^+$ cells. **(C)** The maxmin algorithm is applied to obtain the shape structure of the B lymphocyte point cloud. Using 10^4 selected landmarks gives a plausible representation of the whole B lymphocyte dataset.

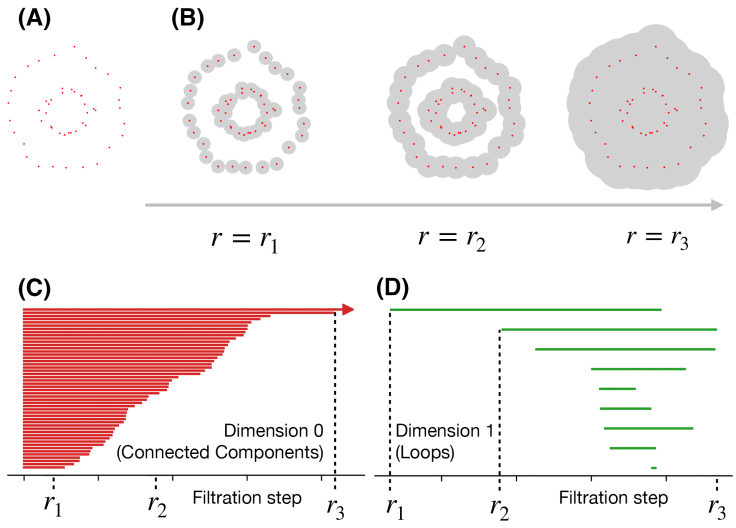


Figure 5.2: **Persistence topological analysis of flow-cytometry data.** **(A)** Example data illustrating the basis of persistence analysis. **(B)** When increasing the radius r of the grey balls centred at data points, topological features arise. For $r = r_1$, a one-dimensional loop appears, and a second one for $r = r_2$. For $r = r_3$, all points are in the same connected component. **(C)** Persistence barcode for dimension 0 of data in (A). The top bar is the only one for $r > r_3$ and persists towards infinity. **(D)** Persistence barcode for dimension 1 of data in (A). Persistence barcodes represent one-dimensional loops, where the longest are "born" at a filtration step $r = r_1$ and $r = r_2$.

used as an input into a variety of machine-learning approaches. Another common representation of these data is through the use of persistence diagrams, usually in coordinates $(r_{\text{birth}}, r_{\text{death}})$, as shown in Figure 5.3(A). In these diagrams, points close to the diagonal represent short-persistent features. To create the persistence images, using coordinates $(r_{\text{birth}}, \rho_r)$ from persistence barcodes (as in Figure 5.3(B)), 2D Gaussian distributions (in this study, with variances 0.01 and 0.05) were generated at each point, resulting in a surface (see e.g. Figure 5.3(C)). Its internal volume can be summed and displayed as a uniformly-spaced grid resulting in persistence images, which are stable under small perturbations to the inputs [252]. To check for differences, in our study we used grids of

resolution 50×50 and 100×100. An example of a persistence image is shown in Figure 5.3(D). Further details of TDA persistence homology group methodology can be found in [234, 252].

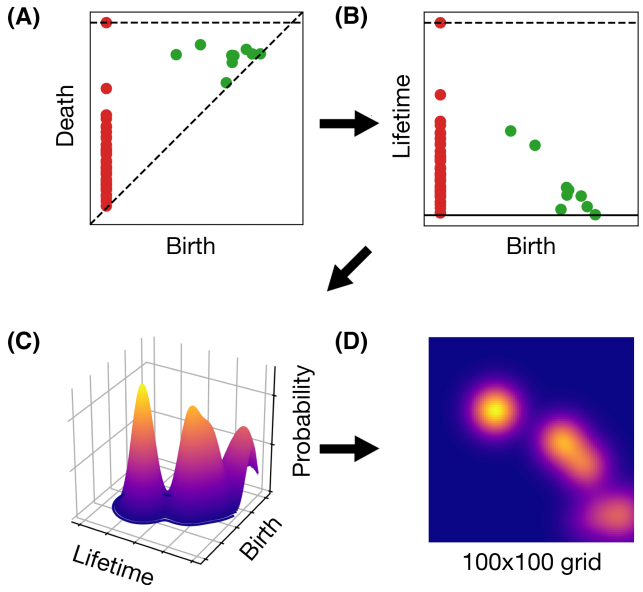


Figure 5.3: **Pipeline for the obtention of persistence images. (A)** From the persistence barcodes, e.g. those in Figure 5.2(C-D), persistence diagrams are constructed. The dashed, diagonal line represents short-persistence barcodes, while the dashed, horizontal line represents infinity. Dimension 0 (red points), and 1 (green points) are computed. **(B)** To obtain a persistence image, the next step is to consider the persistence diagram in coordinates (birth, lifetime). The dashed, horizontal line represents infinity, while the dashed, solid line represents short-persistence barcodes. **(C)** 2D Gaussian distributions can be generated at each point, in this case, for dimension 1 of the persistence diagram (B), thus obtaining a surface. **(D)** A persistence image resulting from summing the volume within the surface (C).

5.1.5 Machine learning methods

For classification and prediction purposes, we used Random Forest techniques, a popular and efficient algorithm based on model aggregation ideas [253]. Support Vector Machines (SVM) [254] were also used for classification by assigning relapsing and non-relapsing labels to the data matrices obtained from persistence images. Two parameters associated to the SVM classification kernel were considered: γ as the curvature of the decision boundary, and c as the trade-off between correct classification and the distance between the decision boundary and support vectors. They were selected by randomly splitting the training set and performing internal cross-validations for a logarithmic range of the parameters $c \in [10^{-2}, 10^{13}]$ and $\gamma \in [10^{-9}, 10^{3}]$. We also used Logistic Regression to construct binary regressions. The models were scored by computing the mean score of fitting the model after 5-fold cross-validation and the coefficient of determination R^2, interpreted as the accuracy of the prediction.

5.1.6 Computing machines

MaxMin algorithm and Persistence Analysis were run on six machines from the Oxford Mathematical Institute, each with 36 cores, with up to 3.9 Ghz speed, and 768 GB of RAM. Figures and other calculations were run on an iMac, running under Mac OS 10.15, with four i5 cores, 3.4 Ghz speed and 32 GB RAM.

5.1.7 Software

Python (3.1) was used for the computations. R (3.6.0) and RStudio (1.2.1335) were used for data curation. Persistence barcodes and images were constructed using Ripser (0.3.2) [255] and Persim packages (0.1.3) in the Python Scikit-TDA toolbox. FCS files were read using the Python CytoFlow library (1.0) and Flow Cytometry libraries from Bioconductor (3.11) in RStudio. FlowJo (Becton Dickinson, 10.6.1) and FACSDiva (Becton Dickinson, 8.0.1) software packages were used for manual gating of FCS data.

5.2 Results

5.2.1 TDA identifies parameters with low information content

We selected immunophenotypic (IPT) parameters that were common between our two patient datasets (dataset 1 and 2), as described in 'Methods' [256] (see Table B.1).

Our first analysis was intended to discard IPT parameters with no essential information on relapse. For dataset 1, we found 15 common IPT characteristics between all patients, while in dataset 2, 18 common IPT markers arose, mostly different to those in dataset 1. All pairwise combinations of these parameters were constructed for each dataset, obtaining 2D projections of the IPT markers. These were analysed via persistence homology in dimensions 0 and 1. This resulted in 105 and 153 persistence barcodes, respectively, for dataset 1 an dataset 2. As shown in Figure 5.2(A)-(B), the radius of any ball centred at each point of any dataset can be increased, and so the lifetime, or persistence, of topological features can be studied. This results in the persistence barcodes shown in Figure 5.2(C)-(D), where intervals are represented with bars denoting the lifetimes of cycles in each dimension.

For each patient in each dataset, we computed the maximum, minimum, standard deviation, mean and median persistence in dimensions 0 and 1. We then performed a Random Forest analysis with cross-validation on each dataset and patient subgroup (i.e. relapsing and non-relapsing), assigning 60% of the patients to a training group, and 40% to the validation group. Receiver operating characteristic (ROC) curves were obtained for the classification of relapsing and non-relapsing patients. We identified IPT pairs with one area under the ROC curve of less than 50% as having very low information and excluded them from the subsequent analysis. The individual IPT markers appearing most frequently in the remaining pairs were CD10, CD123, CD20, CD34, CD38, CD45, CD66c, CD9 and kappa for dataset 1, and CD13, CD15, CD33, CD34, CD45, CD58, CD66c, CD7, CD71, IgM and cyCD3 for dataset 2 (see Tables B.2-B.4 and B.5 -B.6, respectively). The average of the mean areas under the ROC curves for the remaining parameter combinations was 58% in dataset 1, and 57% in dataset 2, thus having low discriminatory power. To increase it, we now consider the analysis of both datasets combined, thus enlarging the number of patients in one combined dataset.

5.2.2 Random Forest analysis of merged datasets provided biologically grounded biomarkers

To enlarge the number of patients included in each analysis, we studied the 11 IPT markers present in both datasets. Their pairwise combinations resulted in 55 persistence barcodes that were analysed via a Random Forest method with cross-validation. The biomarkers with AUC>50% were CD10, CD20, CD38, CD45 and CD7. Only combinations including CD38 or CD20 had a mean AUC>60%.

Since these AUC numbers are low, we would not consider the markers to have strong classification power. However, it is noteworthy that the identified markers are considered to be biologically relevant, as they are usually used on diagnosis together with others such as CD22, CD24, CD34, CD79a, IgM, or TdT. CD10, CD20 and CD45 characterise the shape of the $CD19^+$ B lymphocyte cloud [18]; CD38 is a marker of B-lymphoblast aberrance [219, 257]; while CD7 has been shown to be aberrant in lineage-switching leukaemias [258]. Thus, the five biomarkers were retained for subsequent analyses.

5.2.3 Relapsing patients had more connected components but fewer one-dimensional loops

Persistence analysis was performed for CD10, CD20, CD38 and CD45 in dimensions 0 and 1 (Appendix B shows also results including CD7 in Figures B.1 and B.2).

Statistical analysis of the persistence features for relapsed and non-relapsed patients was performed and the results are summarised in Figure 5.4 (A). The p-values obtained when comparing both groups were larger than 0.05. Thus, in general, these features did not discriminate between the two groups.

Next, we obtained the number of topological features in dimension 0 and 1, the so-called Betti numbers β_0 and β_1, and averaged them for each cohort of patients. We did so by computing, in the persistence barcodes, the mean number of persistence bar values longer than a certain filtration step threshold. This led us to the results shown in Figure 5.4(B). Figure 5.4(C) shows that relapsed patients had a high number of connected components in the range $j \in [0.3, 1.2]$, thus larger β_0 (t-test $p < 0.0003$), and a lower number of one-dimensional loops (β_1), (t-test $p < 0.04$) in the range $j \in [0.11, 0.35]$.

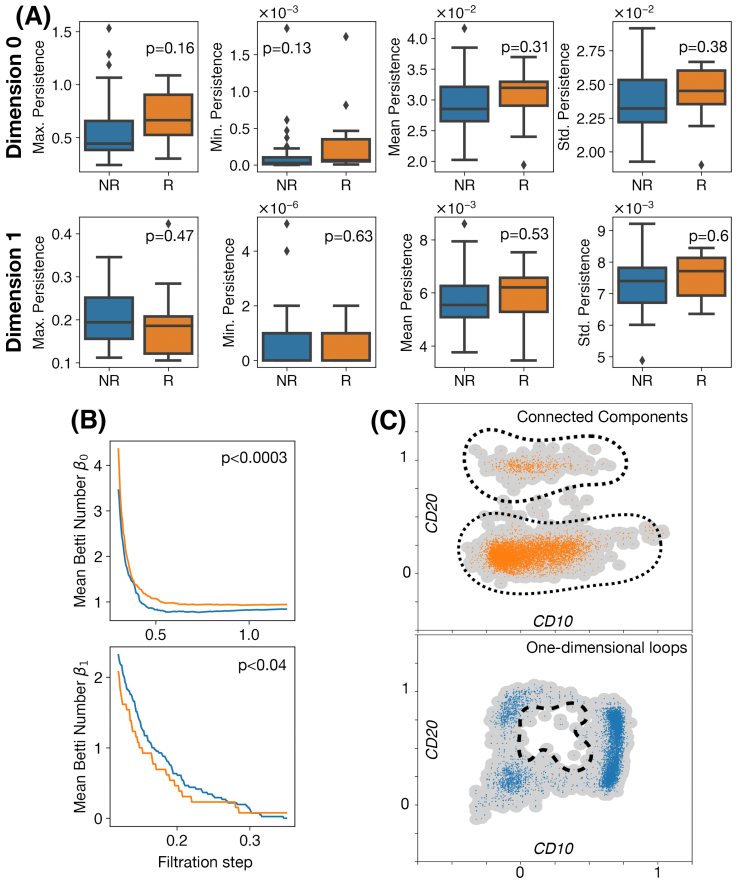


Figure 5.4: **Statistical analysis of persistence topology in flow-cytometry data. (A)** Boxplots of maximal, minimal, mean, and standard deviation persistence of non-relapsed (blue) and relapsed patient data (orange). **(F)** Difference in mean Betti numbers for relapsed (orange lines) and non-relapsed patients (blue lines) for certain filtration step ranges. **(B)** Interpretation of Betti numbers. Relapsed patients (orange) would have a higher number of connected components (dotted lines), while non-relapsed (blue) would have a higher number of one-dimensional loops (dashed lines). The grey areas represent balls of a certain radius r centred at each point.

5.2.4 Persistence images differ in shape between cohorts

Persistence images were obtained by transforming the persistence barcodes into a grid of $n \times n$ pixels, and then placing a 2D Gaussian distribution at each point of the persistence diagrams, as shown in Figure 5.3(A)-(D). These images show information about topological features, are stable under noise and maintain an interpretable connection to the persistence diagrams that they were obtained from. Results are shown for the grid of 100×100 and deviation of 0.05 of the generating 2D Gaussian Distribution (results for 50×50 and a spread of 0.01 are provided in Appendix B). Mean persistence images shown in Figure 5.5(A) are centred around (0,0). The spread influences the width of the regions of high intensity levels. For dimension 0, non-relapsed patients had broader intensity profiles, and relapsed patients less so. For dimension 1, relapsed patients had more spread profiles.

5.2.5 Persistence images allow perfect identification of relapsing patients

Finally, we considered the images as data matrices and used the raw data to perform Logistic Regression (LR) and Support Vector Machine (SVM) analyses. Studies were carried out dependent on three variables: datasets included (1, 2 or both), dimensions of the topological features analysed (0, 1 or 2) and markers included (CD10, CD20, CD38, CD45 with or without CD7).

		Dim. 0		Dim. 1		Dim. 2	
Dataset	Method	Score	Acc.	Score	Acc.	Score	Acc.
Dataset 1	LR	1	1	0.78	-0.29	0.78	-0.29
	SVM	1	1	1	1	0.78	-0.29
Dataset 2	LR	1	1	0.93	0.72	0.72	-0.38
	SVM	1	1	0.93	0.72	0.83	0.17
Both	LR	1	1	0.98	0.9	0.7	-0.32
	SVM	1	1	1	1	0.98	0.9

Table 5.1: Comparison of classification scores for Logistic Regression (LR) and Support Vector Machine (SVM) between datasets and dimensions using biomarkers CD10, CD20, CD38, and CD45.

A summary of the classification results is presented in Table 5.1 (see Appendix B Tables B.8-B.9 and Figures B.3-B.4 for detailed results, including

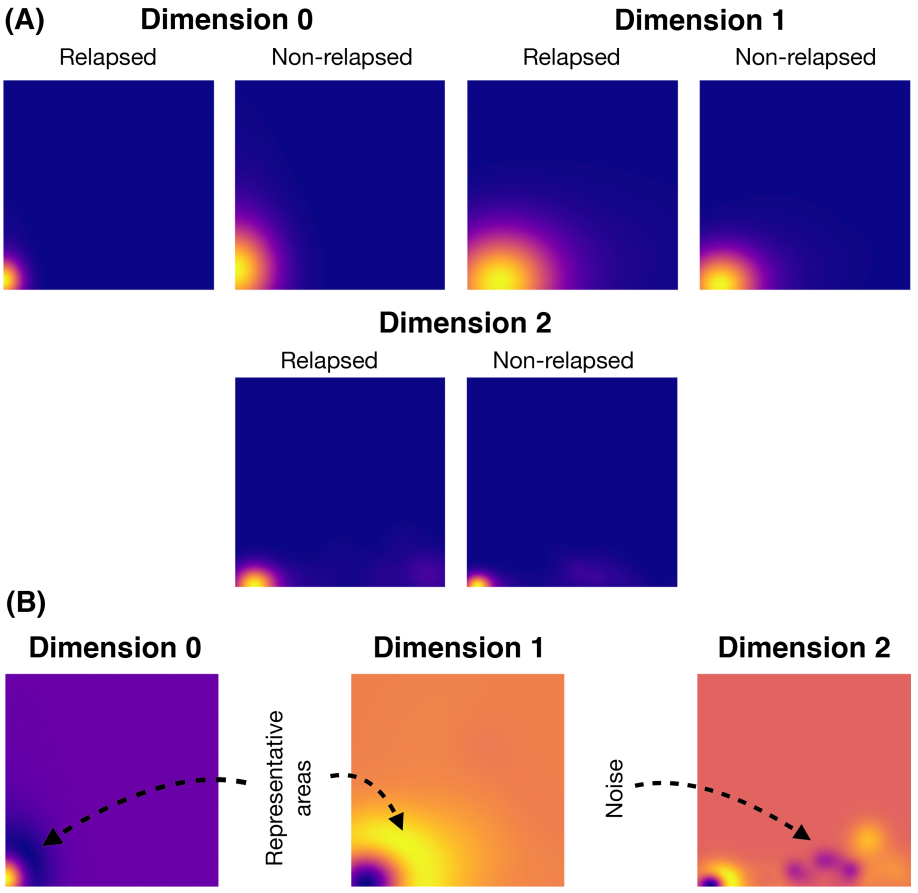


Figure 5.5: **Persistence images for relapsed and non-relapsed patients.** **(A)** Mean persistence images for each patient set, depending on dimension analysed, with a spread of 0.05 in the 2D Gaussian distribution on a 100×100 grid. **(B)** Weights assigned to the pixels of persistence images after SVM classifications. Dimension 0 focuses on the band around the centre, while dimension 1 does it on the centre itself, both dimensions being consistent in the representative areas shown. For dimension 2, other areas are highlighted, and are interpreted as noise, given the classification results shown in Table 5.1.

those with biomarker CD7 and Leave-One-Out cross-validations). We state the mean fitting score of the model after 5-fold cross-validation (Score) and the accuracy of the prediction via the coefficient of determination R^2 (Acc.). Cross-validation was performed within the training sets to obtain representative SVM parameters as described in 'Methods'. Results for the most representative and discriminatory areas are shown in Figure 5.5(B). Perfect classification scores and 100% accuracy were obtained for both datasets and set of biomarkers. Classification results were better for dimensions 0 and 1.

5.3 Discussion and conclusions

Relapse after treatment is a common problem in cancer. Its early detection is one of the main goals of follow-up procedures used in clinics. This is a general problem, but of specific interest in tumour types for which effective second- or third-line treatments are available. In paediatric ALLs, current first-line chemotherapy regimes allow for cure of about 80% of affected children. It is not necessary to proceed to further treatments for patients with no risk of relapse. However, a substantial fraction of them, larger than 20%, ultimately relapse. Early identification of these patients is of great importance, in order to consider other therapeutic options, such as bone marrow transplants or CAR-T cells. However, for those with a high risk of relapse it may be crucial to initiate further treatments sequentially before resistant clones take over and regrow [259, 260]. Also, using more aggressive therapies as a first line treatment could be an option for such patients. Other alternatives could include CAR-T therapies, with the drawback of costing up to 475000 $ per dose [261].

Given the depth of biological data provided by flow cytometry techniques, the use of mathematical analysis to extract relevant information may provide additional insight into the processes governing tumour growth. Our goal was to quantify the differences between relapsed and non-relapsed ALL patients on the basis of data available at diagnosis. We combined topological methods, specifically persistent homology and machine-learning techniques, to distinguish relapsing from non-relapsing patients with complete certainty.

The analysis based on CD10, CD20, CD38 and CD45 expression allowed for a perfect classification of patients according to their relapse status. It is relevant that these markers are available during routine diagnosis. CD10, CD20 and CD45 are used to gate early B cell subpopulations, and CD7 and CD38 are known to have prognostic value.

Connected components and one-dimensional loops distinguished via persist-

ence homology between relapsed and non-relapsed patients, as shown in Figure 5.4(C-D). We raised the possibility for relapsed patients to have a larger number of connected components, implying that isolated leukaemic point clouds could be a leading cause for relapse. Conversely, the number of one-dimensional loops was higher in non-relapsing patients, suggesting an anomalous development in the relapsing ones, with denser and more compact point clouds.

Persistence images allowed us to classify patients using SVMs and LRs. The centre of the distribution at (0,0), as well as the spread of the distribution, for dimensions 0 and 1, were held to be the most differentiating areas between both cohorts (see Figure 5.5(B)). Other zones were highlighted in dimension 2, which had a negative influence on the classification and were interpreted as being noise.

Dataset 1 was on average better classified in comparison to dataset 2. This may be because 50% fewer patients were inclueded in dataset 2 than in dataset 1. However, when the datasets were combined, the classification was excellent in both cases and for every dimension. Additionally, classification scores were in average higher for dimension 0 in comparison to dimension 1, and still lower for dimension 2. We note that, due to the high computational costs of PH in dimension 2, its analysis used 10^3 points, rather than the 10^4 points used in dimensions 0 and 1. Regardless of the method, dimension 0 attained a 100% classification score. Similarly, for dimension 1, the results showed a high score for almost every analysis. The study of two-dimensional loops resulted in a worse classification, with low and even negative determination coefficients. However, SVM analysis for both datasets achieved a 98% classification score as well as a 90% accuracy when considering only CD10, CD20, CD38 and CD45.

This study has several limitations. The limited amount of data could make the results sensitive to overfitting. Despite the robustness to noise of topological methods, it would be necessary to validate the results on different datasets. Also, a more general study could be performed by using bone marrow samples with larger numbers of common IPT markers. In that sense, and even if we obtained the samples and followed the same common hospital protocols to preprocess the data, the use of a single flow cytometer could benefit the accuracy of the study, as these machines differ on data scales and compensation techniques.

This work can be extended in many directions. Firstly, the shape of data after chemotherapy, transplant or CAR-T cell infusion could reveal significant information about the development of resistances to those therapies. Secondly, the flow-cytometry methods can provide additional information on the bone marrow cells and not only on the B-lymphocytes. The topological study could be applied to other haematological conditions, such as lymphomas, T-

cell leukaemias, myeloblastic disorders, etc. Finally, by combining valuable information from the diagnosis, not only IPT cells markers, but also biomolecular or morphology features, as in [262, 263], could improve patient prognosis and relapse prediction, and could be used for therapy personalisation.

We expect the topological point of view to reveal new disease processes hidden in the multiple, rich biomedical data available with current diagnostic methods. Our work can stimulate further studies on the potential of mathematical methodologies to unravel the unknowns of cancer relapse.

The methods, results and conclusions related to this Chapter can be found in Ref. [264].

PART III

Other mathematical works in tumour modelling

CHAPTER 6

Lie Symmetries and mathematical models of tumour development

A huge variety of phenomena are governed by ordinary differential equations (ODEs) and partial differential equations (PDEs). However, there is no general method to solve them. Obtaining solutions for differential equations is one of the greatest problem for both applied mathematics and physics. Several integration methods have been developed to the day to solve special classes of differential equations, specially those focused on physical or biological phenomena.

In this Chapter, we present three applications of the Lie method to obtain solutions of differential equations relevant in terms of its applicability in cell dynamics and tumour invasion.

6.1 The Fisher equation: some studied problems

Reaction-diffusion equations are a fundamental part in modelling the spread of biological populations. Originally introduced in seminal papers by Fisher [265] and Kolmogorov et al. [266], the mathematical framework based on reaction-diffusion equations has proven to be extremely useful for studies on population dynamics [267, 268], as well as mathematical biology in general [267, 269, 270, 271, 272]. The Fisher equation and its extensions are a family of reaction-diffusion models arising prominently also in cancer modelling [273, 274], applications to brain tumour dynamics [275], in the description of propagating crystallisation/polymerisation fronts [276], chemical kinetics [277], geochemistry [278] and many others fields. These equations have already been deeply analysed in the literature [267, 279, 280], in relation to their solutions and travelling

waves for the case of the Fisher equation:

$$u_t = D\, u_{xx} + \rho u(1-u). \tag{6.1}$$

This equation describes the change of the amount of cells $u = u(x,t)$ in time t and space x, for a diffusion term D and a reproduction rate ρ.

The study of this kind of equations is interesting in terms of finding exact solutions. In [281], for example, a numerical analysis was performed. To do so, the Lie classical method is useful to obtain reductions to ODEs and if it is possible, families of exact solutions [282]. One of the most famous and established procedures for obtaining exact solutions of differential equations is the classical symmetries method, also called group analysis. The investigation of symmetries has manifested as one of the most significant and fundamental methods in almost every branch of science, including mathematics and physics. Among the many papers using this method are [283, 284, 285, 286, 287].

In this study, by using equivalence transformations and Lie symmetry groups, we find analytical solutions for several models with biological interpretations. Lie symmetries of the density dependent reaction-diffusion equation

$$u_t = f(u) + \big(g(u)u_x\big)_x \tag{6.2}$$

were calculated in [288], as well as the optimal system of one-dimensional subalgebras of the invariant equation. Several reductions and exacts solutions were also obtained. In [289], some non-trivial conservation laws were constructed for the generalised Fisher Eq. (6.2) associated with symmetries of the differential equations. A non-linear multidimensional reaction-diffusion system with variables diffusivities was also considered in [290]. In this paper, the classical Lie symmetry of this system is calculated. In fact, this equation is not only interesting for cancer models and mutating cells, but also in biochemical reaction kinetics such as the effect of haemoglobin and myoglobin in blood [267].

Over the last decades a lot of attention has been paid on using Lie point symmetry methods to exploit the invariance of the generalised equation

$$u_t = (A(u)u_x)_x + B(u)u_x + C(u). \tag{6.3}$$

In the case $A = 1$, $B = C = 0$, the classical heat equation was firstly studied by S. Lie in [291] in terms of maximal invariance algebra. A complete Lie symmetry classification for the non-linear heat equation (6.3) with $B = C = 0$ was described in [292]. Moreover, for the case $B = 0$ in Eq. (6.3) the Lie symmetry was completely described in [293]. Later, the Lie symmetries of Eq. (6.3) were fully described in [294].

Generalisations of the Fisher equation are necessary to accurately model diffusion and reaction effects. Moreover, there are very few studies that search for exact solutions in cancer models. In [295], an analytical explicit solution was presented for a general two-type birth-death branching process with one-way mutation. In [296], some exact solutions were derived for a model of growth and movement of certain cell cultures as well as solid tumours in response to an arbitrary distribution of nutrients. Therefore, we consider several generalisations of the Fisher Eq. (6.1) and study them to obtain solutions related to tumour dynamics. Thus, we analyse the following cases:

- A generalised Fisher equation with density-space-dependent diffusion in the present manuscript as

$$u_t = \left(g(u)c(x)u_x\right)_x + f(u) \tag{6.4}$$

  which arises in a broad range of biological processes [280] and specifically in cancer modelling problems [281]. To illustrate the latter, a particular case of this mathematical model (6.4) was introduced by [273] to study the complex geometry of the brain and to allow diffusion (or cell motility). Furthermore, Eq. (6.4) has also been studied in [275] for a space-dependent diffusion term, in order to describe malignancy of gliomas as an invasion of grey matter.

- One biological hypothesis claimed is that new mutations may arise as the tumour grows, which is considered as a proliferative advantage [297]. This phenomenon can be introduced into the Fisher's Eq. (6.1) considering the following equation

$$u_t = \triangle u + \left(1 + \delta \int_{\mathbb{R}^n} u\, dx\right) u\,(1-u), n = 1, 2, 3, \tag{6.5}$$

  where $\triangle u$ is a Laplacian term, and the proliferation rate is more generalised, as it considers a new term describing the total mass of the tumour. We aim to expose new solutions of (6.5) with plausible and biological interpretation. By considering the integral in Eq. (6.5) only to be dependant on t, as we are integrating on x, we write (6.5) as

$$u_t = u_{xx} + F(t)\, u\,(1-u), \tag{6.6}$$

  where $F = F(t)$ is the function describing the impact of the whole mass of the tumour, which is actually influencing on its proliferation.

- A generalisation which reads as

$$u_t = \frac{1}{c(x)}\big(c(x)\cdot g(u)u_x\big)_x + f(u). \tag{6.7}$$

where $g(u)$ is the diffusion coefficient dependent on the variable u, with x and t as independent variables, $f(u)$ as an arbitrary function, and $c(x)$ an arbitrary function that depends on the space variable x to account for spatial heterogeneity of the medium. The function $u(x,t)$ denotes the density of the biological population. For particular cases of $f(u)$, $g(u)$ and $c(x)$, this model has recently attracted considerable interest in studies of tumour growth and their applications [269, 273, 274, 275, 281]. This Equation has also been intensively studied: in [298] some non-trivial conservation laws associated to the symmetries were obtained for $g = k\cdot f_u$ and f, c arbitrary functions; in [299] the classical Lie method was applied to derive some non-trivial conservation laws for this equation. Symmetry reductions and exact solutions for Eq. (6.7) were obtained using classical and potential symmetries in [300].

Many mathematical models do not take into account the spatial heterogeneity of the medium. For example, grey and white matter of brain tissue was taken into account in [301], where grey matter is composed of neuronal and glial cell bodies that control brain activity, while the cortex is a coat of grey matter that covers the brain [267] and it is connected to other grey matter regions by white matter fibre bundles. The importance of considering spatial heterogeneity in ecological dispersal has been emphasised in studies where the pattern and scale of movement have been characterised, such as the dispersal of forest beetles [302]. Interface problems arise in the setting of various physical and engineering problems [303, 304] and references therein. In recent years many mathematicians and researchers have been working on developing numerical methods to find numerical solutions for non-linear partial differential equations (PDEs), including those representing interface problems [305, 306, 307, 308]. Analytical solutions are rare and difficult to obtain, hence, the search for exact solutions to non-linear PDEs plays a fundamental role in the analysis of nonlinear physical phenomena. However, exact solutions can be found in several papers [283, 309, 310] in which the authors determine a number of exact solutions for particular cases of Eq. (6.7) using classical Lie point techniques.

The applicability of mathematical models has risen exponentially in recent

years. Biosciences have taken an interest in many realistic mathematical models, as they are useful to explain several natural phenomena, such as cell invasion, climate change, or even tumour development. As the complexity increases within the development of mathematical models, which may take into account an overwhelming number of biological factors, computer sciences and mathematical methods may shed some light on their proper descriptions. The rise of computer simulations can therefore be supremely useful in biological field experiments, as they become a highly efficient tool to facilitate and expedite research.

Thus, the structure of this Chapter goes as follows: firstly, the Lie classical method is reviewed generally in Section 6.2 in order to obtain solutions for differential equations. We apply this method in order to obtain a group classification for Equations (6.4), (6.6) and (6.7). Then, respectively in Sections 6.3, 6.4 and6.5, we focus on cases with a special biological meaning, and then obtained some exact solutions.

6.2 Lie symmetries and reductions

Lie classical method is based on the determination of the point symmetry group of a differential equation, i.e., the largest group of transformations acting on dependent and independent variables of the equation so that it maps solutions of the equation into other solutions.
An infinitesimal point symmetry of Eqs. (6.4), (6.6) or (6.7) will be given by a generator of the form

$$\mathbf{X} = \tau(t,x,u)\partial_t + \xi(t,x,u)\partial_x + \eta(t,x,u)\partial_u. \tag{6.8}$$

Each equation would admit a Lie point symmetry provided that

$$pr(\mathbf{X})(\Delta) = 0 \qquad \text{when} \qquad \Delta = 0, \tag{6.9}$$

where $\Delta = \Delta_1, \Delta_2$ or Δ_3 can be each of the Eqs. (6.4), (6.6) or (6.7), where

$$\begin{array}{l} \Delta_1 = u_t - \big(g(u)c(x)u_x\big)_x - f(u), \\ \Delta_2 = u_t - u_{xx} - F(t)\,u\,(1-u), \\ \Delta_3 = u_t - \frac{1}{c(x)}\big(c(x)\cdot g(u)u_x\big)_x - f(u), \end{array} \tag{6.10}$$

respectively. The term $pr(\mathbf{X})$ is the prolongation of the vector field (6.8). Exponentiation of the point symmetry generator (6.8) produces a one-parameter symmetry transformation group

$$(t,x,u) \to (t^*,x^*,u^*) = \exp(\epsilon\mathbf{v})(t,x,u) \tag{6.11}$$

with ϵ the group parameter and

$$(t^*, x^*, u^*)|_{\epsilon=0} = (t, x, u) \tag{6.12}$$

the identity transformation. The explicit form of (6.8) can be obtained by solving the system

$$\frac{\partial t^*}{\partial \epsilon} = \tau(t^*, x^*, u^*), \quad \frac{\partial x^*}{\partial \epsilon} = \xi(t^*, x^*, u^*), \quad \frac{\partial u^*}{\partial \epsilon} = \eta(t^*, x^*, u^*), \tag{6.13}$$

with initial conditions

$$t^*|_{\epsilon=0} = t, \quad x^*|_{\epsilon=0} = x, \quad u^*|_{\epsilon=0} = u. \tag{6.14}$$

The infinitesimal action of a point symmetry (6.8) on solution $u(t, x)$ of Eq. $\Delta = 0$ is given by $u = u(t, x) \to u^* = u^*(t, x)$, where

$$u^* = u + \epsilon\Big(\eta(t, x, u) - \tau(t, x, u)u_t - \xi(t, x, u)u_x\Big) + O(\epsilon^2) \tag{6.15}$$

which corresponds to a generator

$$\tilde{\mathbf{X}} = P\,\partial_u, \qquad P = \eta - \tau\, u_t - \xi\, u_x, \tag{6.16}$$

called the characteristic form of the infinitesimal point symmetry (6.8). The invariance condition (6.9) is then

$$pr(\tilde{\mathbf{X}})(\Delta) = 0 \qquad \text{when} \qquad \Delta = 0 \tag{6.17}$$

for

$$pr(\tilde{\mathbf{X}}) = pr(\mathbf{X}) - \tau D_t - \xi D_x \qquad \text{when} \qquad \Delta = 0. \tag{6.18}$$

We obtain a set of determining equations for the infinitesimals $\xi = \xi(x, t, u)$, $\tau = \tau(x, t, u)$ and $\eta = \eta(x, t, u)$ by splitting the symmetry determining Eq. (6.17). This method will be used in the next Sections to obtain solutions of each of the Eqs. (6.4), (6.6) and (6.7).

6.3 A Fisher equation with a proliferation term dependant on density and space

We focus now in obtaining symmetries from Eq. (6.4). It admits a Lie point symmetry provided that

$$pr^{(2)}\mathbf{v}(\Delta) = 0 \qquad \text{when} \qquad \Delta = 0,$$

where $\Delta = u_t - f(u) - (g(u)c(x)u_x)_x$ and $pr^{(2)}\mathbf{v}$ is the second prolongation of the vector field (6.8). We obtain a set of determining equations for the infinitesimals $\xi = \xi(x,t,u)$, $\tau = \tau(x,t,u)$ and $\eta = \eta(x,t,u)$. From the determining system, we get that $\xi = \xi(x,t)$, $\tau = \tau(t)$, where η, τ, ξ, g, f and c must satisfy the following equations:

$$\begin{aligned}
&c\,g_u\,\eta + c\,\tau_t\,g - 2\,c\,\xi_x\,g = 0,\\
&c\,g\,\eta_{u\,u} + c\,g_u\,\eta_u + c\,g_{u\,u}\,\eta + c\,\tau_t\,g_u - 2\,c\,\xi_x\,g_u = 0,\\
&2\,c\,g_u\,\eta_x + 2\,c\,g\,\eta_{u\,x} + c_x\,g_u\,\eta + c_x\,\tau_t\,g - c\,\xi_{x\,x}\,g - c_x\,\xi_x\,g + \xi_t = 0,\\
&-c\,g\,\eta_{x\,x} - c_x\,g\,\eta_x + f\,\eta_u + \eta_t - f_u\,\eta - \tau_t\,f = 0.
\end{aligned} \tag{6.19}$$

After solving the determining equations, we can distinguish different cases in which the symmetries are admitted by Eq. (6.4) for functional forms of $c(x)$, $f(u)$ and $g(u)$, where $c' \neq 0$, $f' \neq 0$, $g' \neq 0$. We distinguish as well the corresponding generators and group transformations, which are given in Appendix C.1.

In this section we will focus on case 4 from Appendix C.1, as function $c = c(x)$ is arbitrary, and functions $f(u) = f_1\,(u - g_2) + f_2\,(g_2 - u)^{g_1+1}$ and $g(u) = g_3(g_2 - u)^{g_1}$ have a biological interest in terms of modelling, respectively, cancer cell proliferation as a Verhulst's law of growth [267, 273, 311], and the diffusion term as a typical glioma invasion [267, 275, 311]. By using the generator $\mathbf{X}_4$, we obtain the similarity variable and similarity solution

$$z = x, \quad u = e^{f_1\,t}\,h(z) + g_2, \tag{6.20}$$

and the ODE_4

$$h_{zz} + \frac{g_1\,h_z^2}{h} + \frac{c_z\,h_z}{c} - \frac{f_2\,h}{c\,g_3} = 0. \tag{6.21}$$

If we set $h(z) = -\sqrt{v(z)}$, we obtain that (6.21) is equivalent to

$$v_{zz} - \frac{v_z^2}{2\,v}(g_1 - 1) + \frac{c_z\,v_z}{c} - \frac{2\,f_2\,v}{g_3\,c}. \tag{6.22}$$

We set $g_1 = 1$ as it provide us a linear density diffusion term, which is a Maltusian rate of growth [312]. Then, Eq. (6.22) is transformed into

$$v_{zz} + \frac{c_z\,v_z}{c} - \frac{2\,f_2\,v}{g_3\,c} = 0. \tag{6.23}$$

It can be easily proved that a first integral of Eq. (6.23) is the Ricatti equationo

$$w_z + w^2 + \frac{c_z}{c}w - 2\frac{f_2}{g_3\,c} = 0, \tag{6.24}$$

with $w = w(z)$ and the change $v(z) = \exp(\alpha(z))$ for $\alpha'(z) = w(z)$.

Besides, we want $c = c(x)$ to have an asymptotic behaviour (for large x) related to $\tanh(x)$, which has biological interest as it models single and multiple sharp transition regions [281]. With $g_1 = 1$, we search a solution of (6.24) such as

$$w(z) = \frac{1}{K_1} \tanh\left(\frac{z + K_2}{K_1}\right), \quad K_1 \neq 0, \tag{6.25}$$

so that $c = c(x)$ becomes

$$c(x) = \frac{2\, f_2\, {K_1}^2}{g_3} + \frac{K_3 \sqrt{1 - \left(\tanh\left(\frac{x+K_2}{K_1}\right)\right)^2}}{\tanh\left(\frac{x+K_2}{K_1}\right)} - K_4, \quad K_1 \neq 0, x \geq 0. \tag{6.26}$$

where $K_4 = \operatorname{arctanh}\left(\frac{K_3\, g_3}{\sqrt{4\, {f_2}^2 {K_1}^4 + {K_3}^2 {g_3}^2}}\right) K_1 - K_2$ as the diffusion term cannot be negative [267]. In this case, the asymptotic behaviour of $c = c(x)$ is the following:

$$\lim_{x \to \infty} c(x) = \frac{2\, f_2\, {K_1}^2}{g_3} - K_4. \tag{6.27}$$

Therefore we have provided a one-parameter family of exact solutions of the Eq. (6.4)

$$u(x,t) = g_2 - \frac{e^{f_1\, t}}{\sqrt[4]{1 - \tanh\left(\frac{x+K_2}{K_1}\right)^2}}, \tag{6.28}$$

for each $K_1 \neq 0$ and $c = c(x)$ as in Eq. (6.26).

The carrying capacity in this equation can be seen as g_2, and we obtained for the solution (6.28) that

$$\lim_{t \to \infty} u = g_2. \tag{6.29}$$

This implies that in any region of the space the solutions assume the value of the limit concentration of cells. This is shown in Figure 6.1. Correspondingly, the density-diffusion and growth function asymptotically disappear, this is

$$\lim_{t \to \infty} f = 0, \quad \lim_{t \to \infty} g = 0. \tag{6.30}$$

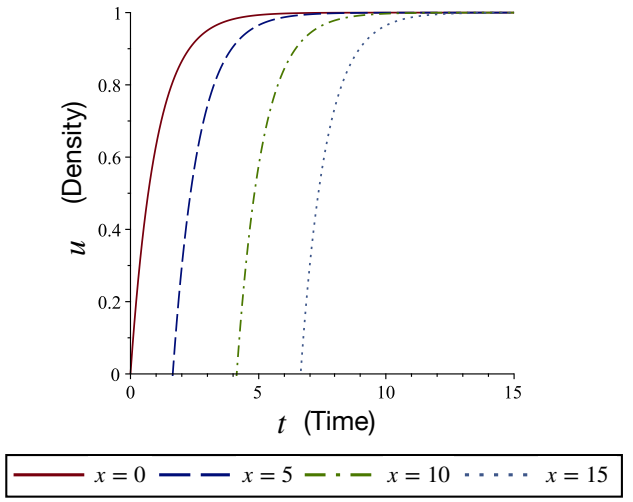


Figure 6.1: **Graphs of population density solutions** (6.28)**.** The results are shown for $g_2 = 1$, $K_1 = 1$, $K_2 = -1$ over different times t and displacement x given. The asymptotic behaviour can be observed.

6.4 A Fisher equation with a proliferation term involving tumour development

From the determining system, we get that $\tau = \tau(t)$, where η, τ, ξ and $F = F(t)$ must satisfy the following equations

$$
\begin{aligned}
&\xi_{u,u} = 0,\\
&\tau_{u,u} = 0,\\
&\eta_{u,u} - 2\xi_{x,u} = 0,\\
&\xi_u + \tau_{x,u} = 0,\\
&3\,F\,u\,\xi_u(1-u) + 2\,\eta_{x,u} - \xi_{x,x} + \xi_t = 0,\\
&F\,u\,\tau_u(1-u) - 2\,\xi_x - \tau_{x,x} + \tau_t = 0,\\
&(F(\eta_u - 2\xi_x) - \tau\,F')u^2 + (\tau\,F' + F(\,2\xi_x - \eta_u - 2\,\eta\,F))u + \eta\,F - \eta_t + \eta_{x,x} = 0.
\end{aligned}
\tag{6.31}
$$

After solving the determining equations, we can distinguish different cases in which the symmetries are admitted by Eq. (6.6) for functional forms of $F(t)$. We distinguish as well the corresponding generators in Appendix C.2.

We now consider a special case in order to obtain exact solutions of Eq.

(6.6). For this case, we consider the tumour mass as a form of tanh, as it may model transition regions [281]. This is, we take $F = F(t)$ as

$$F(t) = a\,\tanh(b\,t) + c, \qquad a, b, c \in \mathbb{R}. \tag{6.32}$$

Considering the Case 4, $F = F(t)$ as in Eq. (6.32) verifies Eq. (C.36) only if $a, c = \pm b$. Eq. (C.25) is then verified for

$$\hat{F}(t) = \pm b\,\tanh(b\,t) \pm b. \tag{6.33}$$

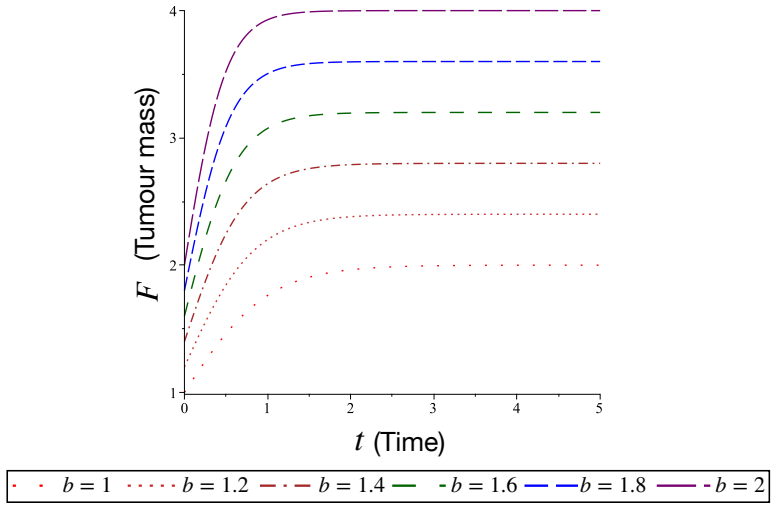


Figure 6.2: **Function $\hat{F}(t)$ in Eq.** (6.33). The results are shown for different values of b, where b can be interpreted as the influence of the mass into the proliferation term.

The use of this structure may be suitable as an upper bounded mass, and mass tends to grow quickly as tanh. Considering the positive signs of (6.33), two symmetries are obtained for Eq. (6.6), which are $\mathbf{X}_1$ and

$$\mathbf{X}_{4*} = \partial_t + b\,u(\tanh(b\,t) - 1)\partial_u. \tag{6.34a}$$

We consider then the similarity variable and similarity solution

$$z = x, \qquad u = \frac{h(z)e^{-b\,t}}{\sqrt{1 - \tanh(b\,t)^2}}, \tag{6.35}$$

yielding the reduction

$$b\,h(z)^2 - 2\,b\,h(z) - h_{zz}. \tag{6.36}$$

A particular solution of Eq. (6.36) is

$$h(z) = 3\,\tanh\left(\sqrt{\frac{b}{2}}\,(x+k_1)\right)^2 - 1, \qquad k_1 \in \mathbb{R}, \tag{6.37}$$

so that, for F as in Eq. (6.33), we have obtained a two-parameter family of exact solutions of Eq. (6.6), which is

$$\hat{u}(t,x) = \frac{\left(3\,\tanh\left(\sqrt{\frac{b}{2}}\,(x+k_1)\right)^2 - 1\right)\,e^{-b\,t}}{\sqrt{1-\tanh(b\,t)^2}}, \qquad k_1 \in \mathbb{R}. \tag{6.38}$$

The shape of this solution is consistent with the solutions found in the previous Section. The behaviour of solution (6.38) is shown in Figures 6.3. In Figure 6.3 (A), it is observed that whenever x grows, the tumour density grows along. In Figures 6.3(A)-(B), it can be observed how over time the tumour density tends to stabilise. In both images it is shown the effect of parameter b, considered as the impact of the mass into the proliferation term.

6.5 A Fisher equation describing a tumour interface problem

The Lie symmetries of Eq. (6.7) depend on the form of the arbitrary elements (functions $f(u)$, $g(u)$, and $c(x)$). The search for these symmetries is more difficult than looking for symmetries of a specific partial differential equation. In this section, we define the Lie group classification for Eq. (6.7) in the case $f\,g_u\,c' \neq 0$ (this is, each term must be non-zero). In order to simplify this analysis, we use the equivalence transformations admitted by Eq. (6.7). Although the direct method (used first by Lie [313]) to calculate equivalence transformations has the benefit of finding the most general equivalence group, as it involves considerable computational difficulties, we have decided to use the Lie infinitesimal method, which was introduced by Ovsiannikov in Ref. [292]. A more detailed description and examples of both methods can be found in [314]. For the sake of simplicity, we prefer to introduce the new function as follows:

$$\alpha(x) = \frac{c'(x)}{c(x)}. \tag{6.39}$$

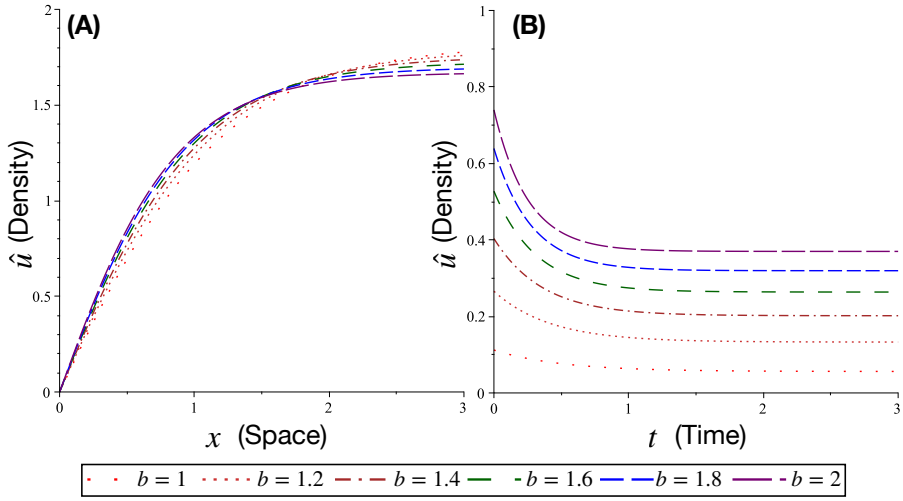


Figure 6.3: **Graphs of $\hat{u}(t,x)$ in Eq.** (6.38). The results are shown **(A)** for a given $t = t_0$ **(B)** for a given $x = x_0$.

As such, Eq. (6.7) can be written as

$$u_t = f(u) + \alpha g u_x + g_u u_x^2 + g u_{xx}. \tag{6.40}$$

An infinitesimal generator of the equivalence transformations of equation (6.40) has the form

$$\mathbf{Y} = \Xi^1 \partial_t + \Xi^2 \partial_x + \phi \partial_u + \mu^1 \partial_f + \mu^2 \partial_\alpha + \mu^3 \partial_g, \tag{6.41}$$

wherein the infinitesimal components Ξ^1, Ξ^2, and ϕ, depend on t, x, and u, while the infinitesimal components μ^i, $(i = 1, 2, 3)$ can also depend on f, g, and α. The Lie infinitesimal criterion [292] requires invariance with respect to a suitable prolongation of Y (the interested reader can refer to [292, 315] for details on how the operator can be extended) of the following equations

$$u_t = f(u) + \alpha g u_x + g_u u_x^2 + g u_{xx}, \tag{6.42}$$
$$f_t = f_x = 0, \tag{6.43}$$
$$g_t = g_x = 0, \tag{6.44}$$
$$\alpha_t = \alpha_u = 0, \tag{6.45}$$

where conditions (6.43)-(6.45) represent the so called auxiliary conditions and provide the functional dependence of functions f, g, and α. Solving the corresponding determining system, we find in [316] that class (6.40) admits a continuous group of equivalence transformations generated by the following operators:

$$\begin{aligned} &\mathbf{Y}_1 = \partial_t, \quad \mathbf{Y}_2 = \partial_x, \quad \mathbf{Y}_3 = \partial_u, \\ &\mathbf{Y}_4 = t\partial_t - f\partial_f - g\partial_g, \quad \mathbf{Y}_5 = x\partial_x - \alpha\partial_\alpha + 2g\partial_g, \quad \mathbf{Y}_6 = u\partial_u + f\partial_f. \end{aligned} \tag{6.46}$$

The finite form of these equivalence transformations is

$$\tilde{t} = \epsilon_0 + \epsilon_1 t, \quad \tilde{x} = \epsilon_2 + \epsilon_3 x, \quad \tilde{u} = \epsilon_4 + \epsilon_5 u, \quad \tilde{f} = \frac{\epsilon_1}{\epsilon_5} f, \quad \tilde{\alpha} = \epsilon_3 \alpha, \quad \tilde{g} = \frac{\epsilon_1}{\epsilon_3^2} g. \tag{6.47}$$

In [316], these equivalence transformations are considered to simplify the symmetry analysis of class (6.40).

In this study, we present the Lie group classification of (6.40) module over this group of equivalence transformations admitted by this class of equations. The associated Lie algebra of infinitesimal symmetries is the set of vector fields of the form

$$\mathbf{v} = \xi^1 \frac{\partial}{\partial t} + \xi^2 \frac{\partial}{\partial t} + \eta \frac{\partial}{\partial u}. \tag{6.48}$$

The requirement that this operator leaves Eq. (6.40) invariant yields to the overdetermined linear system (the so called determining system) for the infinitesimals $\xi^1(t,x,u)$, $\xi^2(t,x,u)$ and $\eta(t,x,u)$.

Having determined the infinitesimals, the form of invariant solutions is found by solving the invariant surface condition

$$\Phi \equiv \xi^1 \frac{\partial u}{\partial t} + \xi^2 \frac{\partial u}{\partial x} - \eta = 0. \tag{6.49}$$

From the determining system, we get $\xi^1 = \xi^1(t)$, $\xi^2 = \xi^2(t,x)$, and then $\xi^1(t)$, $\xi^2(t,x)$ and $\eta(t,x,u)$ must satisfy the following equations, depending on $f(u)$, $g(u)$, and $\alpha(x)$,

$$\begin{aligned} &g\,\xi^1_t + g_u\,\eta - 2g\,\xi^2_x = 0, \\ &\eta_u\,g_u\,g + \eta_{uu}\,g^2 + \eta\,g\,g_{uu} - \eta\,g_u^2 = 0, \\ &\alpha\,\xi^2_x\,g - \xi^2_{xx}\,g + \alpha_x\,\xi^2\,g + 2\,\eta_{xu}\,g + \xi^2_t + 2\,\eta_x\,g_u = 0, \\ &\alpha\,g^2\,\eta_x + g^2\,\eta_{xx} - f\,g\;\eta_u - g\,\eta_t + g\,f_u\,\eta + 2f\,g\,\xi^2_x - f\,g_u\,\eta = 0. \end{aligned}$$

For arbitrary values of f, g, and α, the only symmetry generator admitted by (6.40) is

$$\mathbf{v}_1 = \partial_t. \tag{6.50}$$

Moreover, whenever the function $\alpha(x)$ is constant, Eq. (6.40) also admits the symmetry generator

$$\mathbf{v}_2 = \partial_x. \tag{6.51}$$

In order to have extra symmetry generators, the function g can not be arbitrary. We distinguish the following cases depending on the function g

1. $g = g_0 u^{g_1}$, with $g_0 = \pm 1$, $g_1 \neq 0, -4/3$.
2. $g = g_0 u^{-4/3}$, with $g_0 = \pm 1$.
3. $g = g_0 e^{u g_1}$, with $g_0 = \pm 1$ $g_1 \neq 0$.

For each of these cases we consider only the forms of the functions f and α which yield extra symmetry generators. We have provided the corresponding results (forms of the functions f, α, and extra generators) in the tables from Appendix C.3.

6.5.1 Analytical exact solutions of biological interest

In this section, we get analytical exact solutions and study their behaviour in a number of particular cases of Eq. (6.7) that present several relevant applications in the field of mathematical biology. In [317], Belmonte-Beitia proved the existence of upper and lower bounds for traveling waves solutions for a particular case of Eq. (6.7) in which $c(x) =$ constant. He considered the following equation

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[(1-u)\frac{\partial u}{\partial x}\right] + \rho u(1-u), \tag{6.52}$$

which is a mathematical model for glioma growth and invasion.
In [273, 301, 318], Swanson et al. have based their mathematical models on a variant of the Fisher-Kolmogorov equation. In those papers, the authors have investigated how the proliferation and dispersal of glioma cells combine to generate increasing degrees of cellularity, mitoses, hypoxia-induced neoangiogenesis and necrosis. To account for the spatial heterogeneity of the brain tissue, the authors made the diffusion coefficient a function of the spatial variable x, differentiating regions of grey and white matter [273, 301]. Macro-

and microscopic patterns of growth have suggested that glioma cells tend to migrate along white matter tracts in the brain and, indeed, in vitro experiments have shown that cells exhibit greater motility in white matter than in grey [319]. Swanson [320] reformulated the model in order to more accurately reflect the spatial limitation of cellular proliferation and the inherent heterogeneity in the brain by introducing the carrying capacity of the tissue, and by allowing the diffusion coefficient to depend on the tissue environment x. In [321], Konukoglu et al. considered a reaction-diffusion model using a modified anisotropic eikonal equation and proposed a parameter estimation method using time series of medical images.

Taking into account the biological perspective, we have considered the special form of f and g based purely on the classical definition of cancer as uncontrolled proliferation of cell with the potential for invasion and metastasis [267].

Therefore, considering Eq. (6.7), we set proliferation and diffusion terms as

$$f(u) = ku\left(1 - \frac{u}{u_*}\right), \tag{6.53}$$

$$g(u) = \rho\left(1 - \frac{u}{u_*}\right), \tag{6.54}$$

where the proliferation rate is $k > 0$, ρ is the diffusion rate, and u_* is the limit concentration of cells that a certain volume of tissue can hold (i.e. the carrying capacity of the tissue). Moreover, $c(x)$ is related to a tanh form, modelling a single transition region [281].

Finally, the mathematical equation for cancer cell density considered here is the following:

$$u_t = \frac{1}{c(x)}\left[c(x)\underbrace{\rho\left(1 - \frac{u}{u_*}\right)}_{\text{diffusion}}u_x\right]_x + \underbrace{ku\left(1 - \frac{u}{u_*}\right)}_{\text{proliferation}}; \tag{6.55}$$

where $u(x,t)$ denotes the density of cells. We have assumed that this equation has a single interfacial transition in the diffusion coefficient, making it a very good model to describe solutions of biological interest. We have taken the proliferation term as Verhulst's law of growth which is used to model cancer cell proliferation [267, 273] and the diffusion term with the typical form g [267, 317, 318]. Scaling the cell density so that the carrying capacity becomes unity (this corresponds to the transformation $u = u^*\tilde{u}$, but in the following we will

omit the accent mark for sake of readability), setting $\rho = 1$ and making the following change of variables

$$t = t, \quad x = x, \quad u = (1 - v)\,,$$

Eq. (6.55) can be written as

$$v_t = k\left(v^2 - v\right) + \frac{1}{c(x)}\left[c(x)vv_x\right]_x. \tag{6.56}$$

This equation falls under the second case in Table 1 from Appendix C.3, with $g_0 = 1$, $g_1 = 1$, $f_0 = k$, and $f_1 = -k$. In this case, when $\alpha(x) = \frac{c'(x)}{c(x)}$ does not satisfy condition (C.59), Eq. (6.56) only admits the additional generator

$$\mathbf{v}_4 = e^{kt}\partial_t - ke^{kt}v\partial_v. \tag{6.57}$$

Then we look for a solution with the form

$$v(x,t) = U(x)e^{-kt}, \tag{6.58}$$

where $U(x)$ is a solution of the following equation

$$kU^2 + \frac{c'}{c}UU' + UU'' + U'^2 = 0. \tag{6.59}$$

Setting $U = \sqrt{V}$ we get the linear equation

$$V'' + \frac{c'}{c}V' + 2kV = 0. \tag{6.60}$$

The general solution of Eq. (6.60) is

$$V(x) = V_0(x)\left(c_1 + c_2\int \frac{e^{-F}}{V_0^2(x)}\,dx\right) \tag{6.61}$$

where $F = \int \frac{c'(x)}{c(x)}\,dx$ and $V_0(x)$ is a nontrivial particular solution, if $V_0(x) \neq 0$, of (6.62).
If we consider $c'(x)/c(x) = c_0\tanh(x)$, with c_0 as a constant, then Eq. (6.60) becomes

$$V'' + c_0\tanh(x)V' + 2kV = 0, \tag{6.62}$$

whose solution can be found in terms of associate Legendre functions

$$V(x) = c_1 \frac{\text{LegendreP}\left(\frac{c_0-2}{2}, \frac{\sqrt{c_0^2-8k}}{2}, \tanh(x)\right)}{(\cosh(x))^{c_0/2}} + c_2 \frac{\text{LegendreQ}\left(\frac{c_0-2}{2}, \frac{\sqrt{c_0^2-8k}}{2}, \tanh(x)\right)}{(\cosh(x))^{c_0/2}}. \tag{6.63}$$

Thus, solutions of Eq. (6.55) will have the following form:

$$u = 1 - e^{-kt}\sqrt{V(x)}. \tag{6.64}$$

From now on, we consider special values of the constant c_0:

1. Setting $c_0 = 2$, the transformation

$$V(x) = \frac{w(x)}{\cosh(x)} \tag{6.65}$$

maps Eq. (6.62) into

$$w'' + (2k-1)w = 0, \tag{6.66}$$

whose general solution depends on the value of k. Taking into account that $k > 0$, we have the following solutions for equations (6.66)

$$w(x) = c_1 + c_2 x, \quad \text{if} \quad k = \frac{1}{2}, \tag{6.67}$$

$$w(x) = c_1 \sin(x\sqrt{2k-1}) + c_2 \cos(x\sqrt{2k-1}), \quad \text{if} \quad k > \frac{1}{2}, \tag{6.68}$$

$$w(x) = c_1 \sinh(x\sqrt{1-2k}) + c_2 \cosh(x\sqrt{1-2k}), \quad \text{if} \quad k < \frac{1}{2} \tag{6.69}$$

Accordingly, we obtain solutions of Eq. (6.55):

$$u = 1 - e^{-\frac{t}{2}}\sqrt{\frac{c_1 + c_2 x}{\cosh(x)}}, \quad k = \frac{1}{2}, \tag{6.70}$$

$$u = 1 - e^{-kt}\sqrt{\frac{c_1 \sin(x\sqrt{2k-1}) + c_2 \cos(x\sqrt{2k-1})}{\cosh(x)}}, \quad k > \frac{1}{2} \tag{6.71}$$

$$u = 1 - e^{-kt}\sqrt{\frac{c_1 \sinh(x\sqrt{1-2k}) + c_2 \cosh(x\sqrt{1-2k})}{\cosh(x)}}, \; k < \frac{1}{2} \tag{6.72}$$

2. Setting $c_0 = 2n$ with n a positive integer, and $k = \frac{n^2}{2}$ the solutions (6.63) of (6.62) become

$$V(x) = \frac{c_1 \text{LegendreP}\left(n-1, 0, \tanh(x)\right) + c_2 \text{LegendreQ}\left(n-1, 0, \tanh(x)\right)}{(\cosh(x))^n}, \tag{6.73}$$

and the corresponding solutions of Eq. (6.55) are given by (6.64). For $n = 2$ we obtain the solution (6.70).

3. Setting $c_0 = 1$ the solution of (6.62) is

$$\begin{aligned} V(x) =& c_1 \frac{\text{LegendreP}\left(\frac{-1}{2}, \frac{i\sqrt{8k-1}}{2}, \tanh(x)\right)}{\sqrt{\cosh(x)}} + \\ & c_2 \frac{\text{LegendreQ}\left(\frac{-1}{2}, \frac{i\sqrt{8k-1}}{2}, \tanh(x)\right)}{\sqrt{\cosh(x)}}, \end{aligned} \tag{6.74}$$

and the corresponding solutions of Eq. (6.55) are given by (6.64).

6.5.2 Discussion

In this study we have explored cell motility near interfaces. Given the interpretability of the results, we discuss this last solutions from Eq. (6.7) in more depth. We have been working under the assumption that the proliferation rate $k > 0$, and one can easily see that, fixing x, the asymptotic behaviour of all solutions obtained (6.70), (6.71) and (6.72) is

$$\lim_{t\to\infty} u = 1. \tag{6.75}$$

Recalling that we scaled the cell density so that the carrying capacity u_* is unity, we then obtain

$$\lim_{t\to\infty} u = u_*. \tag{6.76}$$

This is the case in any region of the space where the solutions asymptotically assume the limiting value of the cell concentration. Accordingly, the diffusion and also the proliferation disappear asymptotically, giving us

$$\lim_{t\to\infty} f = 0, \ \lim_{t\to\infty} g = 0. \tag{6.77}$$

Furthermore, the solutions asymptotically approach the value of the limiting concentration of cells for large values of x. To demonstrate this, we observe

that, for solutions (6.70) and (6.71), we easily get $\lim_{x\to\infty} u = 1$.
As mentioned above, this is equivalent to

$$\lim_{x\to\infty} u = u_*. \tag{6.78}$$

Taking into account that $k > 0$, in the solutions (6.72), and if $0 < 1 - 2k < 1$, then condition (6.78) will also hold. Solution (6.64) behaves the same way, because function $V(x)$, given by (6.73), is infinitesimal when x approaches infinity.
In order to provide a biological meaningful solution and discussion, we focus on the family (6.72) of solutions of Eq. (6.7). We have $c_0 = 2$ and set $c_1 = c_2 = \frac{1}{2}$. Thus, we obtain a one-parameter family of solutions of Eq. (6.7) as

$$u = 1 - e^{-kt}\sqrt{\frac{\sinh(x\sqrt{1-2k}) + \cosh(x\sqrt{1-2k})}{2\cosh(x)}},\ k < \frac{1}{2}, \tag{6.79}$$

where k is a free parameter. For this family of solutions, large values of x and t tend to approach the limiting value $u_* = 1$.
The family (6.79) of solutions provided can be considered as a model for an interface in brain tumours [273, 275]. For this case, we have provided some simulations where we set a single transition region in the centre of a one-dimensional spatial domain for $x \in (-20, 20)$ and time as $t \in (0, 100)$.
As we set $c'(x)/c(x) = c_0 \tanh(x)$, the interface would be located at $x_0 = 0$. The solutions provided succeed in modelling a higher cellular density for $x < 0$ and a low one for $x > 0$. These zones would represent respectively the grey and white matter as in [273, 275]. For a fixed $t = t_0$, whenever $k \to \frac{1}{2}$ the cellular density u increases and decreases at the same speed. However, when k decreases, cellular density grows at a slower rate when $x > 0$ in comparison to whenever $x < 0$. As u decreases, diffusion (6.53) and proliferation (6.54) would increase whenever $x \to 0^-$, always depending on the parameter k. This would be consistent with the fact that, when passing through the interface, diffusion increases in a great factor [275]. However, the proliferation rate would disappear when $k \to 0$. Furthermore, now for a fixed $x = x_0$, the density grows faster along with k, reaching its maximal speed whenever k is closer to $\frac{1}{2}$.
Depending on the value k, different representations of tumour invasion at an interface level can be considered. With higher k there is a lower diffusion and higher proliferation rate, and over time tumour density grows quickly. With lower k, diffusion increases and proliferation decreases, so that over time density grows slower. These results are provided in Figures (6.4) and (6.5).

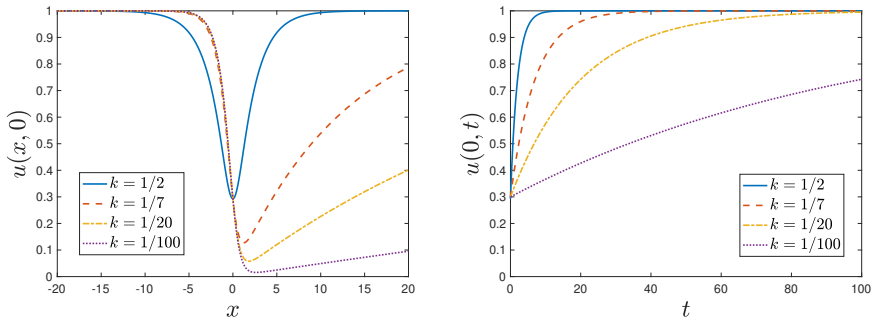


Figure 6.4: **Asymptotic behaviour of the solution $u = u(x,t)$ in Eq. (6.79) for fixed values of x and t.** The function $u = u(x,t)$ is shown when Eq. (6.79) holds. The left figure shows the asymptotic behaviour of $u(x,0)$, while the right shows the asymptotic behaviour of $u(0,t)$. The interface is centred at $x = 0$, and the parameter k models different cell density behaviours. For large values of x and t, the solution u approaches the limit $u_* = 1$. Units of x and t are considered as *mm* and *days*, respectively, as in [281].

6.6 Conclusions

In this Section, we have examined generalised Fishers equation that models biological invasions from the point of view of Lie symmetries. The study of non-linear phenomena has been a continuous source of new problems and has motivated the introduction of new methods in the areas of mathematical analysis, partial differential equations, and other disciplines, thereby becoming one of the most active areas of mathematical research over the last decades. The investigation of exact solutions of non-linear PDEs plays an essential role in the analysis of non-linear phenomena, particularly in obtaining analytical solutions for interface problems. The Lie symmetry method greatly simplifies many non-linear problems. Exact solutions are difficult to investigate in general. The combination of Lie group theory and equivalence transformation yields exact solutions.

In this Section we have classified the Lie symmetries of the generalised Fisher equations (6.4), (6.6) and, (6.7). Firstly, a Fisher Eq. (6.4) with a density-space dependent reaction-diffusion term was presented, which can be considered as an essential part of cancer modelling and cell dynamics. By applying the classical

Lie group method, we obtained a symmetry classification for Eq. (6.4). We have obtained a reduction of order of the ODE_4 derived from (6.4). In particular, we have found a one-parameter family of solutions with biological meaning. The generalised Fisher's Eq. (C.25) has been studied in terms of a function F describing the proliferation dynamics. By applying the classical Lie point symmetry method, we have derived a symmetry classification for (C.25). This was also performed for Eq. (6.6), as we also obtained several reductions. We have provided a one-parameter family of solutions in Eq. (6.38) and simulated solutions which described tumour dynamics.

Lastly, for Eq. (6.7)., we have studied the functions $f(u)$, $g(u)$ and $c(x)$, for which the principal Lie algebra was extended. We simplified this analysis by using equivalence transformations. Moreover, we have provided a number of particular cases of Eq. (6.7) of biological interest, such as tumour progression at their interfaces. As such, we have considered the special form of f and g based purely on the classical definition of cancer as the uncontrolled proliferation of cell with the potential for invasion and metastasis, and the function $c(x)$ is specifically in tanh form, modelling a single transition region [281].

For the corresponding equations of the generalised Fisher equation that models biological invasions, specially for Eq. (6.7), we have obtained analytical solutions for a model of tumour progression at their interfaces. Finally, we have qualitatively studied a special case which simulates the cell density behaviour around an interface. This study is consistent with the biological description of cell density and motility between the grey and white matter of the brain. These results can be used as predictive tools or as a means of understanding tumour growth dynamics.

The methods, results and conclusions related to this Chapter can be found in Refs. [322, 323, 324].

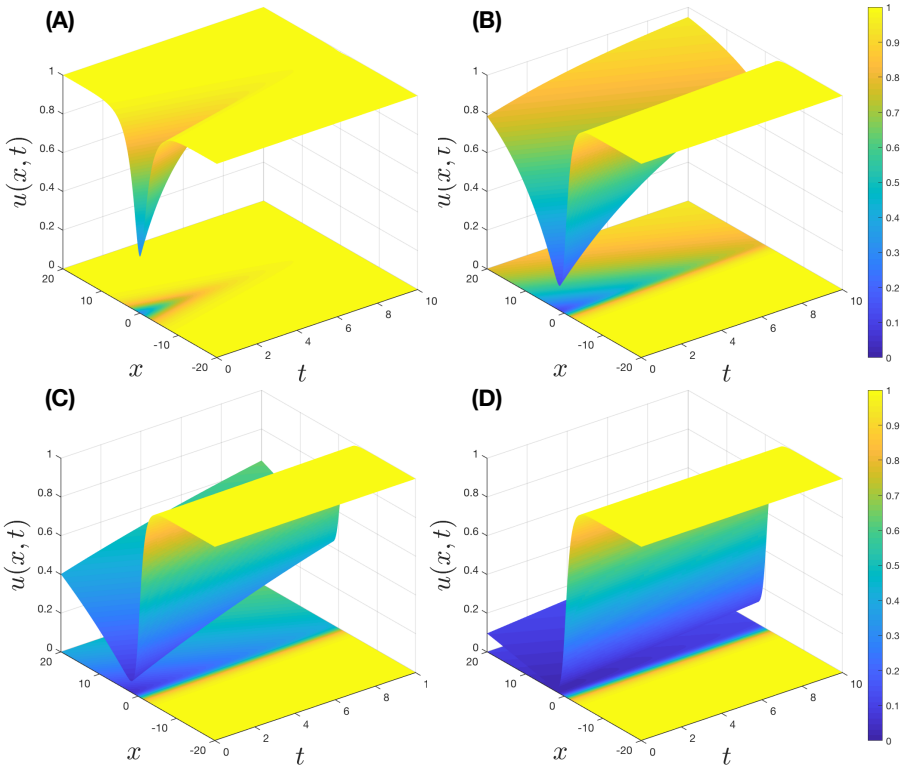


Figure 6.5: **Asymptotic behaviour of the solution $u = u(x,t)$ in Eq. (6.79) for different values of x and t** Solution u is represented for different values of k and modelled by both surface representations as well as density coloured plots. The simulations provided correspond to the values **(A)** $k = \frac{1}{2}$, **(B)** $k = \frac{1}{7}$, **(C)** $k = \frac{1}{20}$ and **(D)** $k = \frac{1}{100}$. The interface level is centred at $x_0 = 0$, so that a density difference is modelled. Whenever k decreases, the density recovery rate when passing through the interface is lower, as well as the density minimal value. This is consistent with a higher diffusion and lower proliferation levels. The parameter k would then model different scenarios for tumour invasion, with a lower k for very infiltrative tumours and a higher k for non-diffusive but proliferating tumours. Units of x and t are considered as mm and $days$, respectively, as in [281].

PART IV

Conclusions and open problems

CHAPTER 7

Thesis conclusions and open problems

Thesis conclusions

In this thesis it has been studied the plausibility of mathematical models to describe cancer dynamics and the potential of relapse prediction.

The extracted conclusions are the following:

1. Mathematical models have proved to be an essential asset in biomedicine. Haematological diseases are well suited to mathematical modelling, not only with differential equations, but also with stochastic models or other techniques. Therefore, there is a huge amount of data to combine with the mathematical models already in the current literature. Even so, these models may not be sufficient to characterise specific disease behaviours in leukaemia diagnosis: one could take, for example, acute lymphoblastic leukaemia dynamics as a particularly undeveloped issue, as studies of chronic myeloid leukaemia appear to us to have attracted more attention. This is probably because myeloid malignancies are most common in adults.

2. We have constructed and studied several non-linear compartmental models describing B cell lymphocyte reconstitution. These simple models describe the process of B-cell generation as portrayed by bone marrow data, and we consider it a first step in a deeper exploration of the phenomena associated with B-cell development. We verified mathematical and biological consistence, opening the door to interesting mathematical

research like the existence of bifurcations or the conditions for global stability, something that finds immediate application in cases of immune reconstitution. Studies of this kind can function as a source of hypothesis generation in biomedical research, for example when contrasting mouse versus human dynamics.

3. We have designed an intuitive algorithm allowing to identify on diagnosis patients with potential of relapse versus those with no risk of relapse in B-cell childhood acute lymphoblastic leukaemia. The most consistent result was the association between a lower expression of CD38 marker and relapse. CD38 is a surface receptor present in a broad variety of immune cells. Our results aligned with the evidence presented in the literature, suggesting that a higher frequency of low CD38 expressing B cells could be an early indicator of relapse risk.

4. We have combined topological methods, specifically persistent homology, and machine-learning techniques, to distinguish relapsing from non-relapsing patients with complete certainty. The analysis based on CD10, CD20, CD38 and CD45 expression allowed for a perfect classification of patients according to their relapse status. It is relevant that these markers are available during routine diagnosis and were obtained via an unsupervised Random Forest method. Classification was performed by using Support Vector Machine and Logistic Regression applied to persistence images of both cohorts of patients. Besides, connected components and one-dimensional loops distinguished via persistence homology between relapsed and non-relapsed patients.

5. We have examined from the point of view of Lie symmetries generalised Fisher's equations that model biological invasions. We have obtained analytical solutions for models that can describe tumour progression. Finally, we have qualitatively studied a special case which simulates the cell density behaviour around an interface. This study is consistent with the biological description of cell density and motility between the grey and white matter of the brain. These results can be used as predictive tools or as a means of understanding tumour growth dynamics.

To conclude, we would like to emphasize the multidisciplinary work that has been initiated with the development of this thesis. By creating these collaborations, we were able to comprehend several biomedical issues from the hand of immunology and haematology experts (Jerez, Niño Jesús, or Virgen

del Rocío Hospital): firstly, biological processes, such as the development of leukaemia invasion and progression; secondly, medical procedures, such as bone marrow transplants; finally, we highlight flow cytometry techniques, including the gating procedures studied to account for leukaemic clones. We note that, even if both the mathematical and biological understanding is quite a remarkable and complex issue, the preprocessing of the data and the understanding of its structure can be considered also as an essential asset to obtain plausible conclusions. In this sense, when considering leukaemia data, we were able to consider quite specific mathematical methods to capture tumour dynamics with differential equations (University of Heidelberg, University of Cádiz), and to classify patients in terms of relapse risk, by both topological and machine learning methods (University of Oxford, University of Oviedo). All this work supposed a considerable number of main collaborators and specialists which are listed below in Table 7.1. We include as well other collaborators in Table 7.2, who would be helping in our future work, not only by widening the knowledge specified above, but also by offering new data to be included in this line of research.

As a final statement, and despite the importance of the models presented, we highlight that the only way to integrate them into clinical practice successfully is through collaboration between mathematicians, biomedical scientists and clinicians. This can lead to new questions and conclusions for both mathematical models and biological problems. The development of such a useful weapon against cancer should be unified, so that the models can be helpful for the actual observation and treatment of disease in patients, beyond the theoretical framework. Mathematical models require refinement in terms of being included in hospital protocols, as a diagnostic or prognostic tool and this can only be achieved by cooperation between the mathematical and medical world.

Open problems and future research

Given the results obtained in this thesis, new questions and problems arise which could be considered as future research:

1. Firstly, when it comes to the mathematical modelling of lymphopoiesis, the choice of three compartments could be refined or expanded following a more detailed characterisation of the cells. A mathematical model where surface markers vary continuously might be able to capture this variation of maturation stages. Secondly, we described signalling as a systemic phenomenon. While this was enough to recapitulate known

Collaborator	Field	Location
Cristina Blázquez-Goñi	Haematology	Jerez Hospital, Jerez de la Frontera, Spain
Juan Francisco Rodríguez-Gutiérrez	Immunology	Jerez Hospital, Jerez de la Frontera, Spain
Lourdes Hermosín-Ramos	Molecular Biology	Jerez Hospital, Jerez de la Frontera, Spain
Águeda Molinos-Quintana	Paediatrics	Virgen del Rocío Hospital, Sevilla, Spain
Teresa Caballero-Velázquez	Immunology	Virgen del Rocío Hospital, Sevilla, Spain
Manuel Ramírez-Orellana	Haematology	Hospital Infantil del Niño Jesús, Madrid, Spain
Ana Castillo-Robleda	Immunology	Hospital Infantil del Niño Jesús, Madrid, Spain
Víctor M. Pérez-García	Mathematics	University of Castilla-La Mancha, Ciudad Real, Spain
María Rosa-Durán	Mathematics	University of Cádiz, Cádiz, Spain
Álvaro Martínez-Rubio	Mathematics	University of Cádiz, Cádiz, Spain
Juan Luis Fernández-Martínez	Mathematics	University of Oviedo, Oviedo, Spain
Anna Marciniak-Czochra	Mathematics	University of Heidelberg, Heidelberg, Germany
Thomas Stiehl	Mathematics and Medicine	University of Heidelberg, Heidelberg, Germany
Helen Byrne	Mathematics	University of Oxford, United Kingdom
Bernadette J. Stolz	Mathematics	University of Oxford, United Kingdom
María Luz Gandarias-Núñez	Mathematics	University of Cádiz, Cádiz, Spain
Rita Tracinà	Mathematics	University of Catania, Catania, Italy

Table 7.1: Main collaborators in the work presented in this thesis.

Collaborator	Field	Location
Odelaisy León-Triana	Mathematics	University of Castilla-La Mancha, Ciudad Real, Spain
Soukaina Sabir	Mathematics	Mohammed V University of Rabat, Rabat, Morocco
Gabriel F. Calvo	Mathematics	University of Castilla-La Mancha, Ciudad Real, Spain
Juan Belmonte-Beitia	Mathematics	University of Castilla-La Mancha, Ciudad Real, Spain
Mariano Torrisi	Mathematics	University of Catania, Catania, Italy
Masood Khalique	Mathematics	North-West University, Potchefstroom, South Africa
Antonio Pérez-Martínez	Mathematics	La Paz Hospital, Madrid, Spain
Gema Salas-Camacho	Haematology	Virgen de las Nieves Hospital, Granada, Spain
Jose Luis Fuster-Soler	Haematology	Virgen de la Arrixaca Hospital, Murcia, Spain
Alfredo Minguela-Puras	Immunology	Virgen de la Arrixaca Hospital, Murcia, Spain
Inmaculada Marchante-Cepillo	Haematology	Puerta del Mar Hospital, Cádiz, Spain
Almudena Sampalo-Lainz	Immunology	Puerta del Mar Hospital, Cádiz, Spain
Roberto Raynero-Mellado	Haematology	Virgen de la Salud Hospital, Toledo, Spain
Nerea Domínguez-Pinilla	Haematology	Virgen de la Salud Hospital, Toledo, Spain

Table 7.2: Other collaborators for the future work presented in this thesis.

B-cell behaviour, a more detailed description including two or more types of signalling is desirable. The model would benefit from longitudinal data coming from immune reconstitution of the B-cell line. In this regard, flow cytometry analyses of both peripheral blood and bone marrow in routine follow-up would allow for a more precise parametrisation and enable the hypotheses presented above to be contrasted.

2. A mathematical model of leukaemia could be studied considering the basics of lymphopoiesis. It could include the interaction between constantly renewing B cells, treatment (whether its chemotherapy or new therapies like CAR-T cells) and leukaemic cells in the bone marrow. By doing this, we could shed some light on the clinical features that have been observed in clinical trials and thus provide a platform for testing hypothesis related to the success or failure of the therapy. Besides, by clustering flow cytometry data we could target the stage of the leukaemic clone, while high-dimensional classification methods could improve risk stratification. This could provide new tools to account for the proportion of leukaemic cells along time and thus be related to other classification techniques regarding relapse that we have considered in this thesis.

3. The nature of the data used in the second part of the study is a recurrent concern in artificial intelligence in haematology. Apart from having only 54 patients, the set of relapsing patients represented only 25% of the whole dataset and that unbalance could introduce biases in the analysis and make the results sensitive to overfitting. In the future, as we increase the size of our dataset, it would be better to increase the number of patients to carry out a 50/50 analysis. Further, there is the issue of data variability, given that it was collected retrospectively, belonging to patients from different years and hospitals. This highlights the importance of the preprocessing routine, which is also amenable to improvements in order to ensure the comparability of the samples.

4. Concerning the mathematical methods for the analysis of biomarkers, we highlight that we can miss information that can be clinically relevant and that refers not to the frequency of cells or intensity of expression, but to the presence or absence of a given subpopulation. Again, we can expect to solve or damp the accuracy of prediction by complementing this work with the inclusion of other clinical data like cytogenetics and molecular biology information, also relevant in the prognostic assessment of haematological diseases.

5. As for the topological study of the data, we could extend this work in many ways. Firstly, the shape of data after chemotherapy, transplant or CAR-T cell infusion could reveal significant information about the development of resistances to those therapies. Moreover, a simple application of the topological methods could compare the shape of the data in the moment of the diagnosis versus whenever a patient relapses, in order to obtain topological features that could distinguish them. Lastly, the flow-cytometry methods can provide additional information on the bone marrow cells and not only on the B-lymphocytes. The topological study could be applied to other haematological conditions, such as lymphomas, T-cell leukaemias, myeloblastic disorders, etc.

6. Regarding Lie symmetries methods, we highlight the importance of the theoretical studies presented, but again, a factual model should include patients' data for it to be applicable at the clinical world. However, these methods could be plausibly applied to the cancer-related mathematical models worked in the future. This could be done so in order to comprehend some dynamical behaviours that are intrinsic to the model itself, thus obtaining solutions that may not be the most suitable ones but may shed some light on the understanding of cellular interaction, mobility or invasion features.

CHAPTER 8

Conclusiones de la tesis y vías de ampliación

Conclusiones de la tesis

En esta tesis se ha estudiado la capacidad de los modelos matemáticos para describir la dinámica del cáncer y el potencial de la predicción de recaídas.

Las conclusiones que se derivan son las siguientes:

1. Los modelos matemáticos se manifiestan como una herramienta fundamental en problemas biomédicos. Las enfermedades hematólogicas se adecuan perfectamente al modelaje matemático, no sólo mediante ecuaciones diferenciales, sino también con modelos estocásticos y otras técnicas. Existe así una enorme cantidad de datos que poder combinar con modelos matemáticos ya existentes en la literatura. De igual modo, estos modelos pueden no ser lo suficientemente aptos para caracterizar comportamientos específicos de la enfermedad en el diagnóstico de la leucemia: se puede considerar, por ejemplo, a la dinámica de la leucemia linfoblástica aguda como un problema especialmente poco desarrollado.

2. Se han construido y estudiado múltiples modelos compartimentales no-lineales que muestran la reconstitución de linfocitos B. Estos modelos describen la regeneración de células B tal y como se observa en datos de médula ósea. Lo consideramos así un primer paso en una exploración más amplia de este fenómeno, especialmente asociado al desarrollo de este tipo de células. Se ha verificado la consistencia tanto matemática como

biológica de los resultados, abriendo así la posibilidad a investigaciones matemáticas de otra índole, como la existencia de bifurcaciones o condiciones para la estabilidad global. Esta cuestión puede encontrar una aplicación inmediata en los casos de la reconstitución inmune. Estudios de este tipo pueden funcionar como una fuente de hipótesis en el ámbito biomédico, por ejemplo, al contrastar dinámica entre modelos murinos y humanos.

3. Se ha diseñado un algoritmo intuitivo capaz de identificar en el momento de diagnosis a los pacientes con potencial de recaída en leucemia linfoblástica aguda pediátrica tipo B. El resultado más consistente es el de la asociación entre una baja expresión del marcador CD38 y la recaída en la enfermedad. Este marcador de superficie se presenta en una amplia variedad de células del sistema inmune. Nuestros resultados se ajustan a las evidencias presentes en la literatura, sugiriendo así que una alta frecuencia de células con bajo CD38 puede ser un indicador temprano de riesgo de recaída.

4. Se han combinado métodos topológicos, específicamente de homología de persistencia, con técnicas de "machine-learning", para distinguir entre pacientes que recaen o no con seguridad plena. El análisis basado en la expresión de CD10, CD20, CD38 y CD45 permite una clasificación perfecta de los pacientes según la posibilidad o no de recaída. Es importante mencionar que estos marcadores se encuentran disponibles en el diagnóstico de rutina y se han obtenido a través de un método de "Random-forest" de manera no-supervisada. La clasificación se ha realizado a través de métodos como "Support Vector Machine" y de Regresión Logística, aplicados a imágenes de persistencia de ambas cohortes de pacientes. Además, la existencia de componentes conexas y de bucles unidimensionales es capaz de distinguir a través de esta homología de persistencia entre pacientes con y sin recaída.

5. Se han examinado, desde el punto de vista de las simetrías de Lie, ecuaciones de Fisher generalizadas que modelan la invasión biológica. Se han obtenido soluciones analíticas de modelos que son capaces de describir la progresión de un tumor. Finalmente, se ha estudiado cualitativamente un caso especial que simula el comportamiento de la densidad celular alrededor de una interfaz tumoral. Este estudio es consistente con la descripción de la densidad celular y su movilidad entre la materia gris y blanca cerebral. Estos resultados son de utilidad considerándolos como

herramientas predictivas o como medio para comprender la dinámica de crecimiento tumoral.

Para concluir, se quiere enfatizar en el carácter multidisciplinar que ha requerido el desarrollo de esta tesis. Mediante estas colaboraciones, se han podido comprender múltiples problemas biomédicos de la mano de expertos en inmunología y hematología (Hospital de Jerez, Niño Jesús o Virgen del Rocío): en primer lugar, procesos biológicos, como lo es el desarrollo de la leucemia, su invasión y progresión; procedimientos médicos, como el de trasplante de médula ósea; finalmente, otros como las técnicas de citometría de flujo, incluyendo así los procedimientos de selección para el conteo de clones leucémicos. Queremos remarcar que, aunque la comprensión tanto matemática como biológica son un problema complejo y de interés, el propio preprocesado de los datos, así como entender su estructura interna, se puede considerar también como un elemento necesario para poder obtener conclusiones sólidas. En este aspecto, al considerar datos de leucemia, se ha podido contar con métodos matemáticos específicos que son capaces de reflejar, por un lado, la dinámica tumoral mediante ecuaciones diferenciales (Universidad de Heidelberg, Universidad de Cádiz), y, por otro lado, la capacidad de clasificación para el riesgo de recaída, tanto por métodos topológicos como de "machine learning" (Universidad de Oxford, Universidad de Oviedo). Todo este trabajo conlleva un elevado número de colaboradores principales que se encuentran en la Tabla 8.1. Incluimos a su vez otros colaboradores en la Tabla 8.2, que estarían en disposición de colaborar en proyectos futuros, no sólo al ampliar los estudios mencionados anteriormente, sino también a la hora de proporcionar nuevos datos que incluir en esta línea de investigación.

Como conclusión final, y a pesar de la importancia de los modelos teóricos presentados, el modo de integrarlos convenientemente en la práctica clínica se basa en la creación de colaboraciones entre matemáticos, científicos biomédicos y personal clínico. Esto puede llevar al planteamiento de nuevas preguntas y conclusiones tanto de modelos matemáticos como de problemas biológicos. El desarrollo de un arma contra el cáncer, como es la de los modelos matemáticos, debe ser unificado, de manera que puedan ser útiles para la observación del mundo real y para el tratamiento de enfermedades, alejándose del entorno puramente teórico. Los modelos matemáticos requieren de una optimización para poder así ser incluidos en protocolos hospitalarios, sea como herramienta de diagnóstico o de pronóstico, y esto solo puede llevarse a cabo a través de la cooperación entre el mundo médico y matemático.

Colaborador	**Campo**	**Lugar**
Cristina Blázquez-Goñi	Hematología	Hospital de Jerez, Jerez de la Frontera, España
Juan Francisco Rodríguez-Gutiérrez	Inmunología	Hospital de Jerez, Jerez de la Frontera, España
Lourdes Hermosín-Ramos	Biología Molecular	Hospital de Jerez, Jerez de la Frontera, España
Águeda Molinos-Quintana	Pediatría	Hospital Virgen del Rocío, Sevilla, España
Teresa Caballero-Velázquez	Inmunología	Hospital Virgen del Rocío, Sevilla, España
Manuel Ramírez-Orellana	Hematología	Hospital Infantil del Niño Jesús, Madrid, España
Ana Castillo-Robleda	Inmunología	Hospital Infantil del Niño Jesús, Madrid, España
Víctor M. Pérez-García	Matemáticas	Universidad de Castilla-La Mancha, Ciudad Real, España
María Rosa-Durán	Matemáticas	Universidad de Cádiz, Cádiz, España
Álvaro Martínez-Rubio	Matemáticas	Universidad de Cádiz, Cádiz, España
Juan Luis Fernández-Martínez	Matemáticas	Universidad de Oviedo, Oviedo, España
Anna Marciniak-Czochra	Matemáticas	Universidad de Heidelberg, Heidelberg, Germany
Thomas Stiehl	Matemáticas y Medicina	Universidad de Heidelberg, Heidelberg, Germany
Helen Byrne	Matemáticas	Universidad de Oxford, Reino Unido
Bernadette J. Stolz	Matemáticas	Universidad de Oxford, Reino Unido
María Luz Gandarias-Núñez	Matemáticas	Universidad de Cádiz, Cádiz, España
Rita Tracinà	Matemáticas	Universidad de Catania, Catania, Italy

Table 8.1: Colaboradores principales del trabajo presentado en esta tesis.

Colaborador	Campo	Lugar
Odelaisy León-Triana	Matemáticas	Universidad de Castilla-La Mancha, Ciudad Real, España
Soukaina Sabir	Matemáticas	Mohammed V Universidad de Rabat, Rabat, Marruecos
Gabriel F. Calvo	Matemáticas	Universidad de Castilla-La Mancha, Ciudad Real, España
Juan Belmonte-Beitia	Matemáticas	Universidad de Castilla-La Mancha, Ciudad Real, España
Mariano Torrisi	Matemáticas	Universidad de Catania, Catania, Italia
Masood Khalique	Matemáticas	Universidad North-West , Potchefstroom, Sudáfrica
Antonio Pérez-Martínez	Matemáticas	Hospital de La Paz, Madrid, España
Gema Salas-Camacho	Hematología	Hospital Virgen de las Nieves, Granada, España
Jose Luis Fuster-Soler	Hematología	Hospital Virgen de la Arrixaca, Murcia, España
Alfredo Minguela-Puras	Inmunología	Hospital Virgen de la Arrixaca, Murcia, España
Inmaculada Marchante-Cepillo	Hematología	Hospital Puerta del Mar, Cádiz, España
Almudena Sampalo-Lainz	Inmunología	Hospital Puerta del Mar, Cádiz, España
Roberto Raynero-Mellado	Hematología	Hospital Virgen de la Salud, Toledo, España
Nerea Domínguez-Pinilla	Hematología	Hospital Virgen de la Salud, Toledo, España

Table 8.2: Otros colaboradores principales para el trabajo futuro a desarrollar presentado en esta tesis.

Vías de ampliación e investigación futura

Dados los resultados obtenidos en esta tesis, surgen nuevas preguntas y problemas que se puede considerar como proyectos futuros:

1. En lo que respecta al modelaje matemático de la linfopoiesis, el hecho de haber distinguido en tres compartimentos celulares podría ser refinado y/o expandido siguiendo una caracterización más detallada del tipo celular. Un modelo matemático que considere una continuidad en los marcadores de superficie de las células puede ser capaz de capturar esta variación en los estados de maduración. En segundo lugar, se ha descrito a la señalización celular como un fenómeno sistémico. Aunque esto haya sido suficiente para recabar el ya conocido comportamiento de las células B, sería conveniente una descripción más detallada, que pueda incluir dos o más tipos de señalización. El modelo se beneficiaría de datos longitudinales provenientes de reconstitución inmune de la línea B. En este aspecto, los análisis de seguimiento y rutina de citometría de flujo tanto de sangre periférica como de médula ósea permitirían una parametrización más precisa capaz de contrastar las hipótesis presentadas con anterioridad.

2. Un nuevo modelo matemático de leucemia se podría estudiar considerando la base de la linfopoiesis. Podría incluir la interacción tanto entre células B en continua renovación, como con el tratamiento (tanto si es quimioterapia como nuevas terapias como las células CAR-T) y junto a las células leucémicas en la médula ósea. A través de esto, podríamos esclarecer las características clínicas que se observan en los ensayos clínicos y así proporcionar una plataforma de testeo de hipótesis asociadas al éxito o fallo de la terapia. Además, a través de métodos de "clustering", se podría seleccionar el estadio del clon leucémico, mientras que otras técnicas de clasificación en datos de citometría de flujo podría mejorar la asignación del riesgo del paciente. Esto puede proporcionar nuevas herramientas que consideren la proporción de células leucémicas a lo largo del tiempo y relacionarlo así con otras técnicas de clasificación de recaída que se hayan considerado en esta tesis.

3. La naturaleza de los datos usados en la segunda parte de este estudio es un problema recurrente en la inteligencia artificial asociada a la hematología. Además de únicamente contar con 54 pacientes, el conjunto de pacientes con recaída representa solo el 25% del total de los datos, siendo esta descompensación un factor de parcialidad en el análisis, haciendo los

resultados sensibles a “overfitting”. En el futuro, con el incremento del tamaño de nuestra muestra de datos, sería aconsejable incrementar el número de pacientes y hacer un análisis 50/50. Más allá de esto, se da el problema de la variabilidad de los datos, dado que se han recolectado de manera retrospectiva, siendo de pacientes de años y hospitales diferentes. Esto realza la importancia de la rutina de preprocesado, que es sensible también a mejora para asegurar la comparabilidad entre las muestras.

4. Según los modelos matemáticos de análisis de biomarcadores, se recalca que se ha podido perder información que puede ser clínicamente relevante y que se refiere no tanto a la frecuencia del número de células o a su intensidad de expresión en el marcaje celular, sino a la presencia o ausencia de una subpoblación específica. Así, se puede solventar o mejorar la capacidad de precisión de la predicción al complementar este trabajo con la inclusión de otros datos clínicos como información citogenética o de biología molecular, ya que se considera también relevante en la evaluación de enfermedades hematológicas.

5. Conforme a los estudios topológicos de los datos, se podría extender su trabajo de múltiples modos. Primero, la forma de los datos tras quimioterapia, transplante de progenitores hematopoyéticos, o infusión de células CAR-T podrían revelar información acerca del desarrollo de resistencias a dichas terapias. Además, una simple aplicación de los métodos topológicos podría comparar la forma de los datos en el momento de diagnosis y en el momento de recaída del paciente, para así obtener características topológicas para distinguirlos. Finalmente, los métodos de citometría de flujo pueden proporcionar información, ya no solo de los linfocitos B, sino también de otras células de la médula ósea. Este estudio topológico podría ser aplicado a otras condiciones hematológicas, como pueden ser linfomas, leucemias tipo T o leucemias mieloides.

6. En cuanto a los métodos de simetrías de Lie, recalcamos la importancia de los estudios teóricos presentados, aunque un modelo real debería de incluir datos de pacientes para poder ser aplicable en el mundo clínico. Sin embargo, estos métodos podrían ser aplicados convenientemente a modelos matemáticos relacionados con el cáncer que se desarrollen en el futuro. Esto podría ser realizado para entender los comportamientos dinámicos que fuesen intrínsecos al modelo, obteniendo así soluciones que pueden no ser las más apropiadas pero que pueden arrojar algo de luz en

la comprensión de interacciones celulares, movilidad o características de invasión.

CHAPTER 9

Publications and Congress contributions

Publications and preprints derived from this thesis

[P1] Authors: Salvador Chulián, Alvaro Martínez-Rubio, María Rosa, Víctor M Pérez-García.
Title: Mathematical models of Leukaemia and its treatment: A review.
Journal: arXiv, submitted to *SEMA Journal*.
Year: 2020.
Category: Mathematics, Applied.
DOI: arXiv preprint 2011.05881.

[P2] Authors: Salvador Chulián, Alvaro Martínez-Rubio, Anna Marciniak-Czochra, Thomas Stiehl, Cristina Blázquez Goñi, Juan Francisco Rodríguez Gutiérrez, Manuel Ramírez Orellana, Ana Castillo Robleda, Víctor M Pérez-García, María Rosa.
Title: Dynamical properties of feedback signalling in B lymphopoiesis: A mathematical modelling approach.
Journal: arXiv, submitted to *Journal of Theoretical Biology*.
Year: 2020.
Category: Biology, Mathematical & Computational Biology, Biology Miscellaneous. Quality. Impact-factor: 2.327 Ranking: Q2.
DOI: arXiv preprint 2007.13526.

[P3] Authors: Salvador Chulián, Álvaro Martínez-Rubio , Víctor M. Pérez-García, María Rosa, Cristina Blázquez-Goñi, Juan Francisco Rodríguez-

Gutiérrez, Lourdes Hermosín-Ramos, Águeda Molinos Quintana, Teresa Caballero-Velázquez, Manuel Ramírez-Orellana, Ana Castillo-Robleda, Juan Luis Fernández-Martínez.
Title: *High-dimensional analysis of single-cell flow cytometry data predicts relapse in Childhood Acute Lymphoblastic Leukemia.*
Journal: *Preprints, submitted to Cancers.*
Year: 2020.
Category: Oncology. Quality. Impact-factor: 6.126, Ranking: Q1.
DOI: preprint 202010.0557.v1.

[P4] Authors: Salvador Chulián, Bernadette J. Stolz, Álvaro Martínez-Rubio, Cristina Blázquez Goñi, Juan Francisco Rodríguez Gutiérrez, Águeda Molinos Quintana, Teresa Caballero-Velázquez, Manuel Ramírez-Orellana, Ana Castillo-Robleda, María Rosa, Víctor M. Pérez-García, Helen Byrne.
Title: The shape of cancer relapse: Topological data analysis predicts recurrence in paediatric acute lymphoblastic leukaemia.
Journal: Preprint, in process for *Sciences Advances.*
Year: 2020.
Category: Multidisciplinary Sciences.
Quality. Impact-factor: 13.117, Ranking: Q1.

[P5] Authors: Salvador Chulián, María Rosa, María Luz Gandarias.
Title: Reductions and symmetries for a generalized Fisher equation with a diffusion term dependent on density and space.
Journal: *Journal of Computational and Applied Mathematics.*
Volume: 354. Pages: 689-698.
Year: 2019.
Category: Mathematics, Applied.
Quality. Impact-Factor: 2.037, Ranking: Q1.
DOI: 10.1016/j.cam.2018.11.018

[P6] Authors: Salvador Chulián, María Rosa, María Luz Gandarias.
Title: Symmetries and solutions for a Fisher equation with a proliferation term involving tumor development.
Journal: *Mathematical Methods in Applied Sciences.*
Volume: 43. Pages: 2076–2084.
Year: 2020.
Category: Mathematics, Mathematics Applied.

Quality. Impact-Factor: 1.626, Ranking:, Q1.
DOI: 10.1002/mma.6105.

[P7] Authors: Salvador Chulián, María Rosa, María Luz Gandarias, Rita Tracinà.
Title: Application of Lie point symmetries to the resolution of an interface problem in a generalized Fisher equation
Journal: Physica D: Nonlinear phenomena.
Volume: 405. Pages: 132411.
Year: 2020.
Category: Physics, Mathematics Applied.
Quality. Impact-Factor: 1.810, Ranking:, Q1.
DOI: 10.1016/j.physd.2020.132411.

Other publications

[P8] Authors: Odelaisy León-Triana, Soukaina Sabir, Gabriel F. Calvo, Juan Belmonte-Beitia, Salvador Chulián, Álvaro Martínez-Rubio, María Rosa, Antonio Pérez-Martínez, Manuel Ramírez Orellana, Víctor M. Pérez-García.
Title: CAR T cell therapy in B-cell acute lymphoblastic leukaemia: Insights from mathematical models.
Journal: Communications in Nonlinear Science and Numerical Simulation.
Volume: 94 Pages: 105570.
Year: 2021.
Cathegory: Mathematics, applied.
Quality. Impact-factor: 4.115, Ranking: Q1.
DOI: 10.1016/j.cnsns.2020.105570.

[P9] Authors: Salvador Chulián, María Rosa, Pilar Azcárate.
Title: A tool for content analysis in textbooks: first-degree equations.
Journal: Volume: 103. Pages:25-33.
Year: 2018.
Category: Basic and experimental sciences: Mathematicss. Psicology and education: applied didactics.
Quality. Impact-Factor: Dialnet 0.052, Ranking: Education 2019 182/226.
Available online on this link.

Congress contributions and invited talks

[O1] Authors: S. Chulián.
Title: TDA analysis of flow cytometry data in acute lymphoblastic leukaemia patients.
Location: Mathematical Institute, University of Oxford, United Kingdom.
Date: September 2020.

[O2] Authors: S. Chulián.
Title: Blood cancer: The mathematics of Lymphoblastic Leukemia.
Congress: Therapy Optimization in Glioblastoma 2019: An integrative human data-based approach using mathematical models. Location: Córdoba, Spain.
Date: September 2019.

[O3] Authors: S. Chulián.
Title: Mathematical Models and Data Analysis on Lymphoblastic Leukemia.
Congress: ICIAM 2019: 9th International Congress on Industrial and Applied Mathematics.
Location: Valencia, Spain.
Date: July 2019.

[O4] Authors: S. Chulián.
Title: Symmetries and solutions for a Fisher equation with a proliferation term involving tumor development.
Congress: 19th International Conference on Mathematical Methods in Science and Engineering.
Location: Rota, Cádiz, Spain.
Date: July 2019.

[O5] Authors: S. Chulián.
Title: New biomarkers for recurrence in LLA via discriminant analysis.
Congress: XII National Congress of the Spanish Society of Haematology and Paediatric Oncology.
Location: Jerez de la Frontera, Cádiz, Spain.
Date: June 2019.

[O6] Authors: S. Chulián.
Title: A tool for content analysis in textbooks: first-degree equations.

Congress: I Ibero-american Congress for Teachers 2018.
Location: Algeciras, Cádiz, Spain.
Date: December 2018.

[O8] Authors: S. Chulián.
Title: Symmetry analysis for a generalized Fisher equation with density-space-dependent diffusion.
Congress: 18th International Conference on Mathematical Methods in Science and Engineering.
Location: Rota, Cádiz, Spain.
Date: July 2018.

Appendices

APPENDIX A

Stability analysis for models 3.6

A.1 Stability analysis for non-trivial equilibria in model A

A.1.1 Model A1

We recall the model from Section **??**. From Eq. (3.22) we obtained the steady states P_i^{A1} for $i = 1, 2, 3$.

Let us consider stability for the equilibrium point P_3^{A2}. We obtain the characteristic equation

$$\lambda^3 + b_2\lambda^2 + b_1\lambda + b_0 = 0, \tag{A.1}$$

where

$$b_2 = \alpha_2 + \alpha_3 - \frac{\alpha_1\rho_2}{\rho_1}, \tag{A.2a}$$

$$b_1 = \alpha_3\left(\alpha_2 - \frac{\alpha_1^2\rho_2}{\rho_1^2}\right), \tag{A.2b}$$

$$b_0 = \frac{\alpha_1\alpha_3(\alpha_1 - \rho_1)(\alpha_1\rho_2 - \alpha_2\rho_1)}{\rho_1^2}. \tag{A.2c}$$

Using the Routh-Hurwitz Criterion, for P_3^{A1} to be positive and stable we must have

$$b_2b_1 - b_0 > 0, b_2 > 0, b_0 > 0. \tag{A.3}$$

The positivity conditions found in Eq. (3.28) yield $b_0 > 0, b_2 > 0$. Furthermore, the stability condition $b_2b_1 - b_0 > 0$ is satisfied if $\rho_1 \leq \rho_2$. If $\rho_1 > \rho_2$, the stability criterion is equivalent to satisfying either

$$\alpha_1^2\rho_2 \leq \rho_1(\alpha_1^2 - \alpha_1\rho_1 + \alpha_2\rho_1) \tag{A.4}$$

or

$$\frac{\alpha_2\rho_1^2\big(\alpha_1^2+\rho_1(\alpha_2+\alpha_3)\big)+\alpha_1^3\rho_2^2}{\alpha_1\rho_1} > \alpha_1\rho_2\,(\alpha_T-\rho_1)+\alpha_2\rho_1(\rho_1+\rho_2), \quad \text{(A.5)}$$

where $\alpha_T=\sum_{i=1}^{3}\alpha_i$.

A.1.2 Model A2

We recall the model from Section **??**. We obtained in Eq. (3.29) the steady states P_i^{A2} for $i=1,2,3$.

We use the Routh-Hurwitz Criterion to study the stability of P_3^{A2}. We obtain the characteristic equation

$$\lambda^3+b_2\lambda^2+b_1\lambda+b_0=0, \quad \text{(A.6)}$$

where

$$b_2=\frac{(\alpha_2\alpha_3(\alpha_2+\alpha_3)\rho_1^2+\alpha_1\ \rho_1(\alpha_2^2\rho_1+\alpha_2\alpha_3(3\rho_1-2\rho_2)}{\rho_1\beta}+ \quad \text{(A.7a)}$$
$$+\frac{\alpha_3^2(\rho_1-\rho_2))-\alpha_1^2(\ \alpha_3(\rho_1-\rho_2)\rho_2+\alpha_2\rho_1(\alpha_3\ \ +\rho_2))}{\rho_1\beta},$$
$$b_1=\frac{\alpha_3\big(\alpha_2^2\alpha_3\rho_1^3+2\alpha_1\alpha_2\rho_1^2(\alpha_2\rho_1+\alpha_3(\rho_1-\rho_2))-\alpha_1^3\rho_1(\rho_1-\rho_2)(\alpha_2+\rho_2)\big)}{\rho_1^2\beta}+ \quad \text{(A.7b)}$$
$$+\frac{\alpha_3\alpha_1^4(\rho_1-\rho_2)\rho_2}{\rho_1^2\beta}-\frac{\alpha_3\alpha_1^2\rho_1(\alpha_2^2\rho_1+\alpha_3(\rho_1-\ \rho_2)\rho_2+\alpha_2\rho_1(\alpha_3-\rho_1+2\rho_2))}{\rho_1^2\beta},$$
$$b_0=\frac{\alpha_1\alpha_3(\alpha_1-\rho_1)(\alpha_1\rho_2-\alpha_2\rho_1)}{\rho_1^2}. \quad \text{(A.7c)}$$

Positivity conditions for P_3^{A2} as in Eq. (3.35) and in Eq. (3.36) yield $b_0>0$. Finally, stability conditions $b_2>0$ and $b_2b_1-b_0>0$ result in

$$\frac{(\alpha_2\alpha_3(\alpha_2+\alpha_3)\rho_1^2+\alpha_1\ \rho_1(\alpha_2^2\rho_1+\alpha_2\alpha_3(3\rho_1-2\rho_2)}{\rho_1\beta}+$$
$$+\frac{\alpha_3^2(\rho_1-\rho_2))-\alpha_1^2(\alpha_3(\rho_1-\rho_2)\rho_2+\alpha_2\rho_1(\alpha_3\ \ +\rho_2))}{\rho_1\beta}>0 \quad \text{(A.8a)}$$

and

$$\frac{(r_1\rho_1^2 + r_2\rho_1 + r_3(\rho_1 - \rho_2))(r_4\rho_1^3 + r_5\rho_1^2 + r_6(\rho_1 - \rho_2)))}{\beta^3} + \\ + \alpha_1(\rho_1 - \alpha_1)\rho_1^2(\alpha_1\rho_2 - \alpha_2\rho_1) > 0 \tag{A.8b}$$

where

$$\begin{aligned} r_1 &= \alpha_1\alpha_2^2 + 3\alpha_1\alpha_2\alpha_3 + \alpha_2\alpha_3(\alpha_2 + \alpha_3), \\ r_2 &= -(\alpha_1^2\alpha_2 + 2\alpha_1\alpha_2\alpha_3)\rho_2 - \alpha_1^2\alpha_2\alpha_3, \\ r_3 &= \alpha_1\alpha_3(\alpha_3\rho_1 - \alpha_1\rho_2), \\ r_4 &= \alpha_1^2\alpha_2 + 2\alpha_1\alpha_2^2 + \alpha_2^2\alpha_3, \\ r_5 &= -\alpha_1^2\alpha_2^2 - \alpha_1^2\alpha_2\alpha_3 - 2\alpha_1^2\alpha_2\rho_2, \\ r_6 &= 2\alpha_1\alpha_2\alpha_3\rho_1^2 + \alpha_1^4\rho_2 - \rho_1\alpha_1^2(\alpha_1\alpha_2 + \rho_2(\alpha_1 + \alpha_3)). \end{aligned} \tag{A.8c}$$

A.2 Stability analysis for models B

A.2.1 Model B1

Let us consider Eqs. (3.8) with last stage signalling $s_\alpha = s_1(t)$ as in Eq. (3.4), this is, $N = C_3$. The steady states for this model are

$$P_1^{B1} = (0, 0, 0), \tag{A.9a}$$

$$P_2^{B1} = \left(0, \frac{\alpha_3(\alpha_2 - \rho_2)}{k\alpha_2\rho_2}, \frac{\alpha_2 - \rho_2}{k\rho_2}\right), \tag{A.9b}$$

$$P_3^{B1} = \left(\frac{\alpha_3(\alpha_1 - \rho_1)(\alpha_2\rho_1 - \alpha_1\rho_2)}{k\alpha_1\alpha_2\rho_1^2}, \frac{\alpha_3(\alpha_1 - \rho_1)}{k\alpha_2\rho_1}, \frac{\alpha_1 - \rho_1}{k\rho_1}\right). \tag{A.9c}$$

The Jacobian matrix is $J_{B1}(C_1, C_2, C_3) = J_{B1}$ such that

$$J_{B1} = \begin{pmatrix} \rho_1 - \dfrac{\alpha_1}{C_3k + 1} & 0 & \dfrac{\alpha_1 C_1 k}{(C_3k + 1)^2} \\ \dfrac{\alpha_1}{C_3k + 1} & \rho_2 - \dfrac{\alpha_2}{C_3k + 1} & \dfrac{\alpha_2 C_2 k - \alpha_1 C_1 k}{(C_3k + 1)^2} \\ 0 & \dfrac{\alpha_2}{C_3k + 1} & -\dfrac{\alpha_2 C_2 k + \alpha_3}{(C_3k + 1)^2} \end{pmatrix}. \tag{A.10}$$

Substituting P_i^{B1} for $i = 1, 2, 3$ in Eq. (A.10) we obtain the eigenvalues governing linear stability. First, for P_1^{B1}, we obtain the same eigenvalues as for P_1^{A1}, i.e.

$$\lambda_{1,1}^{B1} = -\alpha_3, \tag{A.11a}$$

$$\lambda_{1,2}^{B1} = \rho_1 - \alpha_1, \tag{A.11b}$$

$$\lambda_{1,3}^{B1} = \rho_2 - \alpha_2. \tag{A.11c}$$

For P_2^{B1} we get

$$\lambda_{2,1}^{B1} = \rho_1 - \frac{\alpha_1 \rho_2}{\alpha_2}, \tag{A.12a}$$

$$\lambda_{2,2}^{B1} = -\frac{\alpha_3 \rho_2 + \sqrt{\alpha_3}\rho_2 \sqrt{4\alpha_2 + \alpha_3 - 4\rho_2}}{2\alpha_2}, \tag{A.12b}$$

$$\lambda_{2,3}^{B1} = \frac{\alpha_3 \rho_2 - \sqrt{\alpha_3}\rho_2 \sqrt{4\alpha_2 + \alpha_3 - 4\rho_2}}{2\alpha_2}. \tag{A.12c}$$

Finally, for P_3^{B1}, we obtain the characteristic equation

$$\lambda^3 + b_2\lambda^2 + b_1\lambda + b_0 = 0, \tag{A.13}$$

where

$$b_2 = \frac{\alpha_2\rho_1 + \alpha_3\rho_1 - \alpha_1\rho_2}{\alpha_1}, \tag{A.14a}$$

$$b_1 = \frac{\alpha_3\rho_1(\alpha_2\rho_1 + \rho_2(\rho_1 - 2\alpha_1))}{\alpha_1^2}, \tag{A.14b}$$

$$b_0 = \frac{\alpha_3(\alpha_1 - \rho_1)\rho_1^2(\alpha_1\rho_2 - \alpha_2\rho_1)}{\alpha_1^3}. \tag{A.14c}$$

Considering positivity conditions for P_2^{B1} and P_3^{B1}, we find that $\lambda_{1,i}^{B1} < 0$ for $i = 1, 2, 3$, and therefore P_1^{B1} is always stable. From the positivity conditions we also get

$$\frac{\rho_1}{\rho_2} > \frac{\alpha_1}{\alpha_2}. \tag{A.15a}$$

which implies $\lambda_{2,1}^{B1} > 0$ and therefore P_2^{B1} is unstable.

Stability of this equilibrium P_3^{B1} can be analysed by the Routh-Hurwitz Criterion from Eq. (A.3). However, given its own positivity conditions, we get $b_0 < 0$, implying P_3^{B1} is always unstable.

A.2.2 Model B2

Let us now consider Eqs. (3.8) with signalling coming from all cellular compartments $s_\alpha = s_1(t)$ as in Eq. (3.5), this is, $N = \sum_{i=1}^{3} C_i$. The steady states of the model are

$$P_1^{B2} = (0, 0, 0), \tag{A.16a}$$

$$P_2^{B2} = \left(0, \frac{\alpha_3(\alpha_2 - \rho_2)}{k(\alpha_2 + \alpha_3)\rho_2}, \frac{\alpha_2(\alpha_2 - \rho_2)}{k(\alpha_2 + \alpha_3)\rho_2}\right), \tag{A.16b}$$

$$P_3^{B2} = \left(\frac{\alpha_3(\alpha_1 - \rho_1)(\alpha_2\rho_1 - \alpha_1\rho_2)}{\rho_1 k\beta}, \frac{\alpha_1\alpha_3(\alpha_1 - \rho_1)}{k\beta}, \frac{\alpha_1\alpha_2(\alpha_1 - \rho_1)}{k\beta}\right), \tag{A.16c}$$

where

$$\beta = (\alpha_2\alpha_3\rho_1 + \alpha_1(\alpha_2\rho_1 + \alpha_3(\rho_1 - \rho_2))) . \tag{A.17}$$

The Jacobian matrix of Eqs. (3.8) with signal s given by Eq. (3.5) is $J_{B2} = J_{B2}(C_1, C_2, C_3)$ such that

$$J_{B2} = s^2 \begin{pmatrix} C_1 k\alpha_1 - \frac{\alpha_1}{s} + \frac{\rho_1}{s^2} & C_1 k\alpha_1 & C_1 k\alpha_1 \\ \alpha_1 + kR_1 & kR_2 - \frac{\alpha_2}{s} + \frac{\rho_2}{s^2} & kR_2 \\ kR_3 & \alpha_2 + kR_4 & -\alpha_3 - kR_5 \end{pmatrix}, \tag{A.18}$$

where

$$R_1 = C_2(\alpha_1 + \alpha_2) + C_3\alpha_1, \tag{A.19a}$$

$$R_2 = C_2\alpha_2 - C_1\alpha_1, \tag{A.19b}$$

$$R_3 = C_3\alpha_3 - C_2\alpha_2, \tag{A.19c}$$

$$R_4 = C_1\alpha_2 + C_3(\alpha_2 + \alpha_3), \tag{A.19d}$$

$$R_5 = C_1\alpha_3 + C_2(\alpha_2 + \alpha_3). \tag{A.19e}$$

Substituting P_i^{B2} for $i = 1, 2, 3$ in Eq. (A.18) we obtain the eigenvalues governing the linear stability. Specifically, for P_1^{B2}, we again get

$$\lambda_{1,1}^{B2} = -\alpha_3, \tag{A.20a}$$

$$\lambda_{1,2}^{B2} = \rho_1 - \alpha_1, \tag{A.20b}$$

$$\lambda_{1,3}^{B2} = \rho_2 - \alpha_2. \tag{A.20c}$$

For P_2^{B2}, we get the eigenvalues

$$\lambda_{2,1}^{B2} = \rho_1 - \frac{\alpha_1\rho_2}{\alpha_2}, \tag{A.21a}$$

$$\lambda_{2,2}^{B2} = -\frac{\alpha_3^2\rho_2 + \alpha_3\rho_2^2 + h^*(\alpha_2,\alpha_3,\rho_2)}{2\alpha_2(\alpha_2+\alpha_3)}, \tag{A.21b}$$

$$\lambda_{2,3}^{B1} = \frac{-\alpha_3^2\rho_2 - \alpha_3\rho_2^2 + h^*(\alpha_2,\alpha_3,\rho_2)}{2\alpha_2(\alpha_2+\alpha_3)}. \tag{A.21c}$$

where $h^* = h^*(\alpha_2,\alpha_3,\rho_2)$ such that

$$h^* = \sqrt{\alpha_3}\rho_2\sqrt{4\alpha_2(\alpha_2^2+\alpha_2(2\alpha_3-\rho_2)+\alpha_3(\alpha_3-2\rho_2))+\alpha_3(\alpha_3-\rho_2)^2}. \tag{A.22}$$

Finally, for P_3^{B2}, we obtain the characteristic equation

$$\lambda^3 + b_2\lambda^2 + b_1\lambda + b_0 = 0, \tag{A.23}$$

where

$$b_2 = \frac{\alpha_2\alpha_3\rho_1^2(\alpha_2+\alpha_3+\rho_1)+\alpha_1\rho_1(\alpha_2^2\rho_1+\alpha_2\alpha_3(\rho_1-2\rho_2)}{\alpha_1\beta} + \tag{A.24a}$$
$$+\frac{\alpha_3^2(\rho_1-\rho_2))-\alpha_1^2(\alpha_2\rho_1+\alpha_3(\rho_1-\rho_2))\rho_2}{\alpha_1\beta},$$

$$b_1 = \frac{\alpha_3\rho_1(\alpha_2\rho_1^2(\alpha_3\rho_1+\alpha_2(\alpha_3+\rho_1))+\alpha_1^3(\rho_1-\rho_2)\rho_2}{\alpha_1^2\beta} + \tag{A.24b}$$
$$-\frac{\alpha_1\alpha_2\rho_1(\rho_1^2-2\alpha_3\rho_2-\rho_1\rho_2)-\alpha_1^2(\alpha_2\rho_1^2+(\alpha_3+\rho_1)(\rho_1-\rho_2)\rho_2)}{\alpha_1^2\beta},$$

$$b_0 = \frac{\alpha_3(\alpha_1-\rho_1)\rho_1^2(\alpha_1\rho_2-\alpha_2\rho_1)}{\alpha_1^3}. \tag{A.24c}$$

Every equilibrium stability is influenced by the positivity conditions of the other points. From the positivity of P_2^{B2}, we get that

$$\alpha_2 > \rho_2. \tag{A.25}$$

Two different scenarios arise from the positivity conditions of P_3^{B2}; either

$$\beta > 0, \tag{A.26a}$$

$$\alpha_1 > \rho_1, \tag{A.26b}$$

$$\alpha_2\rho_1 > \alpha_1\rho_2; \tag{A.26c}$$

or

$$\beta < 0, \tag{A.27a}$$
$$\alpha_1 < \rho_1, \tag{A.27b}$$
$$\alpha_2\rho_1 < \alpha_1\rho_2. \tag{A.27c}$$

Whenever Eq. (A.26) holds, equilibrium P_1^{B2} is stable (mainly $\alpha_1 < \rho_1$, as Eq.(A.25) is true whenever $P_2^{B2} > 0$). Moreover, equilibrium P_2^{B2} would also be stable whenever Eq. (A.27) holds and also

$$|\mathcal{R}(h^*)| < \alpha_3^2\rho_2 + \alpha_3\rho_2^2 \tag{A.28}$$

where h^* is defined as in Eq. (A.22). However, the stability of P_1^{B2} and P_2^{B2} is biologically uninteresting. Focusing on the non-trivial state, with the above constraints Eq.(A.26) or Eq.(A.27) and the Routh-Hurtwitz criterion, we have $b_0 < 0$. Therefore, P_3^{B2} is positive but always unstable.

A.3 Summary of stability conditions

We summarise in Table A.1 the conclusions of the mathematical analysis regarding stability of the non-trivial state.

Steady State	Model A1 $s_\rho = s_1, s_\alpha = 1$	Model A2 $s_\rho = s_2, s_\alpha = 1$	Model B1 $s_\rho = 1, s_\alpha = s_1$	Model B2 $s_\rho = 1, s_\alpha = s_2$
P_1^j	Unstable	Unstable	Stable	Conditionally stable
P_2^j	Unstable	Conditionally stable	Unstable	Conditionally stable
P_3^j	Conditionally Stable	Conditionally Stable	Unstable	Unstable

Table A.1: Steady-state stability for every model from Eq. (3.6) under conditions of positivity of the non-trivial steady state. Index j stands for the four models considering the different feedback regulations: $A1$ for cell proliferation regulation, all cell feedback; $A2$ for proliferation regulation, late cell feedback; $B1$ for transition rate regulation, late cell feedback; and $B2$ for transition rate regulation, all cell feedback.

APPENDIX B

Supplementary information for topological data analysis

Dataset 1	CD10 CD123 CD13 CD19 CD20 CD24 CD33 CD34 CD38 CD45 CD66c CD7 CD9 Kappa Lambda	Dataset 2	CD10 CD13 CD15 CD19 CD20 CD22 CD24 CD33 CD34 CD38 CD45 CD58 CD66c CD7 CD71 HLADR IGM cyCD3	Dataset 1&2	CD10 CD13 CD19 CD20 CD24 CD33 CD34 CD38 CD45 CD66c CD7

Table B.1: **IPT markers included in the study**. Sets of IPT markers in common for all patients in each dataset are presented in this table.

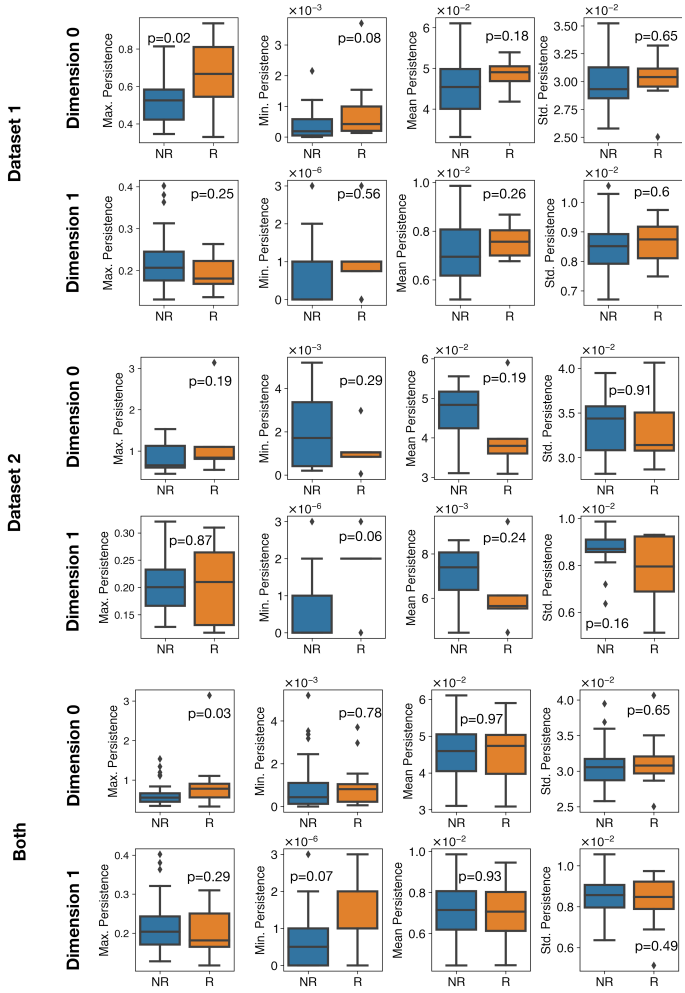


Figure B.1: **Topological feature analysis with 5 biomarkers for relapsed (orange) and non-relapsed (blue) patients**. Results are computed for markers CD10, CD20, CD38, CD45 and CD7. The features included are maximal persistence, minimal persistence, mean persistence, and the analysis is performed in dimensions 0 and 1.

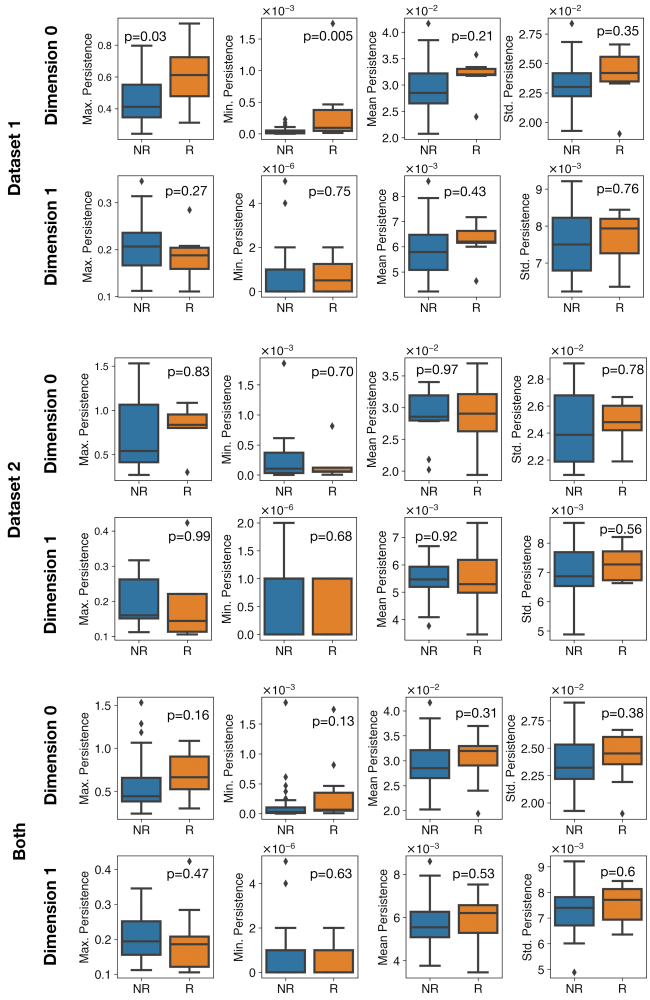


Figure B.2: **Topological feature analysis with 4 biomarkers for relapsed (orange) and non-relapsed (blue) patients**. Results are computed for markers CD10, CD20, CD38, and CD45. The features included are maximal persistence, minimal persistence, mean persistence, and the analysis is performed in dimensions 0 and 1.

Marker comb.	AUC	OOB	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
CD10-CD123	0.56	0.25	0.17	0.91	0.33	0.76	0.09	0.83	0.67	0.72
CD10-CD13	0.61	0.26	0.25	0.95	0.75	0.8	0.05	0.75	0.25	0.77
CD10-CD19	0.75	0.33	0.35	0.97	0.89	0.84	0.03	0.65	0.11	0.84
CD10-CD20	0.86	0.27	0.24	0.92	0.49	0.84	0.08	0.76	0.51	0.79
CD10-CD24	0.8	0.21	0.36	0.95	0.68	0.84	0.05	0.64	0.32	0.81
CD10-CD33	0.87	0.22	0.39	0.97	0.91	0.84	0.03	0.61	0.09	0.83
CD10-CD34	0.77	0.28	0.35	0.87	0.51	0.82	0.13	0.65	0.49	0.75
CD10-CD38	0.78	0.2	0.41	0.88	0.56	0.82	0.12	0.59	0.44	0.76
CD10-CD45	0.7	0.3	0.2	0.86	0.29	0.78	0.14	0.8	0.71	0.69
CD10-CD66c	0.9	0.2	0.51	0.88	0.58	0.86	0.12	0.49	0.42	0.77
CD10-CD7	0.94	0.17	0.52	0.98	0.91	0.87	0.02	0.48	0.09	0.87
CD10-CD9	0.81	0.18	0.51	0.91	0.63	0.87	0.09	0.49	0.37	0.81
CD10-KAPPA	0.93	0.19	0.53	0.95	0.84	0.86	0.05	0.47	0.16	0.84
CD10-LAMBDA	0.92	0.16	0.4	0.95	0.81	0.84	0.05	0.6	0.19	0.81
CD123-CD13	0.93	0.18	0.45	0.98	0.92	0.85	0.02	0.55	0.08	0.85
CD123-CD19	0.77	0.2	0.16	0.98	0.67	0.77	0.02	0.84	0.33	0.77
CD123-CD20	0.64	0.25	0.08	0.94	0.22	0.77	0.06	0.92	0.78	0.73
CD123-CD24	0.58	0.19	0.14	0.91	0.2	0.78	0.09	0.86	0.8	0.72
CD123-CD33	0.69	0.17	0.22	0.91	0.51	0.78	0.09	0.78	0.49	0.73
CD123-CD34	0.68	0.29	0.28	0.9	0.31	0.81	0.1	0.72	0.69	0.75
CD123-CD38	0.69	0.24	0.3	0.87	0.39	0.81	0.13	0.7	0.61	0.73
CD123-CD45	0.6	0.28	0.15	0.88	0.37	0.78	0.12	0.85	0.63	0.71
CD123-CD66c	0.58	0.37	0.37	0.84	0.33	0.82	0.16	0.63	0.67	0.72
CD123-CD7	0.6	0.28	0.16	0.92	0.61	0.78	0.08	0.84	0.39	0.73
CD123-CD9	0.6	0.33	0.11	0.85	0.09	0.76	0.15	0.89	0.91	0.67
CD123-KAPPA	0.56	0.26	0.19	0.91	0.45	0.79	0.09	0.81	0.55	0.74
CD123-LAMBDA	0.7	0.27	0.37	0.87	0.49	0.84	0.13	0.63	0.51	0.73
CD13-CD19	0.82	0.22	0.32	0.96	0.8	0.84	0.04	0.68	0.2	0.82
CD13-CD20	0.71	0.33	0.25	0.88	0.37	0.84	0.12	0.75	0.63	0.75
CD13-CD24	0.74	0.21	0.41	0.9	0.53	0.84	0.1	0.59	0.47	0.78
CD13-CD33	0.79	0.23	0.5	0.96	0.79	0.86	0.04	0.5	0.21	0.84
CD13-CD34	0.65	0.24	0.07	0.95	0.29	0.75	0.05	0.93	0.71	0.72
CD13-CD38	0.75	0.16	0.4	0.96	0.78	0.84	0.04	0.6	0.22	0.82
CD13-CD45	0.71	0.28	0.34	0.88	0.34	0.82	0.12	0.66	0.66	0.75
CD13-CD66c	0.76	0.16	0.36	0.92	0.67	0.81	0.08	0.64	0.33	0.76
CD13-CD7	0.67	0.29	0.2	0.92	0.36	0.8	0.08	0.8	0.64	0.75
CD13-CD9	0.52	0.34	0.26	0.88	0.3	0.81	0.12	0.74	0.7	0.73
CD13-KAPPA	0.71	0.21	0.33	0.9	0.61	0.81	0.1	0.67	0.39	0.75
CD13-LAMBDA	0.77	0.23	0.31	0.92	0.41	0.85	0.08	0.69	0.59	0.8
CD19-CD20	0.79	0.29	0.16	0.98	0.67	0.8	0.02	0.84	0.33	0.8
CD19-CD24	0.83	0.22	0.45	0.94	0.74	0.87	0.06	0.55	0.26	0.83
CD19-CD33	0.85	0.23	0.46	0.95	0.76	0.84	0.05	0.54	0.24	0.82
CD19-CD34	0.69	0.21	0.22	0.94	0.66	0.77	0.06	0.78	0.34	0.73
CD19-CD38	0.86	0.23	0.38	0.98	0.91	0.83	0.02	0.62	0.09	0.83
CD19-CD45	0.79	0.2	0.54	0.9	0.6	0.89	0.1	0.46	0.4	0.81
CD19-CD66c	0.87	0.16	0.58	0.94	0.74	0.89	0.06	0.42	0.26	0.86
CD19-CD7	0.68	0.21	0.38	0.9	0.49	0.83	0.1	0.62	0.51	0.77
CD19-CD9	0.82	0.24	0.51	0.9	0.52	0.88	0.1	0.49	0.48	0.81
CD19-KAPPA	0.81	0.24	0.31	0.95	0.64	0.81	0.05	0.69	0.36	0.79
CD19-LAMBDA	0.69	0.2	0.16	0.97	0.63	0.77	0.03	0.84	0.37	0.75
CD20-CD24	0.7	0.23	0.42	0.82	0.44	0.84	0.18	0.58	0.56	0.73
CD20-CD33	0.7	0.24	0.27	0.88	0.41	0.81	0.12	0.73	0.59	0.73
CD20-CD34	0.66	0.22	0.11	0.91	0.3	0.75	0.09	0.89	0.7	0.71

Table B.2: **Random Forest classification results for dataset 1 (1/3) for different marker combinations**. AUC: Area under the ROC curve. OOB: Out of bag error. TPR: True positive rate. TNR: True negative rate. PPV: Positive predictive value. NPV: Negative predictive value. FPR: False positive rate. FNR: False negative rate. FDR: False discovery rate. ACC: Accuracy (Coefficient of determination).

Marker comb.	AUC	OOB	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
CD20-CD38	0.68	0.32	0.43	0.89	0.56	0.84	0.11	0.57	0.44	0.79
CD20-CD45	0.75	0.21	0.31	0.94	0.57	0.81	0.06	0.69	0.43	0.77
CD20-CD66c	0.56	0.32	0.15	0.88	0.33	0.78	0.12	0.85	0.67	0.71
CD20-CD7	0.54	0.36	0.25	0.84	0.37	0.8	0.16	0.75	0.63	0.71
CD20-CD9	0.73	0.27	0.18	0.92	0.32	0.78	0.08	0.81	0.68	0.74
CD20-KAPPA	0.82	0.13	0.5	0.93	0.64	0.86	0.07	0.5	0.36	0.82
CD20-LAMBDA	0.62	0.3	0.19	0.87	0.37	0.77	0.13	0.81	0.63	0.7
CD24-CD33	0.78	0.15	0.58	0.84	0.53	0.89	0.16	0.42	0.47	0.78
CD24-CD34	0.84	0.18	0.56	0.9	0.7	0.87	0.1	0.44	0.3	0.8
CD24-CD38	0.88	0.16	0.63	0.94	0.83	0.9	0.06	0.37	0.17	0.86
CD24-CD45	0.72	0.21	0.36	0.87	0.65	0.83	0.13	0.64	0.35	0.74
CD24-CD66c	0.74	0.36	0.11	0.94	0.4	0.79	0.06	0.89	0.6	0.76
CD24-CD7	0.77	0.14	0.39	0.96	0.82	0.83	0.04	0.61	0.18	0.81
CD24-CD9	0.69	0.24	0.35	0.87	0.5	0.82	0.13	0.65	0.5	0.74
CD24-KAPPA	0.75	0.28	0.38	0.88	0.62	0.83	0.12	0.62	0.38	0.76
CD24-LAMBDA	0.87	0.18	0.66	0.96	0.74	0.91	0.04	0.34	0.26	0.89
CD33-CD34	0.75	0.23	0.33	0.87	0.52	0.82	0.13	0.67	0.48	0.73
CD33-CD38	0.96	0.2	0.46	0.97	0.8	0.86	0.03	0.54	0.2	0.85
CD33-CD45	0.82	0.21	0.48	0.87	0.55	0.86	0.13	0.52	0.45	0.78
CD33-CD66c	0.76	0.25	0.52	0.9	0.68	0.85	0.1	0.48	0.32	0.79
CD33-CD7	0.56	0.27	0.27	0.86	0.25	0.8	0.14	0.73	0.75	0.7
CD33-CD9	0.79	0.21	0.46	0.95	0.74	0.88	0.05	0.54	0.26	0.86
CD33-KAPPA	0.68	0.27	0.32	0.9	0.46	0.81	0.1	0.68	0.54	0.75
CD33-LAMBDA	0.74	0.28	0.48	0.8	0.47	0.84	0.2	0.52	0.53	0.72
CD34-CD38	0.82	0.23	0.32	0.95	0.74	0.83	0.05	0.68	0.26	0.81
CD34-CD45	0.78	0.27	0.53	0.85	0.52	0.88	0.15	0.46	0.48	0.78
CD34-CD66c	0.81	0.21	0.28	0.91	0.55	0.82	0.09	0.72	0.45	0.76
CD34-CD7	0.64	0.29	0.27	0.83	0.35	0.81	0.17	0.73	0.65	0.7
CD34-CD9	0.8	0.2	0.48	0.91	0.69	0.84	0.09	0.52	0.31	0.8
CD34-KAPPA	0.63	0.3	0.25	0.94	0.58	0.8	0.06	0.75	0.42	0.76
CD34-LAMBDA	0.84	0.24	0.31	0.99	0.9	0.82	0.01	0.69	0.1	0.82
CD38-CD45	0.8	0.22	0.47	0.89	0.57	0.87	0.11	0.52	0.43	0.8
CD38-CD66c	0.71	0.29	0.26	0.96	0.78	0.82	0.04	0.74	0.22	0.8
CD38-CD7	0.81	0.17	0.46	0.95	0.81	0.84	0.05	0.54	0.19	0.83
CD38-CD9	0.75	0.16	0.37	0.92	0.57	0.84	0.08	0.63	0.43	0.79
CD38-KAPPA	0.81	0.23	0.48	0.94	0.76	0.86	0.06	0.52	0.24	0.82
CD38-LAMBDA	0.87	0.23	0.59	0.92	0.69	0.89	0.08	0.41	0.31	0.84
CD45-CD66c	0.61	0.28	0.18	0.88	0.48	0.77	0.12	0.82	0.52	0.71
CD45-CD7	0.65	0.35	0.29	0.88	0.51	0.82	0.12	0.71	0.49	0.74
CD45-CD9	0.81	0.26	0.33	0.93	0.71	0.82	0.07	0.67	0.29	0.78
CD45-KAPPA	0.63	0.3	0.16	0.88	0.26	0.8	0.12	0.84	0.74	0.73
CD45-LAMBDA	0.76	0.14	0.18	0.99	0.8	0.75	0.01	0.82	0.2	0.75
CD66c-CD7	0.66	0.24	0.16	0.96	0.56	0.78	0.04	0.84	0.44	0.76
CD66c-CD9	0.58	0.3	0.13	0.86	0.17	0.78	0.14	0.87	0.83	0.7
CD66c-KAPPA	0.73	0.22	0.33	0.94	0.72	0.8	0.06	0.67	0.28	0.77
CD66c-LAMBDA	0.65	0.3	0.14	0.93	0.37	0.8	0.07	0.86	0.63	0.75
CD7-CD9	0.65	0.26	0.3	0.94	0.68	0.82	0.06	0.7	0.32	0.79
CD7-KAPPA	0.73	0.2	0.28	0.9	0.33	0.79	0.1	0.72	0.67	0.73
CD7-LAMBDA	0.69	0.23	0.29	0.94	0.64	0.82	0.06	0.71	0.36	0.79
CD9-KAPPA	0.81	0.19	0.29	0.98	0.88	0.8	0.02	0.71	0.12	0.79
CD9-LAMBDA	0.68	0.19	0.23	0.86	0.29	0.76	0.14	0.77	0.71	0.67
KAPPA-LAMBDA	0.63	0.32	0.36	0.88	0.4	0.83	0.12	0.64	0.6	0.76

Table B.3: **Random Forest classification results for dataset 1 (2/3) for different marker combinations**. AUC: Area under the ROC curve. OOB: Out of bag error. TPR: True positive rate. TNR: True negative rate. PPV: Positive predictive value. NPV: Negative predictive value. FPR: False positive rate. FNR: False negative rate. FDR: False discovery rate. ACC: Accuracy (Coefficient of determination).

Marker comb.	AUC	OOB	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
CD10-CD13	0.5	0.42	0.2	0.74	0.29	0.69	0.26	0.8	0.71	0.59
CD10-CD15	0.59	0.39	0.15	0.86	0.32	0.72	0.14	0.85	0.68	0.66
CD10-CD19	0.3	0.4	0.0	0.76	0.0	0.64	0.24	1.0	1.0	0.54
CD10-CD20	0.52	0.4	0.25	0.76	0.37	0.71	0.24	0.75	0.63	0.61
CD10-CD22	0.46	0.44	0.1	0.78	0.22	0.67	0.22	0.9	0.78	0.59
CD10-CD24	0.54	0.35	0.15	0.78	0.13	0.7	0.22	0.85	0.87	0.6
CD10-CD33	0.34	0.51	0.05	0.7	0.04	0.64	0.3	0.95	0.96	0.51
CD10-CD34	0.23	0.49	0.05	0.84	0.07	0.69	0.16	0.95	0.93	0.61
CD10-CD38	0.68	0.38	0.3	0.84	0.44	0.76	0.16	0.7	0.56	0.69
CD10-CD45	0.38	0.41	0.1	0.82	0.21	0.68	0.18	0.9	0.79	0.61
CD10-CD58	0.7	0.47	0.15	0.86	0.26	0.72	0.14	0.85	0.74	0.66
CD10-CD66c	0.43	0.35	0.3	0.78	0.37	0.75	0.22	0.7	0.63	0.64
CD10-CD7	0.54	0.38	0.25	0.82	0.43	0.73	0.18	0.75	0.57	0.66
CD10-CD71	0.25	0.53	0.0	0.84	0.0	0.67	0.16	1.0	1.0	0.6
CD10-HLADR	0.45	0.37	0.15	0.84	0.21	0.74	0.16	0.85	0.79	0.64
CD10-IGM	0.29	0.56	0.05	0.74	0.05	0.66	0.26	0.95	0.95	0.54
CD10-cyCD3	0.4	0.37	0.1	0.76	0.18	0.67	0.24	0.9	0.82	0.57
CD13-CD15	0.64	0.34	0.45	0.9	0.76	0.8	0.1	0.55	0.24	0.77
CD13-CD19	0.13	0.53	0.0	0.78	0.0	0.64	0.22	1.0	1.0	0.56
CD13-CD20	0.48	0.44	0.15	0.7	0.12	0.67	0.3	0.85	0.88	0.54
CD13-CD22	0.62	0.38	0.2	0.8	0.25	0.72	0.2	0.8	0.75	0.63
CD13-CD24	0.29	0.48	0.0	0.92	0.0	0.69	0.08	1.0	1.0	0.66
CD13-CD33	0.41	0.4	0.1	0.66	0.04	0.64	0.34	0.9	0.96	0.5
CD13-CD34	0.46	0.46	0.05	0.8	0.05	0.66	0.2	0.95	0.95	0.59
CD13-CD38	0.74	0.41	0.3	0.82	0.44	0.74	0.18	0.7	0.56	0.67
CD13-CD45	0.51	0.4	0.2	0.78	0.2	0.71	0.22	0.8	0.8	0.61
CD13-CD58	0.74	0.42	0.6	0.74	0.52	0.84	0.26	0.4	0.48	0.7
CD13-CD66c	0.76	0.36	0.4	0.86	0.69	0.78	0.14	0.6	0.31	0.73
CD13-CD7	0.77	0.4	0.3	0.88	0.5	0.78	0.12	0.7	0.5	0.71
CD13-CD71	0.72	0.36	0.5	0.8	0.51	0.83	0.2	0.5	0.49	0.71
CD13-HLADR	0.6	0.41	0.45	0.84	0.67	0.81	0.16	0.55	0.33	0.73
CD13-IGM	0.68	0.32	0.35	0.88	0.62	0.77	0.12	0.65	0.38	0.73
CD13-cyCD3	0.61	0.39	0.3	0.88	0.56	0.76	0.12	0.7	0.44	0.71
CD15-CD19	0.74	0.43	0.45	0.86	0.63	0.8	0.14	0.55	0.37	0.74
CD15-CD20	0.79	0.42	0.45	0.92	0.79	0.82	0.08	0.55	0.21	0.79
CD15-CD22	0.47	0.43	0.05	0.9	0.12	0.7	0.1	0.95	0.88	0.66
CD15-CD24	0.47	0.45	0.2	0.86	0.36	0.73	0.14	0.8	0.64	0.67
CD15-CD33	0.57	0.37	0.2	0.76	0.17	0.71	0.24	0.8	0.83	0.6
CD15-CD34	0.9	0.22	0.55	0.92	0.83	0.86	0.08	0.45	0.17	0.81
CD15-CD38	0.43	0.4	0.15	0.76	0.16	0.69	0.24	0.85	0.84	0.59
CD15-CD45	0.66	0.43	0.35	0.8	0.45	0.79	0.2	0.65	0.55	0.67
CD15-CD58	0.62	0.38	0.55	0.7	0.4	0.83	0.3	0.45	0.6	0.66
CD15-CD66c	0.68	0.46	0.35	0.84	0.52	0.77	0.16	0.65	0.48	0.7
CD15-CD7	0.78	0.36	0.25	0.86	0.5	0.75	0.14	0.75	0.5	0.69
CD15-CD71	0.4	0.46	0.15	0.74	0.12	0.68	0.26	0.85	0.88	0.57
CD15-HLADR	0.49	0.41	0.35	0.64	0.23	0.75	0.36	0.65	0.77	0.56
CD15-IGM	0.65	0.36	0.45	0.74	0.36	0.79	0.26	0.55	0.64	0.66
CD15-cyCD3	0.5	0.37	0.25	0.84	0.45	0.74	0.16	0.75	0.55	0.67
CD19-CD20	0.46	0.39	0.15	0.86	0.25	0.72	0.14	0.85	0.75	0.66
CD19-CD22	0.3	0.41	0.0	0.9	0.0	0.69	0.1	1.0	1.0	0.64
CD19-CD24	0.36	0.52	0.2	0.76	0.19	0.71	0.24	0.8	0.81	0.6

Table B.4: **Random Forest classification results for dataset 1 (3/3) for different marker combinations**. AUC: Area under the ROC curve. OOB: Out of bag error. TPR: True positive rate. TNR: True negative rate. PPV: Positive predictive value. NPV: Negative predictive value. FPR: False positive rate. FNR: False negative rate. FDR: False discovery rate. ACC: Accuracy (Coefficient of determination).

Marker comb.	AUC	OOB	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
CD19-CD33	0.56	0.51	0.05	0.84	0.06	0.69	0.16	0.95	0.94	0.61
CD19-CD34	0.37	0.5	0.15	0.78	0.26	0.69	0.22	0.85	0.74	0.6
CD19-CD38	0.46	0.49	0.4	0.6	0.27	0.71	0.4	0.6	0.73	0.54
CD19-CD45	0.48	0.5	0.2	0.72	0.26	0.67	0.28	0.8	0.74	0.57
CD19-CD58	0.8	0.27	0.65	0.82	0.58	0.88	0.18	0.35	0.42	0.77
CD19-CD66c	0.37	0.52	0.15	0.7	0.12	0.66	0.3	0.85	0.88	0.54
CD19-CD7	0.82	0.37	0.4	0.86	0.58	0.82	0.14	0.6	0.42	0.73
CD19-CD71	0.43	0.53	0.15	0.76	0.14	0.69	0.24	0.85	0.86	0.59
CD19-HLADR	0.49	0.48	0.3	0.76	0.35	0.72	0.24	0.7	0.65	0.63
CD19-IGM	0.58	0.49	0.05	0.78	0.11	0.66	0.22	0.95	0.89	0.57
CD19-cyCD3	0.38	0.49	0.05	0.88	0.2	0.69	0.12	0.95	0.8	0.64
CD20-CD22	0.44	0.44	0.1	0.86	0.14	0.71	0.14	0.9	0.86	0.64
CD20-CD24	0.12	0.51	0.0	0.8	0.0	0.66	0.2	1.0	1.0	0.57
CD20-CD33	0.6	0.34	0.1	0.78	0.1	0.68	0.22	0.9	0.9	0.59
CD20-CD34	0.52	0.36	0.1	0.94	0.42	0.73	0.06	0.9	0.58	0.7
CD20-CD38	0.46	0.46	0.05	0.86	0.06	0.69	0.14	0.95	0.94	0.63
CD20-CD45	0.4	0.54	0.25	0.74	0.39	0.69	0.26	0.75	0.61	0.6
CD20-CD58	0.37	0.38	0.1	0.72	0.17	0.66	0.28	0.9	0.83	0.54
CD20-CD66c	0.36	0.52	0.15	0.56	0.07	0.61	0.44	0.85	0.93	0.44
CD20-CD7	0.59	0.36	0.4	0.86	0.6	0.78	0.14	0.6	0.4	0.73
CD20-CD71	0.47	0.43	0.0	0.86	0.0	0.68	0.14	1.0	1.0	0.61
CD20-HLADR	0.56	0.36	0.35	0.78	0.39	0.79	0.22	0.65	0.61	0.66
CD20-IGM	0.61	0.46	0.25	0.82	0.47	0.73	0.18	0.75	0.53	0.66
CD20-cyCD3	0.31	0.46	0.15	0.76	0.14	0.69	0.24	0.85	0.86	0.59
CD22-CD24	0.45	0.45	0.15	0.76	0.17	0.69	0.24	0.85	0.83	0.59
CD22-CD33	0.24	0.48	0.05	0.64	0.02	0.61	0.36	0.95	0.98	0.47
CD22-CD34	0.53	0.5	0.15	0.86	0.31	0.72	0.14	0.85	0.69	0.66
CD22-CD38	0.36	0.51	0.1	0.66	0.08	0.64	0.34	0.9	0.92	0.5
CD22-CD45	0.32	0.51	0.25	0.66	0.17	0.7	0.34	0.75	0.83	0.54
CD22-CD58	0.24	0.47	0.1	0.66	0.06	0.59	0.34	0.9	0.94	0.5
CD22-CD66c	0.74	0.39	0.5	0.82	0.5	0.83	0.18	0.5	0.5	0.73
CD22-CD7	0.48	0.45	0.4	0.82	0.48	0.8	0.18	0.6	0.52	0.7
CD22-CD71	0.84	0.28	0.4	0.88	0.54	0.8	0.12	0.6	0.46	0.74
CD22-HLADR	0.46	0.4	0.05	0.82	0.04	0.68	0.18	0.95	0.96	0.6
CD22-IGM	0.46	0.44	0.15	0.72	0.2	0.65	0.28	0.85	0.8	0.56
CD22-cyCD3	0.36	0.47	0.0	0.78	0.0	0.65	0.22	1.0	1.0	0.56
CD24-CD33	0.22	0.41	0.05	0.9	0.1	0.7	0.1	0.95	0.9	0.66
CD24-CD34	0.85	0.36	0.5	0.88	0.65	0.84	0.12	0.5	0.35	0.77
CD24-CD38	0.45	0.48	0.15	0.6	0.09	0.64	0.4	0.85	0.91	0.47
CD24-CD45	0.48	0.49	0.25	0.78	0.26	0.73	0.22	0.75	0.74	0.63
CD24-CD58	0.29	0.49	0.1	0.78	0.1	0.68	0.22	0.9	0.9	0.59
CD24-CD66c	0.71	0.43	0.6	0.74	0.49	0.85	0.26	0.4	0.51	0.7
CD24-CD7	0.46	0.42	0.35	0.74	0.48	0.73	0.26	0.65	0.52	0.63
CD24-CD71	0.56	0.45	0.25	0.8	0.22	0.74	0.2	0.75	0.78	0.64
CD24-HLADR	0.32	0.56	0.1	0.74	0.09	0.67	0.26	0.9	0.91	0.56
CD24-IGM	0.68	0.38	0.65	0.66	0.42	0.86	0.34	0.35	0.58	0.66
CD24-cyCD3	0.6	0.35	0.3	0.88	0.69	0.75	0.12	0.7	0.31	0.71
CD33-CD34	0.75	0.28	0.6	0.78	0.52	0.86	0.22	0.4	0.48	0.73
CD33-CD38	0.58	0.45	0.05	0.84	0.14	0.68	0.16	0.95	0.86	0.61
CD33-CD45	0.56	0.48	0.25	0.78	0.24	0.74	0.22	0.75	0.76	0.63
CD33-CD58	0.55	0.44	0.3	0.62	0.16	0.71	0.38	0.7	0.84	0.53
CD33-CD66c	0.41	0.54	0.2	0.7	0.15	0.7	0.3	0.8	0.85	0.56

Table B.5: **Random Forest classification results for dataset 2 (1/2) different marker combinations**. AUC: Area under the ROC curve. OOB: Out of bag error. TPR: True positive rate. TNR: True negative rate. PPV: Positive predictive value. NPV: Negative predictive value. FPR: False positive rate. FNR: False negative rate. FDR: False discovery rate. ACC: Accuracy (Coefficient of determination).

Marker comb.	AUC	OOB	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
CD33-CD7	0.78	0.46	0.65	0.7	0.53	0.86	0.3	0.35	0.47	0.69
CD33-CD71	0.65	0.51	0.15	0.94	0.42	0.75	0.06	0.85	0.58	0.71
CD33-HLADR	0.48	0.55	0.2	0.76	0.2	0.7	0.24	0.8	0.8	0.6
CD33-IGM	0.52	0.54	0.4	0.66	0.25	0.74	0.34	0.6	0.76	0.59
CD33-cyCD3	0.43	0.39	0.1	0.84	0.12	0.7	0.16	0.9	0.88	0.63
CD34-CD38	0.33	0.47	0.1	0.72	0.1	0.66	0.28	0.9	0.9	0.54
CD34-CD45	0.37	0.34	0.2	0.74	0.19	0.7	0.26	0.8	0.81	0.59
CD34-CD58	0.39	0.42	0.1	0.86	0.19	0.7	0.14	0.9	0.81	0.64
CD34-CD66c	0.51	0.53	0.1	0.82	0.14	0.69	0.18	0.9	0.86	0.61
CD34-CD7	0.09	0.44	0.05	0.78	0.03	0.61	0.22	0.95	0.97	0.57
CD34-CD71	0.46	0.48	0.15	0.92	0.42	0.73	0.08	0.85	0.58	0.7
CD34-HLADR	0.7	0.4	0.2	0.92	0.58	0.74	0.08	0.8	0.42	0.71
CD34-IGM	0.58	0.35	0.05	0.84	0.2	0.67	0.16	0.95	0.8	0.61
CD34-cyCD3	0.52	0.41	0.1	0.86	0.17	0.71	0.14	0.9	0.83	0.64
CD38-CD45	0.51	0.5	0.15	0.74	0.14	0.68	0.26	0.85	0.86	0.57
CD38-CD58	0.51	0.53	0.25	0.72	0.18	0.72	0.28	0.75	0.82	0.59
CD38-CD66c	0.5	0.46	0.1	0.8	0.22	0.68	0.2	0.9	0.78	0.6
CD38-CD7	0.81	0.37	0.7	0.82	0.63	0.89	0.18	0.3	0.37	0.79
CD38-CD71	0.38	0.55	0.2	0.76	0.23	0.7	0.24	0.8	0.77	0.6
CD38-HLADR	0.51	0.52	0.2	0.7	0.22	0.68	0.3	0.8	0.78	0.56
CD38-IGM	0.6	0.47	0.35	0.8	0.54	0.75	0.2	0.65	0.46	0.67
CD38-cyCD3	0.52	0.46	0.15	0.82	0.17	0.71	0.18	0.85	0.83	0.63
CD45-CD58	0.98	0.19	0.65	0.96	0.92	0.88	0.04	0.35	0.08	0.87
CD45-CD66c	0.53	0.27	0.05	0.9	0.08	0.7	0.1	0.95	0.92	0.66
CD45-CD7	0.68	0.3	0.55	0.66	0.32	0.82	0.34	0.45	0.68	0.63
CD45-CD71	0.37	0.46	0.05	0.7	0.04	0.62	0.3	0.95	0.96	0.51
CD45-HLADR	0.3	0.48	0.0	0.78	0.0	0.65	0.22	1.0	1.0	0.56
CD45-IGM	0.41	0.43	0.25	0.88	0.57	0.74	0.12	0.75	0.43	0.7
CD45-cyCD3	0.74	0.3	0.55	0.76	0.5	0.85	0.24	0.45	0.5	0.7
CD58-CD66c	0.72	0.41	0.45	0.8	0.45	0.82	0.2	0.55	0.55	0.7
CD58-CD7	0.9	0.39	0.2	0.94	0.71	0.75	0.06	0.8	0.29	0.73
CD58-CD71	0.5	0.47	0.15	0.66	0.11	0.64	0.34	0.85	0.89	0.51
CD58-HLADR	0.72	0.4	0.35	0.9	0.58	0.79	0.1	0.65	0.42	0.74
CD58-IGM	0.76	0.4	0.3	0.84	0.46	0.75	0.16	0.7	0.54	0.69
CD58-cyCD3	0.69	0.39	0.1	0.86	0.25	0.7	0.14	0.9	0.75	0.64
CD66c-CD7	0.34	0.45	0.05	0.76	0.04	0.66	0.24	0.95	0.96	0.56
CD66c-CD71	0.66	0.35	0.4	0.8	0.47	0.78	0.2	0.6	0.53	0.69
CD66c-HLADR	0.45	0.44	0.15	0.84	0.26	0.71	0.16	0.85	0.74	0.64
CD66c-IGM	0.53	0.43	0.4	0.8	0.51	0.8	0.2	0.6	0.49	0.69
CD66c-cyCD3	0.27	0.4	0.1	0.64	0.06	0.61	0.36	0.9	0.94	0.49
CD7-CD71	0.52	0.38	0.35	0.88	0.65	0.77	0.12	0.65	0.35	0.73
CD7-HLADR	0.44	0.52	0.2	0.76	0.15	0.72	0.24	0.8	0.85	0.6
CD7-IGM	0.43	0.42	0.25	0.64	0.25	0.62	0.36	0.75	0.75	0.53
CD7-cyCD3	0.66	0.39	0.3	0.8	0.36	0.77	0.2	0.7	0.64	0.66
CD71-HLADR	0.38	0.5	0.0	0.78	0.0	0.64	0.22	1.0	1.0	0.56
CD71-IGM	0.47	0.38	0.25	0.66	0.22	0.67	0.34	0.75	0.78	0.54
CD71-cyCD3	0.5	0.48	0.25	0.84	0.29	0.75	0.16	0.75	0.71	0.67
HLADR-IGM	0.44	0.5	0.1	0.8	0.11	0.69	0.2	0.9	0.89	0.6
HLADR-cyCD3	0.45	0.41	0.1	0.8	0.11	0.69	0.2	0.9	0.89	0.6
IGM-cyCD3	0.82	0.39	0.45	0.74	0.43	0.78	0.26	0.55	0.57	0.66

Table B.6: **Random Forest classification results for dataset 2 (1/2) for different marker combinations**. AUC: Area under the ROC curve. OOB: Out of bag error. TPR: True positive rate. TNR: True negative rate. PPV: Positive predictive value. NPV: Negative predictive value. FPR: False positive rate. FNR: False negative rate. FDR: False discovery rate. ACC: Accuracy (Coefficient of determination).

Marker comb.	AUC	OOB	TPR	TNR	PPV	NPV	FPR	FNR	FDR	ACC
CD10-CD13	0.35	0.34	0.06	0.84	0.07	0.75	0.16	0.94	0.93	0.66
CD10-CD19	0.41	0.35	0.08	0.85	0.12	0.76	0.15	0.92	0.88	0.67
CD10-CD20	0.6	0.33	0.2	0.86	0.31	0.79	0.14	0.8	0.69	0.71
CD10-CD24	0.47	0.34	0.1	0.87	0.16	0.77	0.13	0.9	0.84	0.7
CD10-CD33	0.59	0.33	0.18	0.88	0.32	0.79	0.12	0.82	0.68	0.72
CD10-CD34	0.53	0.37	0.12	0.83	0.12	0.76	0.17	0.88	0.88	0.67
CD10-CD38	0.68	0.25	0.46	0.85	0.53	0.84	0.15	0.54	0.47	0.76
CD10-CD45	0.43	0.36	0.04	0.84	0.11	0.75	0.16	0.96	0.89	0.65
CD10-CD66c	0.5	0.32	0.08	0.89	0.21	0.77	0.11	0.92	0.79	0.71
CD10-CD7	0.63	0.28	0.18	0.87	0.25	0.79	0.13	0.82	0.75	0.71
CD13-CD19	0.31	0.34	0.06	0.88	0.15	0.76	0.12	0.94	0.85	0.69
CD13-CD20	0.32	0.36	0.04	0.79	0.05	0.73	0.21	0.96	0.95	0.62
CD13-CD24	0.54	0.32	0.1	0.88	0.21	0.77	0.12	0.9	0.79	0.7
CD13-CD33	0.48	0.32	0.18	0.86	0.28	0.78	0.14	0.82	0.73	0.7
CD13-CD34	0.45	0.35	0.14	0.81	0.21	0.76	0.19	0.86	0.79	0.66
CD13-CD38	0.58	0.33	0.18	0.92	0.4	0.78	0.08	0.82	0.6	0.74
CD13-CD45	0.54	0.35	0.08	0.91	0.17	0.76	0.09	0.92	0.83	0.71
CD13-CD66c	0.38	0.36	0.08	0.82	0.12	0.74	0.18	0.92	0.88	0.65
CD13-CD7	0.52	0.34	0.12	0.85	0.22	0.76	0.15	0.88	0.78	0.68
CD19-CD20	0.56	0.35	0.12	0.87	0.17	0.76	0.13	0.88	0.83	0.69
CD19-CD24	0.51	0.31	0.14	0.81	0.14	0.75	0.19	0.86	0.86	0.65
CD19-CD33	0.65	0.34	0.18	0.88	0.28	0.77	0.12	0.82	0.72	0.71
CD19-CD34	0.44	0.34	0.08	0.88	0.16	0.75	0.12	0.92	0.84	0.69
CD19-CD38	0.64	0.35	0.16	0.88	0.32	0.77	0.12	0.84	0.68	0.7
CD19-CD45	0.5	0.36	0.22	0.8	0.24	0.77	0.2	0.78	0.76	0.66
CD19-CD66c	0.54	0.37	0.18	0.84	0.28	0.77	0.16	0.82	0.72	0.68
CD19-CD7	0.35	0.33	0.0	0.92	0.0	0.75	0.08	1.0	1.0	0.7
CD20-CD24	0.4	0.35	0.2	0.76	0.23	0.75	0.24	0.8	0.77	0.63
CD20-CD33	0.32	0.36	0.02	0.85	0.03	0.73	0.15	0.98	0.97	0.65
CD20-CD34	0.57	0.3	0.16	0.88	0.34	0.77	0.12	0.84	0.66	0.7
CD20-CD38	0.61	0.3	0.3	0.78	0.29	0.78	0.22	0.7	0.71	0.67
CD20-CD45	0.57	0.3	0.14	0.85	0.19	0.76	0.15	0.86	0.81	0.68
CD20-CD66c	0.43	0.33	0.1	0.82	0.1	0.74	0.18	0.9	0.9	0.65
CD20-CD7	0.7	0.3	0.36	0.86	0.42	0.81	0.14	0.64	0.58	0.74
CD24-CD33	0.52	0.31	0.14	0.86	0.19	0.76	0.14	0.86	0.81	0.69
CD24-CD34	0.5	0.35	0.14	0.85	0.22	0.76	0.15	0.86	0.78	0.68
CD24-CD38	0.62	0.32	0.14	0.92	0.35	0.77	0.08	0.86	0.65	0.73
CD24-CD45	0.44	0.34	0.08	0.91	0.21	0.76	0.09	0.92	0.79	0.71
CD24-CD66c	0.32	0.38	0.08	0.82	0.07	0.74	0.18	0.92	0.93	0.64
CD24-CD7	0.54	0.33	0.24	0.84	0.3	0.78	0.16	0.76	0.7	0.7
CD33-CD34	0.52	0.31	0.12	0.82	0.1	0.75	0.18	0.88	0.9	0.65
CD33-CD38	0.41	0.33	0.08	0.84	0.1	0.75	0.16	0.92	0.9	0.66
CD33-CD45	0.49	0.35	0.1	0.85	0.12	0.75	0.15	0.9	0.88	0.67
CD33-CD66c	0.41	0.33	0.1	0.81	0.15	0.74	0.19	0.9	0.85	0.64
CD33-CD7	0.52	0.35	0.14	0.86	0.21	0.76	0.14	0.86	0.79	0.69
CD34-CD38	0.33	0.35	0.0	0.88	0.0	0.74	0.12	1.0	1.0	0.67
CD34-CD45	0.43	0.37	0.12	0.82	0.12	0.75	0.18	0.88	0.88	0.66
CD34-CD66c	0.54	0.34	0.2	0.88	0.38	0.78	0.12	0.8	0.62	0.71
CD34-CD7	0.38	0.37	0.0	0.84	0.0	0.73	0.16	1.0	1.0	0.64
CD38-CD45	0.67	0.35	0.2	0.89	0.37	0.79	0.11	0.8	0.63	0.72
CD38-CD66c	0.48	0.33	0.1	0.87	0.25	0.75	0.13	0.9	0.75	0.69
CD38-CD7	0.54	0.27	0.22	0.87	0.24	0.79	0.13	0.78	0.76	0.71
CD45-CD66c	0.54	0.35	0.12	0.82	0.12	0.75	0.18	0.88	0.88	0.65
CD45-CD7	0.56	0.32	0.12	0.86	0.27	0.76	0.14	0.88	0.73	0.69
CD66c-CD7	0.39	0.38	0.12	0.84	0.18	0.75	0.16	0.88	0.82	0.67

Table B.7: **Random Forest classification results for both datasets combined, depending on the marker combination**. AUC: Area under the ROC curve. OOB: Out of bag error. TPR: True positive rate. TNR: True negative rate. PPV: Positive predictive value. NPV: Negative predictive value. FPR: False positive rate. FNR: False negative rate. FDR: False discovery rate. ACC: Accuracy (Coefficient of determination).

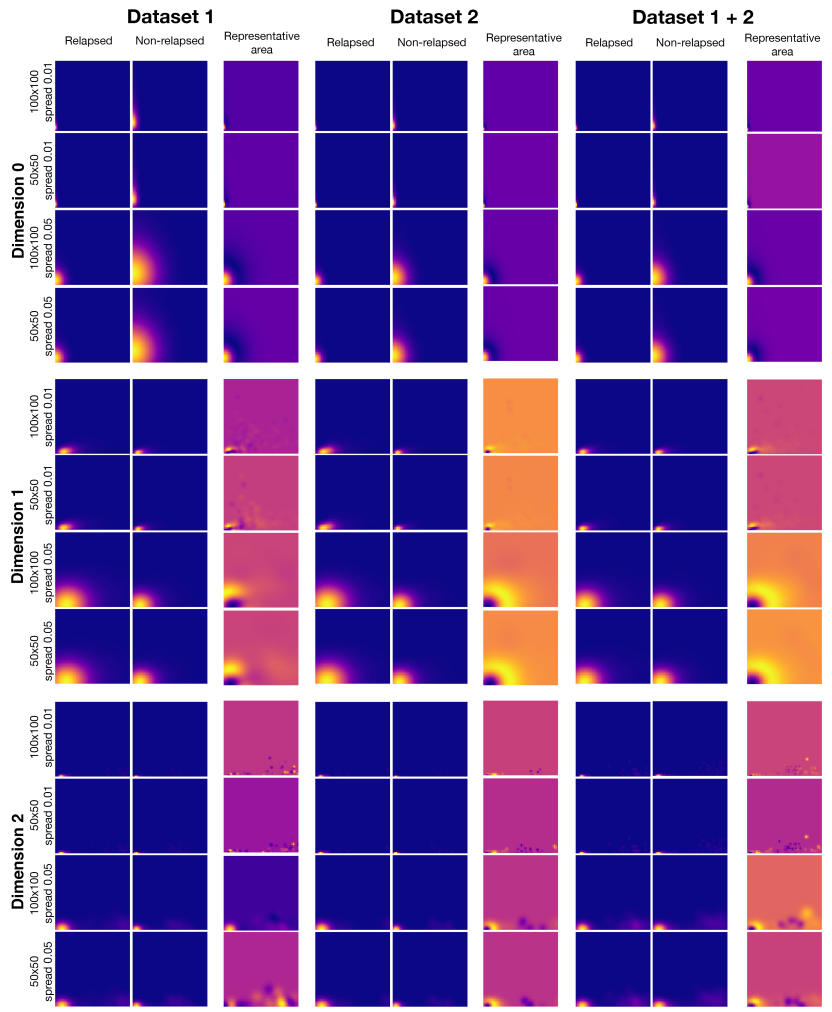


Figure B.3: **Mean persistence images for markers CD10, CD20, CD38 and CD45 for relapsed and non-relapsed patients.** The results are shown considering the dataset (1,2 or both) and dimension analysed (0,1 or 2), which also depends on the choose of persistence image grid (50×50 or 100×100) and spread of the Gaussian 2D distributions (0.01 or 0.05). The representative area between both cohorts are shown for each dataset.

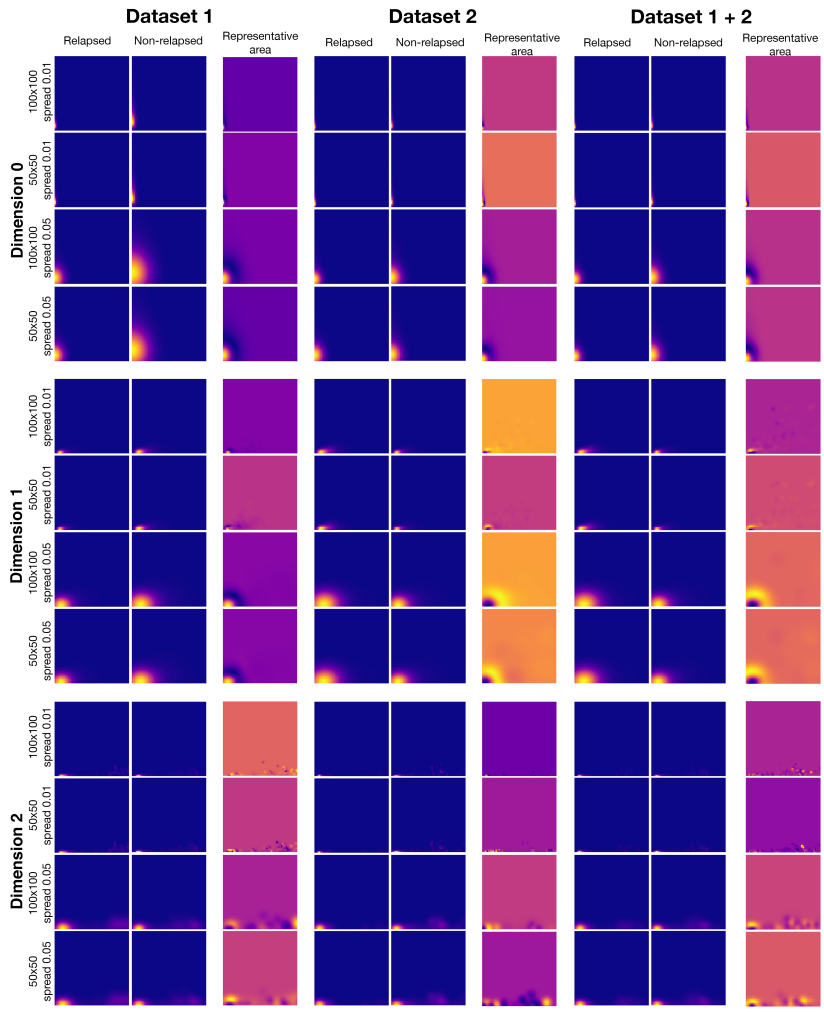


Figure B.4: **Mean persistence images for markers CD10, CD20, CD38, CD45 and CD7 for relapsed and non-relapsed patients.** The results are shown considering the dataset (1,2 or both) and dimension analysed (0,1 or 2), which also depends on the choose of persistence image grid (50 ×50 or 100 ×100) and spread of the Gaussian 2D distributions (0.01 or 0.05). The representative area between both cohorts are shown for each dataset.

				2D Gaussian Dist. spread = 0.01			
				100x100 grid		50x50 grid	
Markers	Dim.	Dataset	Method	Score	Accuracy	Score	Accuracy
10,20,38,45,7	0	1	LR	1	1	1	1
			SVM	1	1	1	1
		2	LR	0.88	0.45	0.88	0.45
			SVM	0.95	0.72	0.95	0.72
		1&2	LR	0.96	0.8	0.96	0.8
			SVM	0.98	0.9	0.98	0.9
	1	1	LR	0.97	0.84	0.97	0.84
			SVM	0.97	0.84	0.97	0.84
		2	LR	0.88	0.45	0.88	0.45
			SVM	0.88	0.45	0.88	0.45
		1&2	LR	0.96	0.8	0.96	0.8
			SVM	0.98 / 1	0.9 / 1	0.98	0.9
	2	1	LR	0.79	-0.28	0.79	-0.28
			SVM	0.79	-0.28	0.75 / 0.78	-0.44 / -0.28
		2	LR	0.78	-0.11	0.78	-0.11
			SVM	0.78	-0.11	0.78	-0.11
		1&2	LR	0.78	-0.21	0.84	0.09
			SVM	0.85	0.19	0.78	-0.21
10,20,38,45	0	1	LR	1	1	1	1
			SVM	1	1	1	1
		2	LR	1	1	1	1
			SVM	1	1	1	1
		1&2	LR	1	1	1	1
			SVM	1	1	1	1
	1	1	LR	0.89	0.36	0.92 / 0.94	0.52 / 0.68
			SVM	0.97	0.84	0.97	0.84
		2	LR	1	1	1	1
			SVM	1	1	1	1
		1&2	LR	0.98	0.9	0.98	0.9
			SVM	1	1	1	1
	2	1	LR	0.78	-0.29	0.75	-0.45
			SVM	0.78	-0.29	0.69	-0.77
		2	LR	0.77	-0.11	0.9	0.45
			SVM	0.9	0.45	0.9	0.45
		1&2	LR	0.78	-0.22	0.8 / 0.81	-0.11 / -0.01
			SVM	0.89 / 0.96	0.39 / 0.8	0.89 / 0.94	0.39 / 0.7

Table B.8: **Classification results of the Support Vector Machine (SVM) and Logistic Regression (LR) for a 2D Gaussian Distribution spread of 0.01.** These results depend on the IPT marker set analysed, dataset, dimension, spread of the 2D Gaussian distribution generated, and size of the grid of the persistence images used. Score is understood as the mean score after a 5-fold cross-validation and accuracy represents the coefficient of determination R^2. Leave-One-Out cross-validation results are also provided in case they were different to the 5-fold cross-validation ones.

				2D Gaussian Dist. spread = 0.05			
				100x100 grid		50x50 grid	
Markers	Dim.	Dataset	Method	Score	Accuracy	Score	Accuracy
10,20,38,45,7	0	1	LR	1	1	1	1
			SVM	1	1	1	1
		2	LR	0.88	0.45	0.88	0.45
			SVM	0.95	0.72	0.95	0.72
		1&2	LR	0.98	0.9	0.98	0.9
			SVM	0.98	0.9	0.98	0.9
	1	1	LR	1	1	1	1
			SVM	1	1	1	1
		2	LR	0.88	0.45	0.88	0.45
			SVM	0.88	0.72	0.94	0.72
		1&2	LR	0.98	0.9	0.98	0.9
			SVM	0.98	0.9	0.98	0.9
	2	1	LR	0.79	-0.28	0.79	-0.28
			SVM	0.89 / 0.94	0.36 / 0.68	0.86 / 0.94	0.20 / 0.68
		2	LR	0.72	-0.38	0.78	-0.11
			SVM	0.73	-0.38	0.73	-0.38
		1&2	LR	0.76	-0.31	0.78	-0.21
			SVM	0.89 / 0.9	0.4 / 0.5	0.87	0.29
10,20,38,45	0	1	LR	1	1	1	1
			SVM	1	1	1	1
		2	LR	1	1	1	1
			SVM	1	1	1	1
		1&2	LR	1	1	1	1
			SVM	1	1	1	1
	1	1	LR	0.78	-0.29	0.94	0.68
			SVM	1	1	1	1
		2	LR	0.93	0.72	0.93	0.72
			SVM	0.93	0.72	0.93	0.72
		1&2	LR	0.98	0.9	0.98	0.9
			SVM	1	1	1	1
	2	1	LR	0.78	-0.29	0.78	-0.29
			SVM	0.78	-0.29	0.78	-0.29
		2	LR	0.72	-0.38	0.72	-0.38
			SVM	0.83	0.17	0.9	0.45
		1&2	LR	0.7	-0.32	0.7	-0.32
			SVM	0.98	0.9	0.98	0.9

Table B.9: **Classification results of the Support Vector Machine (SVM) and Logistic Regression (LR) for a 2D Gaussian Distribution spread of 0.05.** These results depend on the IPT marker set analysed, dataset, dimension, spread of the 2D Gaussian distribution generated, and size of the grid of the persistence images used. Score is understood as the mean score after a 5-fold cross-validation and accuracy represents the coefficient of determination R^2. Leave-One-Out cross-validation results are also provided in case they were different to the 5-fold cross-validation ones.

APPENDIX C

Lie symmetry generators of the equations in study

C.1 Lie symmetry generators for Eq. (6.4)

In this Section we present the corresponding generators for Eq. (6.4). **Case 1**. For $c = c(x), f = f(u)$ and $g = g(u)$ arbitrary functions we get the generator

$$\mathbf{X}_1 = \partial_t. \tag{C.1a}$$

$$(x^*, t^*, u^*) = (x, t + \epsilon, u) \quad \textit{time translation.} \tag{C.1b}$$

Case 2. For $f = f(u), g = g(u)$ arbitrary functions and $c(x) = \frac{1}{4}\left(c_1\, x + c_2\right)^2$ we get the generator $\mathbf{X}_1$ and besides

$$\mathbf{X}_2 = (c_1 x + c_2)\partial_x. \tag{C.2a}$$

$$(x^*, t^*, u^*) = \left(e^{c_1\,\epsilon}\left(x + \frac{c_2}{c_1}\right) - \frac{c_2}{c_1}, t, u\right) \quad \textit{scaling and shift.} \tag{C.2b}$$

Case 3. For $f(u) = f_2(g_2 - u)^{-f_1}$, $g(u) = g_3(g_2 - u)^{g_1}$ and $c(x) = c_3(c_2 - x)^{c_1}$ with arbitrary values of the constants f_1, g_1, such that $f_1 + g_1 + 1 \neq 0$, equation (6.4) admits admits the generator $\mathbf{X}_1$ and the following:

$$\mathbf{X}_3 = (c_2 - x)\partial_x + \frac{(c_1 - 2)(f_1 + 1)t}{f_1 + g_1 + 1}\partial_t + \frac{(c_1 - 2)(u - g_2)}{f_1 + g_1 + 1}\partial_u. \tag{C.3a}$$

$$(x^*, t^*, u^*) = \left(e^{-\epsilon}\,(x - c_2) + c_2, \exp\left(\frac{(c_1 - 2)(f_1 + 1)}{f_1 + g_1 + 1}\epsilon\right)t, \right. \tag{C.3b}$$

$$g_2 + \exp\left(\frac{c_1 - 2}{f_1 + g_1 + 1}\epsilon\right)(u - g_2)\Bigg) \quad \textit{scaling and shift.}$$

3.1. If $f_1 + g_1 + 1 = 0$ then equation (6.4) admits $\mathbf{X}_1$ as a generator for $c(x)$ an arbitrary function.

Case 4. For $f(u) = f_1\,(u - g_2) + f_2\,(g_2 - u)^{g_1+1}$, $g(u) = g_3(g_2 - u)^{g_1}$ and $c = c(x)$ an arbitrary function with arbitrary values of the constants f_1, g_1, g_2 such that $g_1, f_1 \neq 0$, equation (6.4) admits the generator $\mathbf{X}_1$ and the following:

$$\mathbf{X}_4 = e^{-f_1\,g_1\,t}\partial_t + e^{-f_1\,g_1\,t} f_1\,(u - g_2)\,\partial_u. \tag{C.4a}$$

$$(x^*, t^*, u^*) = \Bigg(\frac{1}{f_1 g_1}\ln\left|e^{f_1\,g_1\,t} + f_1 g_1 \epsilon\right|, x, \tag{C.4b}$$

$$g_2 + \exp\left(\frac{f_1\,\epsilon}{f_1\,g_1\,\epsilon + e^{f_1\,g_1\,t}}\right)(u - g_2)\Bigg) \quad \textit{time dilation and shift.}$$

4.1. If $f_1 = 0$, with $f(u) = f_2\,(g_2 - u)^{g_1+1}$, $g(u) = g_3(g_2 - u)^{g_1}$ and arbitrary $c(x)$, then equation (6.4) admits the generator $\mathbf{X}_1$ and as $g_1 \neq 0$,

$$\mathbf{X}_{4a} = t\,\partial_t + \frac{(u - g_2)}{g_1}\partial_u \tag{C.5a}$$

$$(x^*, t^*, u^*) = \left(e^{\epsilon}\,t, x, \exp\left(\frac{\epsilon}{g_1}\right)(u - g_2) + g_2\right) \quad \textit{scaling and shift.} \tag{C.5b}$$

Case 5. For $f(u) = f_2\,(g_2 - u)^{-f_1}$, $g(u) = g_3(g_2 - u)^{g_1}$, with arbitrary values of the constants f_1, f_2, g_1, g_2, g_3, we consider the following subcases:

5.1. For $c(x) = \left(\frac{(2\,g_1+3)(c_1\,x+c_2)}{3\,g_1+4}\right)^{\frac{3\,g_1+4}{2\,g_1+3}}$ with arbitrary values of the constants c_1, c_2 such that $f_1 + g_1 + 1 \neq 0$, $g_1 \neq 0, -\frac{4}{3}, -\frac{3}{2}$, equation (6.4) admits the following generator $\mathbf{X}_1$, and besides:

$$\begin{aligned}\mathbf{X}_5 =& (c_1\,x + c_2)\partial_x + \frac{(f_1+1)\,(g_1+2)c_1\,t}{(2\,g_1+3)\,(g_1+f_1+1)}\partial_t + \\ &- \frac{(g_2-u)\,(g_1+2)c_1}{(2\,g_1+3)\,(g_1+f_1+1)}\partial_u.\end{aligned} \tag{C.6a}$$

$$(x^*, t^*, u^*) = \left(e^{c_1\,\epsilon} \left(x + \frac{c_2}{c_1} \right) - \frac{c_2}{c_1}, \exp\left(\frac{c_1(g_1+2)(f_1+1)}{(2g_1+3)(g_1+f_1+1)}\epsilon \right) t, \right. \tag{C.6b}$$

$$\left. g_2 + \exp\left(\frac{c_1(g_1+2)}{(2g_1+3)(g_1+f_1+1)}\epsilon \right)(u-g_2) \right) \quad \textit{scaling and shift.}$$

5.2. We consider $c\,(x) = c_2 \exp(c_1 x)$ for arbitrary values of the constants c_1, c_2. If $g_1 = -\frac{3}{2}$, and $f_1 \neq \frac{1}{2}$ then equation (6.4) admits the generators $\mathbf{X}_1$ and

$$\mathbf{X}_{5b} = \frac{-2\,t\,(f_1+1)\,c_1}{2\,f_1-1}\partial_t + \partial_x + 2\,\frac{(g_2-u)\,c_1}{2\,f_1-1}\partial_u. \tag{C.7a}$$

$$(x^*, t^*, u^*) = \left(t\,\exp\left(\frac{-2\,(f_1+1)\,c_1\epsilon}{2\,f_1-1} \right), x+\epsilon, \right. \tag{C.7b}$$

$$\left. \exp\left(-\frac{2\,c_1\epsilon}{2\,f_1-1} \right)(u-g_2) + g_2 \right) \quad \textit{scaling and shift.}$$

5.3. We consider $c\,(x) = c_2 \exp(c_1 x)$ for arbitrary values of the constants c_1, c_2. If $g_1 = -\frac{3}{2}$, and $f_1 = \frac{1}{2}$ then equation (6.4) admits the generators $\mathbf{X}_1$ and

$$\mathbf{X}_{5c} = \frac{c_1\,t}{2}\partial_t + \partial_x + c_1\,(u-g_2)\partial_u. \tag{C.8a}$$

$$(x^*, t^*, u^*) = \left(t\,e^{\frac{c_1\,\epsilon}{2}}, x+\epsilon, e^{c_1\,\epsilon}(u-g_2)+g_2 \right) \quad \textit{scaling and shift.} \tag{C.8b}$$

5.4. If $g_1 = -\frac{4}{3}$ or $f_1 + g_1 + 1 = 0$, then equation (6.4) admits the generator $\mathbf{X}_1$ for $c(x)$ an arbitrary function.

Case 6. For $f(u) = f_2\,(g_1\,u + g_2)^{\frac{f_1}{g_1}}$, $g(u) = \left(-\frac{4}{3}\,(g_1\,u+g_2)^{-1} \right)^{\frac{4}{3}}$ and $c(x) = c_3\,(c_2 - x)^{c_1}$ with arbitrary values of the constants f_1, f_2, g_1, g_2, c_1, c_2, c_3 such that $3f_1 + g_1 \neq 0$, $g_1 \neq 0$, $c_1 \neq 0$, equation (6.4) admits the generator

$\mathbf{X}_1$ and the following:

$$\mathbf{X}_6 = (c_2 - x)\partial_x + \frac{3\,(c_1-2)\,(f_1-g_1)\,t}{3\,f_1+g_1}\partial_t - \frac{3\,(u+g_2)\,(c_1-2)\,g_1}{3\,f_1+g_1}\partial_u. \tag{C.9a}$$

$$(x^*,t^*,u^*) = \left(e^{-\epsilon}\,(x-c_2)+c_2, \exp\left(\frac{3\,(c_1-2)\,(f_1-g_1)}{3\,f_1+g_1}\epsilon\right)t, \exp\left(\frac{-3\,(c_1-2)\,g_1}{3\,f_1+g_1}\epsilon\right)(u+g_2)-g_2\right) \quad \textit{scaling and shift.} \tag{C.9b}$$

6.1. If $3f_1 + g_1 = 0$ then equation (6.4) admits the generators $\mathbf{X}_1$ and

$$\mathbf{X}_{6*} = t\,\partial_t + 3\frac{(u-g_2)}{4}\partial_u \tag{C.10a}$$

$$(x^*,t^*,u^*) = \left(e^{\epsilon}\,t, x, \exp\left(\frac{3\epsilon}{4}\right)(u-g_2)+g_2\right) \quad \textit{scaling and shift.} \tag{C.10b}$$

Case 7. For $c = c(x)$ an arbitrary function, $f(u) = f_1\left(u+\frac{g_2}{g_1}\right) + f_2\left(u+\frac{g_2}{g_1}\right)^{-\frac{1}{3}}$ and $g(u) = \left(-\frac{4}{3}\,(g_1\,u+g_2)^{-1}\right)^{\frac{4}{3}}$, with arbitrary values of the constants f_1, f_2, g_1, g_2 such that $g_1 \neq 0$, equation (6.4) admits the generator $\mathbf{X}_1$ and the following:

$$\mathbf{X}_7 = e^{-\frac{4}{3}f_1\,t}\partial_t - \frac{f_1\,(g_1\,u+g_2)\,e^{-\frac{4}{3}f_1\,t}}{g_1}\partial_u. \tag{C.11a}$$

$$(x^*,t^*,u^*) = \left(\frac{-3}{4\,f_1}\ln\left|e^{-\frac{4}{3}f_1\,t} - \frac{4}{3}f_1\epsilon\right|, x, \exp\left(f_1\left(\frac{4}{3}f_1\epsilon - e^{-\frac{4}{3}f_1\,t}\right)\epsilon\right)\left(u+\frac{g_2}{g_1}\right) - \frac{g_2}{g_1}\right) \tag{C.11b}$$

$$\textit{exponential dilation and shift.} \tag{C.11c}$$

Case 8. For $f(u) = f_1\,(u-g_2) + f_2(u-g_2)^{g_1+1}$ and $g(u) = g_3(g_2-u)^{g_1}$, we consider arbitrary values of the constants f_1, f_2, g_1, g_2, and g_3. We distinguish the following subcases:

8.1. We consider $c(x) = \left(\frac{(2\,g_1+3)(c_1\,x+c_2)}{3\,g_1+4}\right)^{\frac{3\,g_1+4}{2\,g_1+3}}$ for arbitrary values of the constants c_1, c_2. If $f_2 = 0$ and $g_1 \neq -2, -\frac{4}{3}, -\frac{3}{2}, -1$, then equation (6.4) admits the generators $\mathbf{X}_1$, $\mathbf{X}_4$ and also

$$\mathbf{X}_{8a} = \frac{g_1\ (2\,g_1+3)\,(c_1\,x+c_2)}{c_1\ (g_1+2)}\partial_x + (u-g_2)\partial_u, \tag{C.12a}$$

$$(x^*, t^*, u^*) = \left(\left(x+\frac{c_2}{c_1}\right)\exp\left(\frac{g_1(2\,g_1+3)}{g_1+2}\epsilon\right) - \frac{c_2}{c_1}, \right. \tag{C.12b}$$

$$\left. t, e^{\epsilon}(u-g_2) - g_2\right) \qquad \textit{scaling and shift.}$$

and

$$\mathbf{X}_{8b} = \frac{2}{g_1\,c_1}(2\,g_1+3)(c_1\,x+c_2)\left(x+\frac{c_2}{c_1}\right)^{-\frac{g_1+1}{2\,g_1+3}}\partial_x + \tag{C.13a}$$

$$+\frac{2}{c_1}(g_1+1)(u-g_2)\left(x+\frac{c_2}{c_1}\right)^{-\frac{g_1+1}{2\,g_1+3}}\partial_u.$$

$$(x^*, t^*, u^*) = \left(\left(\left(x+\frac{c_2}{c_1}\right)^{\frac{g_1+1}{2\,g1+3}} + 2\epsilon + \frac{2\epsilon}{g_1}\right)^{\frac{2\,g_1+3}{g_1+1}} - \frac{c_2}{c_1}, t, \right. \tag{C.13b}$$

$$\left. e^{\left(\frac{2}{c_1}(g_1+1)\,g1\,\epsilon\left(\left(x+\frac{c_2}{c_1}\right)^{\frac{g_1+1}{2\,g_1+3}}\,g_1+2\epsilon\,(g_1+1)\right)^{-1}\right)}(u-g_2)+g_2\right)$$

exponential dilation and shift.

8.2. We consider $c\,(x) = -\frac{f_1\,g_2\,x^2}{2\,g_3} + c_1\,x + c_2$ for arbitrary values of the constants c_1, c_2. If $f_2 = 0$ and $g_1 = -1$, then equation (6.4) admits the generators $\mathbf{X}_1$, $\mathbf{X}_4$ and $\mathbf{X}_{8c}$ for $K = \sqrt{g_3\ ({c_1}^2 g_3 + 2\,c_2\,f_1\,g_2)}$ with

$$r(x) = \operatorname{arctanh}\left(\frac{c_1\,g_3 - f_1\,g_2\,x}{K}\right) \tag{C.14}$$

and

$$\mathbf{X}_{8c} = \frac{1}{2} r(x) \left(f_1 g_2 x^2 - 2 g_3 (c_1 x + c_2) \right) \partial_x + \tag{C.15a}$$
$$+ (g_2 - u) \left((f_1 g_2 x - c_1 g_3) r(x) + K \right) \partial_u.$$

$$(x^*, t^*, u^*) = \frac{1}{f_1 g_2} \left(-\tanh\left(e^{1/2\, K \epsilon} r(x) \right) K + c_1 g_3 \right), t, \tag{C.15b}$$
$$(u - g_2)\, \mathrm{e}^{-\left(-\tanh\left(e^{\frac{K\epsilon}{2}} r(x) \right) K \operatorname{arctanh}\left(\tanh\left(e^{\frac{K\epsilon}{2}} r(x) \right) \right) + K \right) \epsilon} + g_2 \Bigg)$$

oscillatory dilation and shift.

8.3. If $f_2 = 0$ and $g_1 = -2$, then equation (6.4) admits the generators $\mathbf{X}_1$, $\mathbf{X}_{4a}$ and $\mathbf{X}_{8d}$ and $\mathbf{X}_{8e}$ for any $c(x)$ that verifies

$$c''(x) = \frac{c'(x)^2 g_3 - 4\, c(x) f_1}{2\, g_3\, c(x)} \tag{C.16}$$

with

$$\mathbf{X}_{8d} = \partial_x + \frac{(u - g_2) c'(x)}{2\, c(x)} \partial_u. \tag{C.17a}$$

$$(x^*, t^*, u^*) = \left(x + \epsilon, t, \exp\left(\frac{\epsilon\, c'(x + \epsilon)}{2\, c(x + \epsilon)} \right) (u - g_2) + g_2 \right) \tag{C.17b}$$

scaling and shift.

and

$$\mathbf{X}_{8e} = x\, \partial_x + (u - g_2) \left(\frac{x\, c'(x)}{2\, c(x)} - 1 \right) \partial_u. \tag{C.18a}$$

$$(x^*, t^*, u^*) = \left(e^{\epsilon} x, t, \exp\left(\frac{\epsilon\, (x + \epsilon)\, c'(x + \epsilon)}{2\, c(x + \epsilon)} - \epsilon \right) (u - g_2) + g_2 \right) \tag{C.18b}$$

scaling and shift.

8.4. We consider $c(x) = c_2 \exp(c_1 x)$ for arbitrary values of the constants c_1, c_2. If $f_2 = 0$ and $g_1 = -\frac{3}{2}$, then equation (6.4) admits the generators $\mathbf{X}_1$, $\mathbf{X}_4$ and

$$\mathbf{X}_{8f} = \frac{3}{2\, c_1} \partial_x + (u - g_2) \partial_u. \tag{C.19a}$$

$$(x^*, t^*, u^*) = \left(x + \frac{3}{2\,c_1}\epsilon, t, e^{\epsilon}(u - g_2) + g_2 \right) \quad \textit{scaling and shift.} \tag{C.19b}$$

8.5. If $f_2 = 0$ and $g_1 = -\frac{4}{3}$ then equation (6.4) admits $\mathbf{X}_1$ and $\mathbf{X}_4$ as generators with $c(x)$ an arbitrary function.

8.6. If $f_2 \neq 0$ then equation (6.4) admits $\mathbf{X}_1$ and $\mathbf{X}_4$ as generators with $c(x)$ an arbitrary function.

Case 9. For $f(u) = f_1\,(u - g_2) + f_2(u - g_2)^{g_1+1}$, $g(u) = g_3(g_2 - u)^{g_1}$ we distinguish the following subcases:

9.1. For $c(x) = -\frac{(c_1-x)^2}{2(2+g_1)}$, with $g_1 \neq -2, -\frac{4}{3}, -1$, and $f_1 \neq 0$, we obtain the generators $\mathbf{X}_1$, $\frac{1}{f_1}\mathbf{X}_4$ and the following:

$$\mathbf{X}_9 = \frac{x - c_1}{\sqrt{-2(2 + g_1)}}\partial_x. \tag{C.20a}$$

$$(x^*, t^*, u^*) = \left(\exp\left(\ln|x - c_1| + \frac{\epsilon}{\sqrt{-2(2 + g_1)}} \right) + c_1, t, u \right) \tag{C.20b}$$

exponential dilation and shift.

9.2. If $f_1 \neq 0$ and $g_1 = -2, -\frac{4}{3}, -1$, we obtain the generators $\mathbf{X}_1$, and $\mathbf{X}_4$ for arbitrary $c(x)$.

9.3. If $f_1 = 0$ we obtain the generators $\mathbf{X}_1$ and $\mathbf{X}_{4a}$ for arbitrary $c(x)$.

Case 10. For $f(u) = f_1\left(u + \frac{g_2}{g_3}\right) + f_2\left(u + \frac{g_2}{g_3}\right)^{-\frac{1}{3}}$, $g(u) = \left(-\frac{4}{3}\,(g_3\,u + g_2)^{-1}\right)^{\frac{4}{3}}$ and $c(x) = \frac{1}{4}\,(c_1\,x + c_2)^2$ with arbitrary values of the constants f_1, f_2, g_3, g_2, c_1, c_2 we distinguish the following subcases:

10.1. If $g_3 = -1$ then we obtain for equation (6.4) the generators $\mathbf{X}_1$, $\mathbf{X}_4$ for $g_1 = -\frac{4}{3}$, and

$$\mathbf{X}_{10a} = \frac{1}{2}\left(x + \frac{c_2}{c_1}\right)\partial_x. \tag{C.21a}$$

$$(x^*, t^*, u^*) = \left(\exp\left(\frac{\epsilon}{2}\right)\left(x + \frac{c_2}{c_1}\right) - \frac{c_2}{c_1}, t, u\right) \quad \textit{scaling and shift.} \tag{C.21b}$$

10.2. If $g_3 \neq -1$ then we obtain the generators $\mathbf{X}_1$ and $\mathbf{X}_7$

Case 11. For $f(u) = f_2\, e^{f_1\, u}$, $g(u) = g_2\, e^{g_1\, u}$, $c(x) = c_2\, e^{c_1\, x}$ with arbitrary values of the constants f_1, f_2, g_3, g_2, c_1, c_2 we distinguish the following subcases:

11.1. If $f_1 \neq g_1$, then we obtain the generator $\mathbf{X}_1$ and the following

$$\mathbf{X}_{11a} = \frac{c_1\, f_1\, t}{f_1 - g_1}\partial_t + \partial_x - \frac{c_1}{f_1 - g_1}\partial_u. \tag{C.22a}$$

$$(x^*, t^*, u^*) = \left(\exp\left(\frac{c_1\, f_1}{f_1 - g_1}\epsilon\right)\, t, x + \epsilon, u - \frac{c_1}{f_1 - g_1}\epsilon\right) \quad \textit{scaling and shift.} \tag{C.22b}$$

11.2. If $f_1 = g_1$, then we obtain the generator $\mathbf{X}_1$ and the following

$$\mathbf{X}_{11b} = t\,\partial_t - \frac{1}{f_1}\partial_u \tag{C.23a}$$

$$(x^*, t^*, u^*) = \left(e^{\epsilon}\, t, x, u - \frac{1}{f_1}\epsilon\right) \quad \textit{scaling and shift.} \tag{C.23b}$$

C.2 Lie symmetry generators for Eq. (6.6)

In this Section we present the corresponding generators for Eq. (6.6). **Case 1**. For $F(t)$ an arbitrary function we get the generator

$$\mathbf{X}_1 = \partial_x. \tag{C.24}$$

Case 2. For $F = F(t)$ verifying that

$$F'''' = \frac{\sum_{i=0}^{6} F_i(t)}{F_0(t)} \tag{C.25a}$$

where

$$F_0(t) = F^2\,\left(-F^4 + 2\,F\,F'' - 3(F')^2\right), \tag{C.25b}$$
$$F_1(t) = 3\,F^3\,(F''')^2\,, \tag{C.25c}$$
$$F_2(t) = -F'\,F\,\left(11\,F^4 + 14\,F\,F'' - 3(F')^2\right), \tag{C.25d}$$
$$F_3(t) = 8\,F^2\,(F'')^3, \tag{C.25e}$$
$$F_4(t) = 2F\,\left((F')^2 - F^4\right), \tag{C.25f}$$
$$F_5(t) = -F^8 + 40\,F^4\,(F')^2 - 3\,(F')^4, \tag{C.25g}$$
$$F_6(t) = -2\,(F')^2\,F^3\,(15(F')^2 - F^4), \tag{C.25h}$$

with $F_0(t) \neq 0$, then we obtain the characteristic form $\mathbf{P} = (A_1 x + A_2)u_x + B(t, x, u, u_t)$, where $A_1, A_2 \in \mathbb{R}$ and $B = B(t, x, u, u_t)$ depends on the function F such that

$$B(t, x, u, u_t) = A_1 \frac{B_1(t)u_t + B_2(t)u + B_3(t)}{B_0(t)} \tag{C.26a}$$

and

$$B_0(t) = F(F'\left(F^4 + 4\,F\,F'' - 3\,(F')^2\right) - F^2 F'''), \tag{C.26b}$$
$$B_1(t) = 2\,F^2\left(-F^4 + 2\,F\,F'' - 3\,(F')^2\right), \tag{C.26c}$$
$$B_2(t) = 2\,F\left(6\,F'(F\,F'' - (F')^2) - F^2 F'''\right), \tag{C.26d}$$
$$B_3(t) = F\,F'''\left(F' + F^2\right) + F''\left(F^4 - 6F'\,F^2 + (F')^2\right) - \\ - 2F\left((F'')^2 + (F')^2(F^2 - 3\,F')\right), \tag{C.26e}$$

for $B_0(t) \neq 0$.

Case 3. When considering $F_0(t) = 0$ in Eq. (C.25b), we obtain

$$F(t) = \frac{4\,f_1}{(t + f_2)^2\,{f_1}^2 - 4} \tag{C.27}$$

for $f_1, f_2 \in \mathbb{R}$, so that we get the generators $\mathbf{X}_1$,

$$\mathbf{X}_{31} = \partial_t + \frac{G_1(t,u)}{G_0(t)}\partial_u, \qquad \mathbf{X}_{32} = t\partial_t + \frac{x}{2}\partial_x + \frac{G_2(t,u)}{G_0(t)}\partial_u, \tag{C.28}$$

where

$$\begin{aligned} G_0(t) = &\ 4\left((t+f_2)^2 f_1^2 - 4\right), \\ G_1(t,u) = &\ f_1\left(4 + (t+f_2)^2 {f_1}^2 + 4\,(t+f_2)\,(2\,u-1)\,f_1\right), \\ G_2(t,u) = &\ 2\left((t^2-f_2^2)(2u-1){f_1}^2 + 2f_1(2\,t+f_2) + 8u - 4\right) \\ &-f_2\,(t+f_2)^2\,{f_1}^3. \end{aligned}$$

This case arises from the equation

$$k_1^2\,F^2 + 2\,F\,F'' - 3\,(F')^2 - F^4 = 0. \tag{C.29}$$

For $k_1 = 0$ we obtain the solutions $F = F(t)$ as (C.27) and

$$F(t) = \frac{1}{f_2 \pm t}, \quad f_2 \in \mathbb{R}. \tag{C.30}$$

With the change of variables

$$u(t,x) = h(t) + g(t)\cdot v(t,x) \tag{C.31}$$

with

$$h(t) = \frac{F' + F^2}{2\,F^2}, \qquad g(t) = -\frac{1}{F}, \tag{C.32}$$

it yields that Eq. (6.6) can be transformed into

$$v_t = v_{x,x} + v^2, \tag{C.33}$$

For this equation, a symmetry involving a non-Painlevé-type second-order ODE reduction can be obtained [283]. The case for F as in (C.30) will be studied in Section C.2.1.
Whenever $k_1 \neq 0$ in (C.29), with the change of variables (C.31) and

$$h(t) = \frac{k_1\,F + F' + F^2}{2\,F^2}, \qquad g(t) = \frac{k_2}{F}, \tag{C.34}$$

Eq. (6.6) can be transformed into

$$v_t = v_{x,x} - k_1\, v - k_2\, v^2. \tag{C.35}$$

The particular case when $k_1 = -1$ has already been studied as a linear diffusion equation [325], where solutions considering travelling waves can be provided for Eq. (C.35).

Case 4. We now consider whenever $B_0(t) = 0$, as stated in Eq. (C.26b), yielding that for any $F = F(t)$ with

$$F'\left(F^4 + 4\,F\,F'' - 3\,(F')^2\right) - F^2 F''' = 0, \tag{C.36}$$

then F also verifies Eq. (C.25). For any $f_1, f_2, f_3 \in \mathbb{R}$, the solution of Eq. (C.36) is

$$F(t) = \frac{4\,e^{-\sqrt{f_1}(f_3+t)} f_1}{\left(e^{-\sqrt{f_1}(f_3+t)}\right)^2 + 4\,f_2\,e^{-\sqrt{f_1}(f_3+t)} + 4\,{f_2}^2 - 4\,f_1}. \tag{C.37}$$

We obtain the generators $\mathbf{X}_1$ and $\mathbf{X}_4$, this last in terms of F and its derivatives as

$$\mathbf{X}_4 = \partial_t - \left(-\frac{F'\,u}{F} + \frac{F'}{2\,F} + \frac{F''}{2\,F^2} - \frac{(F'')^2}{2\,F^3}\right)\partial_u. \tag{C.38}$$

Case 5. For $F = F(t)$ verifying

$$F''' = \frac{2(F'')^2 F + (-F^4 + 6F^2(F') - (F')^2)F'' + 2(F')^2F^3 - 6(F')^3F}{F(F^2 + F')} \tag{C.39}$$

we obtain the generators $\mathbf{X}_1$ and

$$\mathbf{X}_5 = x\,\partial_x + \frac{-2F(F^2 + F')\partial_t + 2u((F'')F - 2(F')^2)\partial_u}{F^2(F') + (F'')F - (F')^2}. \tag{C.40}$$

For the sake of simplicity, we have assumed that the functions

$$F(t) = \frac{1}{f_1 + f_2\,t}, \tag{C.41}$$

$$F(t) = \frac{f_1\,\exp(f_1\,t + f_2)}{\exp(f_1\,t + f_2) - 1}, \tag{C.42}$$

are not considered in the prior cases, as the function in Eq. (C.42) only holds within Case 1, and the function in Eq. (C.41) will be specifically exposed in the next Subsection.

C.2.1 Generators and reductions for special cases of F

From the general analysis arises the function

$$F(t) = \frac{1}{f_1 + f_2\, t} \tag{C.43}$$

which is studied independently, as it does not fit in the cases studied previously. Its generators are studied in terms of the parameters $f_1, f_2 \in \mathbb{R}$. For these cases, we will consequently obtain the resulting reduced equations, as $F = F(t)$ is given explicitly:

Case 6.1. For $F(t) = \dfrac{1}{f_1 + f_2\, t}$ and $f_2 \in \mathbb{R} - \{1, -1\}$, we obtain the generator $\mathbf{X}_1$ and the generator

$$\mathbf{X}_{61} = 2\left(t + \frac{f_1}{f_2}\right)\partial_t + x\,\partial_x. \tag{C.44}$$

Generator $\mathbf{X}_{61}$ yields a similarity variable and similarity solution of the form

$$z = \frac{\sqrt{\frac{f_1}{f_2} + t}}{x}, \qquad u = h(z), \tag{C.45}$$

obtaining the reduced ODE_{61}:

$$b\, z^4 h_{zz} + h_z\left(\frac{b\, z}{2} - 2\, b\, z^3\right) + h(1-h) = 0. \tag{C.46}$$

Case 6.2. For $f_2 = 1$ we obtain the generators $\mathbf{X}_1$ and

$$\mathbf{X}_{621} \quad = \partial_t + \frac{u}{f_1 + t}\partial_u, \qquad \mathbf{X}_{622} \quad = t\,\partial_t + \frac{x}{2}\partial_x - \frac{f_1\, u}{f_1 + t}\partial_u. \tag{C.47}$$

Generator $\mathbf{X}_{621}$ yields a similarity variable and similarity solution of the form

$$z = t, \qquad u = h(z)(f_1 + z), \tag{C.48}$$

obtaining the reduced ODE_{621}:

$$-\, h_z\,(f_1 + z) - h\, f_1 + h - h^2(f_1 + z) = 0. \tag{C.49}$$

For the generator $\mathbf{X}_{622}$, a similarity variable and similarity solution is obtained:

$$z = \frac{x^2}{t}, \qquad u = h(z)\frac{f_1 + t}{t}, \tag{C.50}$$

yielding the reduced ODE_{622}:

$$4\,h_{z,z}\,z + h_z\,z + 2\,h_z + h(1-h) = 0. \tag{C.51}$$

Case 6.3. For $f_2 = -1$ we obtain the generator $\mathbf{X}_1$ and besides

$$\mathbf{X}_{631} = \partial_t + \frac{1-u}{f_1-t}\partial_u, \qquad \mathbf{X}_{632} = t\,\partial_t + \frac{x}{2}\partial_x + \frac{f_1(1-u)}{f_1-t}\partial_u. \tag{C.52}$$

Generator $\mathbf{X}_{631}$ yields a similarity variable and similarity solution of the form

$$z = t, \qquad u = 1 - h(z)(f_1 - z), \tag{C.53}$$

obtaining the reduced ODE_{631}:

$$h_z\,(f_1 - z) + h\,f_1 + h - h^2(f_1 - z) = 0. \tag{C.54}$$

For the generator $\mathbf{X}_{632}$, a similarity variable and similarity solution is obtained

$$z = \frac{x^2}{t}, \qquad u = 1 - h(z)\frac{t - f_1}{t} \tag{C.55}$$

yielding the same reduce ODE_{632} as in (C.51).

C.3 Lie symmetry generators for Eq. (6.7)

In this Section we present the corresponding generators for Eq. (6.7). For the sake of readability please see the notes below.

Notes:
(1) In this case α, f_0 and g_1 must satisfy the condition

$$H_1(x)^2 H(x) = \text{constant}, \tag{C.56}$$

where

$$H_1(x) = e^{-A}\left(c_4(3g_1+4) + 2(g_1c_1 + c_2)\int e^A\,dx\right), \tag{C.57}$$

$$H(x) = 2g_0((2+g_1)\alpha^2 + (3g_1+4)\alpha_x) - f_0(3g_1+4)^2. \tag{C.58}$$

(2) The constants c_1, c_2 and c_4 are linked to α, f_0 and g_1 by condition (C.56).
(3) In this case, α, f_0, and g_1 must satisfy the condition

$$H_2(x)^2 H(x) = \text{constant}, \tag{C.59}$$

i	α	f	$\mathbf{v}_k$
1.1	$\forall$	$f_0 u^{g_1+1}$	$\mathbf{v}_3 = t\partial_t - \frac{u}{g_1}\partial_u$
1.2	$\forall$	$f_0 u^{g_1+1} + f_1 u,\ f_1 \neq 0$	$\mathbf{v}_4 = \frac{e^{-f_1 g_1 t}}{f_1}\partial_t + e^{-f_1 g_1 t} u \partial_u$
1.3	(1)	$f_0 u^{g_1+1},\ g_1 \neq -4/3$	$\mathbf{v} = c_2\mathbf{v}_3 + c_4\mathbf{v}_{51} + \frac{c_2+c_1 g_1}{g_1(4+3g_1)}\mathbf{v}_{52}$, (2) $\mathbf{v}_{51} = e^{-A}\left(\partial_x - \frac{2\alpha u}{3g_1+4}\partial_u\right)$, with $A = \frac{g_1}{3g_1+4}\int \alpha\, dx$, $\mathbf{v}_{52} = \left(2g_1 e^{-A}\int e^{A}\, dx\right)\partial_x +$ $+4u\left(1 - \frac{g_1 \alpha e^{-A}}{3g_1+4}\int e^{A}\, dx\right)\partial_u$
1.4	(3)	$f_0 u^{g_1+1} + f_1 u,\ f_1 \neq 0,\ g_1 \neq -4/3$	$\mathbf{v}_4, \mathbf{v} = c_4\mathbf{v}_{51} + \frac{c_1}{4+3g_1}\mathbf{v}_{52}$, (4)
1.5	$\frac{\alpha_1}{x}$	$f_0 u^{f_1},\ f_1 \neq g_1 + 1$	$\mathbf{v}_6 = \frac{2(1-f_1)t}{1+g_1-f_1}\partial_t + x\partial_x + \frac{2u}{1+g_1-f_1}\partial_u$

Table C.1: Lie symmetry generators for Eq. (6.7) for $g = g_0 u^{g_1}$.

i	α	f	$\mathbf{v}_k$
2.1	(5)	$f_0 u^{-1/3}$	$\mathbf{v}_3, \mathbf{v}_{50} = \frac{1}{\alpha}\partial_x + \frac{3\alpha_x}{2\alpha^2}u\partial_u$
2.2	(5)	$f_0 u^{-1/3} + f_1 u,\ f_1 \neq 0$	$\mathbf{v}_4, \mathbf{v}_{50}$

Table C.2: Lie symmetry generators for Eq. (6.7) for $g = g_0 u^{-4/3}$.

where $H(x)$ is given by (C.58), and

$$H_2(x) = e^{-A}\left(c_4(3g_1+4) + 2g_1 c_1 \int e^{A}\, dx\right). \tag{C.60}$$

(4) The constants c_1 and c_4 are linked to α, f_0, and g_1 by condition (C.59).
(5) In this case α and f_0 must satisfy the equation

$$3g_0(\alpha^3\alpha_{xx} - 2\alpha^2\alpha_x^2 + 6\alpha_x^3 - 6\alpha\alpha_x\alpha_{xx} + \alpha^2\alpha_{xxx}) - 4f_0\alpha^2\alpha_x = 0. \tag{C.61}$$

(6) In this case α, f_0, and g_1 must satisfy the condition

$$H_3(x)^2 H_5(x) = \text{constant}, \tag{C.62}$$

where

$$H_3(x) = e^{-B}\left(c_5 + \frac{2}{3}c_1 \int e^{B}\, dx\right), \tag{C.63}$$

i	α	f	$\mathbf{v}_k$
3.1	$\forall$	$f_0e^{g_1u}+f_1$, $f_1\neq 0$	$\mathbf{v}_1,\mathbf{v}_7=\dfrac{e^{-f_1g_1t}}{f_1}\partial_t+e^{-f_1g_1t}\partial_u$
3.2	$\forall$	$f_0e^{g_1u}$	$\mathbf{v}_1,\mathbf{v}_8=t\partial_t-\frac{1}{g_1}\partial_u$
3.3	(6)	$f_1+f_0e^{g_1u}$, $f_1\neq 0$	$\mathbf{v}_1,\mathbf{v}_7$, $\mathbf{v}_9=c_5\mathbf{v}_{91}+c_1\mathbf{v}_{92}$, (7) $\mathbf{v}_{91}=e^{-B}\left(\partial_x-\frac{2\alpha}{3g_1}\partial_u\right)$, with $B=\frac{1}{3}\int\alpha\,dx$ $\mathbf{v}_{92}=\left(\frac{2}{3}e^{-B}\int e^B\,dx\right)\partial_x-\frac{4}{9g_1}\left(\alpha e^{-B}\int e^B\,dx-3\right)\partial_u$
3.4	(8)	$f_0e^{g_1u}$	$\mathbf{v}_1,\mathbf{v}_9=c_2\mathbf{v}_8+c_5\mathbf{v}_{91}+c_1\mathbf{v}_{92}$, (9)
3.5	$\dfrac{\alpha_1}{x}$	$f_0e^{f_1u}$	$\mathbf{v}_1,\mathbf{v}_{10}=\dfrac{2t}{f_1-g_1}\partial_t+\dfrac{x}{f_1}\partial_x-\dfrac{2}{f_1(f_1-g_1)}\partial_u$

Table C.3: Lie symmetry generators for Eq. (6.7) for $g=g_0e^{ug_1}$

$$H_5(x)=9g_1f_0-2g_0(3\alpha_x+\alpha^2). \tag{C.64}$$

(7) The constants c_1 and c_5 are linked to α, f_0, and g_1 by condition (C.62).
(8) In this case α, f_0, and g_1 must satisfy the condition

$$H_4(x)^2H_5(x)=\text{constant}, \tag{C.65}$$

where $H_5(x)$ is given by (C.64), and

$$H_4(x)=e^{-B}\left(c_5+\frac{2}{3}(c_1+c_2)\int e^B\,dx\right). \tag{C.66}$$

(9) The constants c_1, c_2, and c_5 are linked to α, f_0 and g_1 by condition (C.65).

Bibliography

[1] F. Bray et al. 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries'. In: *CA: A Cancer Journal for Clinicians* vol. 68, no. 6 (2018), pp. 394–424. DOI: 10.3322/caac.21492.

[2] K. D. Miller et al. 'Cancer treatment and survivorship statistics, 2016'. In: *CA: A Cancer Journal for Clinicians* vol. 66, no. 4 (2016), pp. 271–289. DOI: 10.3322/caac.21349.

[3] E. Desandes and D. P. Stark. 'Epidemiology of Adolescents and Young Adults with Cancer in Europe'. In: *Progress in Tumor Research*. 2016, pp. 1–15. DOI: 10.1159/000447037.

[4] R. L. Siegel, K. D. Miller and A. Jemal. 'Cancer statistics, 2020'. In: *CA: A Cancer Journal for Clinicians* vol. 70, no. 1 (2020), pp. 7–30. DOI: 10.3322/caac.21590.

[5] J. C. Wang and J. E. Dick. 'Cancer stem cells: lessons from leukemia'. In: *Trends in Cell Biology* vol. 15, no. 9 (2005), pp. 494–501. DOI: 10.1016/j.tcb.2005.07.004.

[6] T. Reya et al. 'Stem cells, cancer, and cancer stem cells'. In: *Nature* vol. 414, no. 6859 (2001), pp. 105–111. DOI: 10.1038/35102167.

[7] E. Laurenti and B. Göttgens. 'From haematopoietic stem cells to complex differentiation landscapes.' In: *Nature* vol. 553 (7689 2018), pp. 418–426. DOI: 10.1038/nature25022.

[8] L. Velten et al. 'Human haematopoietic stem cell lineage commitment is a continuous process'. In: *Nature Cell Biology* vol. 19, no. 4 (2017), pp. 271–281. DOI: 10.1038/ncb3493.

[9] A. Fasano and A. Sequeira. *Hemomath*. 30th Oct. 2017. DOI: 10.1007/978-3-319-60513-5.

[10] K. Rehe et al. 'Acute B lymphoblastic leukaemia-propagating cells are present at high frequency in diverse lymphoblast populations'. In: *EMBO Molecular Medicine* vol. 5, no. 1 (2012), pp. 38–51. DOI: 10.1002/emmm.201201703.

[11] D. Bonnet and J. E. Dick. 'Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell'. In: *Nature Medicine* vol. 3, no. 7 (1997), pp. 730–737. DOI: 10.1038/nm0797-730.

[12] K. J. Hope, L. Jin and J. E. Dick. 'Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity'. In: *Nature Immunology* vol. 5, no. 7 (2004), pp. 738–743. DOI: 10.1038/ni1080.

[13] E. Passegue et al. 'Normal and leukemic hematopoiesis: Are leukemias a stem cell disorder or a reacquisition of stem cell characteristics?' In: *Proceedings of the National Academy of Sciences* vol. 100, no. Supplement 1 (2003), pp. 11842–11849. DOI: 10.1073/pnas.2034201100.

[14] D. Hanahan and R. A. Weinberg. 'Hallmarks of Cancer: The Next Generation'. In: *Cell* vol. 144, no. 5 (2011), pp. 646–674. DOI: 10.1016/j.cell.2011.02.013.

[15] M. H. Cheok and W. E. Evans. 'Acute lymphoblastic leukaemia: a model for the pharmacogenomics of cancer therapy'. In: *Nature Reviews Cancer* vol. 6, no. 2 (2006), pp. 117–129. DOI: 10.1038/nrc1800.

[16] H. T. Maecker, J. P. McCoy and R. Nussenblatt. 'Standardizing immunophenotyping for the Human Immunology Project'. In: *Nature Reviews Immunology* vol. 12, no. 3 (2012), pp. 191–200. DOI: 10.1038/nri3158.

[17] F. J. O. Giner and A. Orfao. 'Aplicación de la citometría de flujo al diagnóstico y seguimiento inmunofenotípico de las leucemias agudas'. In: *Medicina Clínica* vol. 118, no. 11 (2002), pp. 423–436. DOI: 10.1016/s0025-7753(02)72408-1.

[18] E. van Lochem et al. 'Immunophenotypic differentiation patterns of normal hematopoiesis in human bone marrow: Reference patterns for age-related changes and disease-induced shifts'. In: *Cytometry* vol. 60B, no. 1 (2004), pp. 1–13. DOI: 10.1002/cyto.b.20008.

[19] J. J. M. van Dongen et al. 'EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes'. In: *Leukemia* vol. 26, no. 9 (2012), pp. 1908–1975. DOI: 10.1038/leu.2012.120.

[20] V. M. Pérez-García et al. 'Applied mathematics and nonlinear sciences in the war on cancer'. In: *Applied Mathematics and Nonlinear Sciences* vol. 1, no. 2 (2016), pp. 423–436. DOI: 10.21042/amns.2016.2.00036.

[21] H. M. Byrne. 'Dissecting cancer through mathematics: from the cell to the animal model'. In: *Nature Reviews Cancer* vol. 10, no. 3 (2010), p. 221. DOI: 10.1038/nrc2808.

[22] P. M. Altrock, L. L. Liu and F. Michor. 'The mathematics of cancer: integrating quantitative models'. In: *Nature Reviews Cancer* vol. 15, no. 12 (2015), p. 730. DOI: 10.1038/nrc4029.

[23] Y. Saeys, S. V. Gassen and B. N. Lambrecht. 'Computational flow cytometry: helping to make sense of high-dimensional immunology data'. In: *Nature Reviews Immunology* vol. 16, no. 7 (2016), pp. 449–462. DOI: 10.1038/nri.2016.56.

[24] N. Aghaeepour et al. 'Critical assessment of automated flow cytometry data analysis techniques'. In: *Nature Methods* vol. 10, no. 3 (2013), pp. 228–238. DOI: 10.1038/nmeth0513-445c.

[25] L. M. Weber and M. D. Robinson. 'Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data'. In: *Cytometry Part A* vol. 89, no. 12 (2016), pp. 1084–1096. DOI: 10.1002/cyto.a.23030.

[26] Y. Saeys, S. V. Gassen and B. Lambrecht. 'Response to Orlova et al. Science not art: statistically sound methods for identifying subsets in multi-dimensional flow and mass cytometry data sets'. In: *Nature Reviews Immunology* vol. 18, no. 1 (2017), pp. 78–78. DOI: 10.1038/nri.2017.151.

[27] E. J. deAndrés-Galiana et al. 'On the prediction of Hodgkin lymphoma treatment response'. In: *Clinical and Translational Oncology* vol. 17, no. 8 (2015), pp. 612–619. DOI: 10.1007/s12094-015-1285-z.

[28] G. D. Clapp and D. Levy. 'A review of mathematical models for leukemia and lymphoma'. In: *Drug Discovery Today: Disease Models* vol. 16 (2015), pp. 1–6. DOI: 10.1016/j.ddmod.2014.10.002.

[29] C. L. Sawyers. 'Chronic Myeloid Leukemia'. In: *New England Journal of Medicine* vol. 340, no. 17 (1999), pp. 1330–1340. DOI: 10.1056/nejm199904293401706.

[30] B. Lowenberg, J. R. Downing and A. Burnett. 'Acute Myeloid Leukemia'. In: *New England Journal of Medicine* vol. 341, no. 14 (1999), pp. 1051–1062. DOI: 10.1056/nejm199909303411407.

[31] T. Stiehl and A. Marciniak-Czochra. 'How to characterize stem cells? contributions from mathematical modeling'. In: *Current Stem Cell Reports* vol. 5, no. 2 (2019), pp. 57–65.

[32] A. Marciniak-Czochra et al. 'Modeling of Asymmetric Cell Division in Hematopoietic Stem Cells—Regulation of Self-Renewal Is Essential for Efficient Repopulation'. In: *Stem Cells and Development* vol. 18, no. 3 (2009), pp. 377–386. DOI: 10.1089/scd.2008.0143.

[33] T. Stiehl et al. 'Cell Division Patterns in Acute Myeloid Leukemia Stem-like Cells Determine Clinical Course: A Model to Predict Patient Survival'. In: *Cancer Research* vol. 75, no. 6 (2015), pp. 940–949. DOI: 10.1158/0008-5472.can-14-2508.

[34] T. Stiehl and A. Marciniak-Czochra. 'Mathematical Modeling of Leukemogenesis and Cancer Stem Cell Dynamics'. In: *Mathematical Modelling of Natural Phenomena* vol. 7, no. 1 (2012), pp. 166–202. DOI: 10.1051/mmnp/20127199.

[35] T. Stiehl et al. 'Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation patterns at diagnosis and relapse'. In: *Journal of The Royal Society Interface* vol. 11, no. 94 (2014), pp. 20140079–20140079. DOI: 10.1098/rsif.2014.0079.

[36] T. Stiehl and A. Marciniak-Czochra. 'Stem cell self-renewal in regeneration and cancer: insights from mathematical modeling'. In: *Current Opinion in Systems Biology* vol. 5 (2017), pp. 112–120. DOI: 10.1016/j.coisb.2017.09.006.

[37] T. Stiehl, C. Lutz and A. Marciniak-Czochra. 'Emergence of heterogeneity in acute leukemias'. In: *Biology Direct* vol. 11, no. 51 (2016). DOI: 10.1186/s13062-016-0154-1.

[38] T. Lorenzi, A. Marciniak-Czochra and T. Stiehl. 'A structured population model of clonal selection in acute leukemias with multiple maturation stages'. In: *Journal of Mathematical Biology* vol. 79, no. 5 (2019), pp. 1587–1621. DOI: 10.1007/s00285-019-01404-w.

[39] T. Walenda et al. 'Feedback Signals in Myelodysplastic Syndromes: Increased Self-Renewal of the Malignant Clone Suppresses Normal Hematopoiesis'. In: *PLoS Computational Biology* vol. 10, no. 4 (2014). Ed. by Q. Nie, e1003599. DOI: 10.1371/journal.pcbi.1003599.

[40] T. Stiehl, A. D. Ho and A. Marciniak-Czochra. 'Mathematical modeling of the impact of cytokine response of acute myeloid leukemia cells on patient prognosis'. In: *Scientific reports* vol. 8, no. 1 (2018), pp. 1–11.

[41] J. G. Milton and M. C. Mackey. 'Periodic haematological diseases: mystical entities or dynamical disorders?' In: *Journal of the Royal College of Physicians of London* vol. 23 (4 1989), pp. 236–241.

[42] M. . Weis et al. 'A data-driven, mathematical model of mammalian cell cycle regulation'. In: *PLoS One* vol. 9, no. 5 (2014), e97130. DOI: 10.1371/journal.pone.0097130.

[43] C. Foley and M. C. Mackey. 'Dynamic hematological disease: a review'. In: *Journal of mathematical biology* vol. 58, no. 1-2 (2009), pp. 285–322. DOI: 10.1007/s00285-008-0165-3.

[44] M. C. Mackey. 'Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis.' In: *Blood* vol. 51 (5 1978), pp. 941–956.

[45] M. C. Mackey et al. 'Periodic Oscillations of Blood Cell Populations in Chronic Myelogenous Leukemia'. In: *SIAM Journal on Mathematical Analysis* vol. 38, no. 1 (2006), pp. 166–187. DOI: 10.1137/04061578x.

[46] C. Colijn and M. C. Mackey. 'A mathematical model of hematopoiesis—I. Periodic chronic myelogenous leukemia'. In: *Journal of Theoretical Biology* vol. 237, no. 2 (2005), pp. 117–132. DOI: 10.1016/j.jtbi.2005.03.033.

[47] C. Colijn and M. C. Mackey. 'A mathematical model of hematopoiesis: II. Cyclical neutropenia'. In: *Journal of Theoretical Biology* vol. 237, no. 2 (2005), pp. 133–146. DOI: 10.1016/j.jtbi.2005.03.034.

[48] A. Safarishahrbijari and A. Gaffari. 'Parameter identification of hematopoiesis mathematical model - periodic chronic myelogenous leukemia'. In: *Współczesna Onkologia* vol. 1 (2013), pp. 73–77. DOI: 10.5114/wo.2013.33778.

[49] L. Pujo-Menjouet and M. C. Mackey. 'Contribution to the study of periodic chronic myelogenous leukemia'. In: *Comptes Rendus Biologies* vol. 327, no. 3 (2004), pp. 235–244. DOI: 10.1016/j.crvi.2003.05.004.

[50] A. Halanay. 'Periodic Solutions in a Mathematical Model for the Treatment of Chronic Myelogenous Leukemia'. In: *Mathematical Modelling of Natural Phenomena* vol. 7, no. 1 (2012), pp. 235–244. DOI: 10.1051/mmnp/20127110.

[51] S. I. Rubinow and J. L. Lebowitz. 'A mathematical model of the acute myeloblastic leukemic state in man'. In: *Biophysical Journal* vol. 16, no. 8 (1976), pp. 897–910. DOI: 10.1016/s0006-3495(76)85740-2.

[52] S. I. Rubinow and J. L. Lebowitz. 'A mathematical model of neutrophil production and control in normal man'. In: *Journal of Mathematical Biology* vol. 1, no. 3 (1975), pp. 187–225. DOI: 10.1007/bf01273744.

[53] S. I. Rubinow and J. L. Lebowitz. 'A mathematical model of the chemotherapeutic treatment of acute myeloblastic leukemia.' In: *Biophysical journal* vol. 16 (11 1976), pp. 1257–1271. DOI: 10.1016/S0006-3495(76)85772-4.

[54] A. S. Fokas, J. B. Keller and B. D. Clarkson. 'Mathematical model of granulocytopoiesis and chronic myelogenous leukemia.' In: *Cancer research* vol. 51 (8 1991), pp. 2084–2091.

[55] M. Fuentes-Garí et al. 'A mathematical model of subpopulation kinetics for the deconvolution of leukaemia heterogeneity'. In: *Journal of The Royal Society Interface* vol. 12, no. 108 (2015), p. 20150276. DOI: 10.1098/rsif.2015.0276.

[56] J. M. Sarker et al. 'An Integrative multi-lineage model of variation in leukopoiesis and acute myelogenous leukemia'. In: *BMC Systems Biology* vol. 11, no. 1 (2017). DOI: 10.1186/s12918-017-0469-2.

[57] A. L. MacLean, S. Filippi and M. P. H. Stumpf. 'The ecology in the hematopoietic stem cell niche determines the clinical outcome in chronic myeloid leukemia'. In: *Proceedings of the National Academy of Sciences* vol. 111, no. 10 (2014), pp. 3883–3888. DOI: 10.1073/pnas.1317072111.

[58] F. Michor et al. 'Dynamics of chronic myeloid leukaemia'. In: *Nature* vol. 435, no. 7046 (2005), pp. 1267–1270. DOI: 10.1038/nature03669.

[59] J. Foo et al. 'Eradication of Chronic Myeloid Leukemia Stem Cells: A Novel Mathematical Model Predicts No Therapeutic Benefit of Adding G-CSF to Imatinib'. In: *PLoS Computational Biology* vol. 5, no. 9 (2009). Ed. by C. Fraser, e1000503. DOI: 10.1371/journal.pcbi.1000503.

[60] F. Michor, Y. Iwasa and M. A. Nowak. 'The age incidence of chronic myeloid leukemia can be explained by a one-mutation model'. In: *Proceedings of the National Academy of Sciences* vol. 103, no. 40 (2006), pp. 14931–14934. DOI: 10.1073/pnas.0607006103.

[61] C. L. Chaffer and R. A. Weinberg. 'How Does Multistep Tumorigenesis Really Proceed?' In: *Cancer Discovery* vol. 5, no. 1 (2015), pp. 22–24. DOI: 10.1158/2159-8290.cd-14-0788.

[62] J. E. Visvader. 'Cells of origin in cancer'. In: *Nature* vol. 469, no. 7330 (2011), pp. 314–322. DOI: 10.1038/nature09781.

[63] N. L. Komarova and D. Wodarz. 'Drug resistance in cancer: Principles of emergence and prevention'. In: *Proceedings of the National Academy of Sciences* vol. 102, no. 27 (2005), pp. 9714–9719. DOI: 10.1073/pnas.0501870102.

[64] H. Cho et al. 'Modelling acute myeloid leukaemia in a continuum of differentiation states'. In: *Letters in Biomathematics* vol. 5, no. sup1 (2018), S69–S98. DOI: 10.1080/23737867.2018.1472532.

[65] Y. Daniel, Y. Ginosar and Z. Agur. 'The universal properties of stem cells as pinpointed by a simple discrete model'. In: *Journal of Mathematical Biology* vol. 44, no. 1 (2002), pp. 79–86. DOI: 10.1007/s002850100115.

[66] M. Deininger, E. Buchdunger and B. J. Druker. 'The development of imatinib as a therapeutic agent for chronic myeloid leukemia'. In: *Blood* vol. 105, no. 7 (2005), pp. 2640–2653. DOI: 10.1182/blood-2004-08-3097.

[67] D. Dingli and F. Michor. 'Successful Therapy Must Eradicate Cancer Stem Cells'. In: *Stem Cells* vol. 24, no. 12 (2006), pp. 2603–2610. DOI: 10.1634/stemcells.2006-0136.

[68] F. Michor. 'Reply: The long-term response to imatinib treatment of CML'. In: *British Journal of Cancer* vol. 96, no. 4 (2007), pp. 679–680. DOI: 10.1038/sj.bjc.6603604.

[69] P. S. Kim, P. P. Lee and D. Levy. 'Dynamics and Potential Impact of the Immune Response to Chronic Myelogenous Leukemia'. In: *PLoS Computational Biology* vol. 4, no. 6 (2008). Ed. by R. J. D. Boer, e1000095. DOI: 10.1371/journal.pcbi.1000095.

[70] D. Paquin et al. 'Strategic Treatment Interruptions During Imatinib Treatment of Chronic Myelogenous Leukemia'. In: *Bulletin of Mathematical Biology* vol. 73, no. 5 (2010), pp. 1082–1100. DOI: 10.1007/s11538-010-9553-0.

[71] A. Olshen et al. 'Dynamics of chronic myeloid leukemia response to dasatinib, nilotinib, and high-dose imatinib'. In: *Haematologica* vol. 99, no. 11 (2014), pp. 1701–1709. DOI: 10.3324/haematol.2013.085977.

[72] M. Helal et al. 'Analysis of mathematical model of leukemia'. In: *ITM Web of Conferences* vol. 4 (2015). Ed. by A. Lakmeche and V. Volpert, p. 01005. DOI: 10.1051/itmconf/20150401005.

[73] B. Aïnseba and C. Benosman. 'Optimal control for resistance and suboptimal response in CML'. In: *Mathematical biosciences* vol. 227, no. 2 (2010), pp. 81–93. DOI: 10.1016/j.mbs.2010.06.005.

[74] N. Komarova. 'Mathematical modeling of cyclic treatments of chronic myeloid leukemia'. In: *Mathematical Biosciences and Engineering* vol. 8, no. 2 (2011), pp. 289–306. DOI: 10.3934/mbe.2011.8.289.

[75] H. Moore and N. K. Li. 'A mathematical model for chronic myelogenous leukemia (CML) and T cell interaction'. In: *Journal of Theoretical Biology* vol. 227, no. 4 (2004), pp. 513–523. DOI: 10.1016/j.jtbi.2003.11.024.

[76] L. Berezansky, S. Bunimovich-Mendrazitsky and A. Domoshnitsky. 'A mathematical model of Imatinib and Interferon-alpha combined treatment of chronic myeloid leukemia'. In: *Functional Differential Equations* vol. 19, no. 3-4 (2012), pp. 257–266.

[77] S. Nanda, H. Moore and S. Lenhart. 'Optimal control of treatment in a mathematical model of chronic myelogenous leukemia'. In: *Mathematical Biosciences* vol. 210, no. 1 (2007), pp. 143–156. DOI: 10.1016/j.mbs.2007.05.003.

[78] I. Roeder and M. Loeffler. 'A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity'. In: *Experimental Hematology* vol. 30, no. 8 (2002), pp. 853–861. DOI: 10.1016/s0301-472x(02)00832-9.

[79] I. Roeder et al. 'Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications'. In: *Nature Medicine* vol. 12, no. 10 (2006), pp. 1181–1184. DOI: 10.1038/nm1487.

[80] I. Roeder, M. Herberg and M. Horn. 'An Age Structured Model of Hematopoietic Stem Cell Organization with Application to Chronic Myeloid Leukemia'. In: *Bulletin of Mathematical Biology* vol. 71, no. 3 (2008), pp. 602–626. DOI: 10.1007/s11538-008-9373-7.

[81] P. S. Kim, P. P. Lee and D. Levy. 'A PDE Model for Imatinib-Treated Chronic Myelogenous Leukemia'. In: *Bulletin of Mathematical Biology* vol. 70, no. 7 (2008), pp. 1994–2016. DOI: 10.1007/s11538-008-9336-z.

[82] G. D. Clapp and D. Levy. 'Incorporating asymmetric stem cell division into the Roeder model for chronic myeloid leukemia'. In: *Mathematical models of tumor-immune system dynamics*. Vol. 107. Springer Proc. Math. Stat. 2014, pp. 1–20. DOI: 10.1007/978-1-4939-1793-8_1.

[83] M. Doumic-Jauffret, P. S. Kim and B. Perthame. 'Stability Analysis of a Simplified Yet Complete Model for Chronic Myelogenous Leukemia'. In: *Bulletin of Mathematical Biology* vol. 72, no. 7 (2010), pp. 1732–1759. DOI: 10.1007/s11538-009-9500-0.

[84] E. Pefani et al. 'Chemotherapy Drug Scheduling for the Induction Treatment of Patients With Acute Myeloid Leukemia'. In: *IEEE Transactions on Biomedical Engineering* vol. 61, no. 7 (2014), pp. 2049–2056. DOI: 10.1109/tbme.2014.2313226.

[85] F. Jost et al. 'Model-based optimal AML consolidation treatment'. In: *arXiv preprint arXiv:1911.08980* (2019).

[86] S. L. Cooper and P. A. Brown. 'Treatment of pediatric acute lymphoblastic leukemia'. In: *Pediatric Clinics* vol. 62, no. 1 (2015), pp. 61–73. DOI: 10.1016/j.pcl.2014.09.006.

[87] F. Lang, B. Wojcik and M. A. Rieger. 'Stem Cell Hierarchy and Clonal Evolution in Acute Lymphoblastic Leukemia'. In: *Stem Cells International* vol. 2015 (2015), pp. 1–13. DOI: 10.1155/2015/137164.

[88] J. C. Panetta, W. E. Evans and M. H. Cheok. 'Mechanistic mathematical modelling of mercaptopurine effects on cell cycle of human acute lymphoblastic leukaemia cells'. In: *British Journal of Cancer* vol. 94, no. 1 (2006), pp. 93–100. DOI: 10.1038/sj.bjc.6602893.

[89] D. Jayachandran et al. 'Optimal Chemotherapy for Leukemia: A Model-Based Strategy for Individualized Treatment'. In: *PLoS ONE* vol. 9, no. 10 (2014). Ed. by F. Pappalardo, e109623. DOI: 10.1371/journal.pone.0109623.

[90] J. C. Panetta et al. 'Methotrexate Intracellular Disposition in Acute Lymphoblastic Leukemia'. In: *Clinical Cancer Research* vol. 8, no. 7 (2002), pp. 2423–2429.

[91] J. C. Panetta et al. 'Modeling Mechanisms of In Vivo Variability in Methotrexate Accumulation and Folate Pathway Inhibition in Acute Lymphoblastic Leukemia Cells'. In: *PLoS Computational Biology* vol. 6, no. 12 (2010). Ed. by J. A. Papin, e1001019. DOI: 10.1371/journal.pcbi.1001019.

[92] T. T. T. Le et al. 'A mathematical model of white blood cell dynamics during maintenance therapy of childhood acute lymphoblastic leukemia'. In: *Mathematical Medicine and Biology: A Journal of the IMA* (2018). dqy017. DOI: 10.1093/imammb/dqy017.

[93] F. Jost et al. 'Model-based simulation of maintenance therapy of childhood acute lymphoblastic leukemia'. In: *arXiv preprint arXiv:1911.08929* (2019).

[94] G. D. Clapp et al. 'Implication of the autologous immune system in BCR-ABL transcript variations in chronic myelogenous leukemia patients treated with Imatinib'. In: *Cancer research* vol. 75, no. 19 (2015), pp. 4053–4062. DOI: 10.1158/0008-5472.CAN-15-0611.

[95] L. G. DePillis and A. Radunskaya. 'A mathematical tumor model with immune resistance and drug therapy: an optimal control approach'. In: *Computational and Mathematical Methods in Medicine* vol. 3, no. 2 (2001), pp. 79–100. DOI: 10.1080/10273660108833067.

[96] S. Nanda, L. G. dePillis and A. Radunskaya. 'B cell chronic lymphocytic leukemia - A model with immune response'. In: *Discrete and Continuous Dynamical Systems - Series B* vol. 18, no. 4 (2013), pp. 1053–1076. DOI: 10.3934/dcdsb.2013.18.1053.

[97] D. S. Rodrigues et al. 'A mathematical model for chemoimmunotherapy of chronic lymphocytic leukemia'. In: *Applied Mathematics and Computation* vol. 349 (2019), pp. 118–133. DOI: https://doi.org/10.1016/j.amc.2018.12.008.

[98] Y. Nishiyama and N. NiShiyama. 'Modeling immunotherapy and outcomes in acute myeloid leukemia'. In: *The Science Reports of Kanazawa University* vol. 61 (2017), pp. 25–38.

[99] Y. Nishiyama, Y. Saikawa and N. Nishiyama. 'Interaction between the immune system and acute myeloid leukemia: A model incorporating promotion of regulatory T cell expansion by leukemic cells'. In: *Biosystems* vol. 165 (2018), pp. 99–105. DOI: 10.1016/j.biosystems.2018.01.006.

[100] D. Kirschner and J. C. Panetta. 'Modeling immunotherapy of the tumor - immune interaction'. In: *Journal of Mathematical Biology* vol. 37, no. 3 (1998), pp. 235–252. DOI: 10.1007/s002850050127.

[101] A. Talkington, C. Dantoin and R. Durrett. 'Ordinary Differential Equation Models for Adoptive Immunotherapy'. In: *Bulletin of Mathematical Biology* vol. 80, no. 5 (2017), pp. 1059–1083. DOI: 10.1007/s11538-017-0263-8.

[102] A. Cappuccio, M. Elishmereni and Z. Agur. 'Cancer Immunotherapy by Interleukin-21: Potential Treatment Strategies Evaluated in a Mathematical Model'. In: *Cancer Research* vol. 66, no. 14 (2006), pp. 7293–7300. DOI: 10.1158/0008-5472.can-06-0241.

[103] L. Preziosi, ed. *Cancer Modelling and Simulation (Chapman & Hall/CRC Mathematical and Computational Biology)*. 2003. DOI: 10 . 1201 / 9780203494899.

[104] M. Kolev. 'A mathematical model of cellular immune response to leukemia'. In: *Mathematical and Computer Modelling* vol. 41, no. 10 (2005), pp. 1071–1081. DOI: 10.1016/j.mcm.2005.05.003.

[105] J. M. Chrobak, M. Bodnar and H. Herrero. 'About a generalized model of lymphoma'. In: *Journal of Mathematical Analysis and Applications* vol. 386, no. 2 (2012), pp. 813–829. DOI: 10.1016/j.jmaa.2011.08.043.

[106] A. Saadatpour et al. 'Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia'. In: *PLoS computational biology* vol. 7, no. 11 (2011), e1002267. DOI: 10.1371/journal.pcbi.1002267.

[107] V. M. Pérez-García et al. 'CAR T cells for T-cell leukemias: Insights from mathematical models'. In: *arXiv preprint arXiv:2004.14291* (2020).

[108] O. León-Triana et al. 'CAR T cell therapy in B-cell acute lymphoblastic leukaemia: Insights from mathematical models'. In: *Communications in Nonlinear Science and Numerical Simulation* vol. 94 (2020), p. 105570. DOI: 10.1016/j.cnsns.2020.105570.

[109] M. S. Khatun and M. H. A. Biswas. 'Modeling the effect of adoptive T cell therapy for the treatment of leukemia'. In: *Computational and Mathematical Methods* vol. 2, no. 2 (2020), e1069. DOI: 10.1002/cmm4.1069.

[110] R. Mostolizadeh, Z. Afsharnezhad and A. Marciniak-Czochra. 'Mathematical model of chimeric anti-gene receptor (CAR) T cell therapy with presence of cytokine'. In: *Numerical Algebra, Control & Optimization* vol. 8, no. 1 (2018), p. 63.

[111] A. M. Stein et al. 'Tisagenlecleucel Model-Based Cellular Kinetic Analysis of Chimeric Antigen Receptor–T Cells'. In: *CPT: Pharmacometrics & Systems Pharmacology* vol. 8, no. 5 (2019), pp. 285–295. DOI: 10.1002/psp4.12388.

[112] L. R. C. Barros, B. de Jesus Rodrigues and R. C. Almeida. 'CAR-T cell Goes on a Mathematical Model'. In: *Journal of Cellular Immunology* vol. 2, no. 1 (2020).

[113] S. Hanson et al. 'Toxicity Management in CAR T cell therapy for B-ALL: Mathematical modelling as a new avenue for improvement'. In: *bioRxiv* (2016). DOI: 10.1101/049908.

[114] D. Hardiansyah and C. M. Ng. 'Quantitative Systems Pharmacology Model of Chimeric Antigen Receptor T-Cell Therapy'. In: *Clinical and Translational Science* vol. 12, no. 4 (2019), pp. 343–349. DOI: 10.1111/cts.12636.

[115] A. A. Toor et al. 'A dynamical systems perspective on chimeric antigen receptor T-cell dosing'. In: *Bone Marrow Transplantation* vol. 54, no. 3 (2019), pp. 485–489.

[116] A. S. Bratus, A. S. Goncharov and I. T. Todorov. 'Optimal control in a mathematical model for leukemia therapy with phase constraints'. In: *Moscow University Computational Mathematics and Cybernetics* vol. 36, no. 4 (2012), pp. 178–182. DOI: 10.3103/S0278641912040024.

[117] Y. Todorov and F. Nuernberg. 'Optimal therapy protocols in the mathematical model of acute leukemia with several phase constraints'. In: *Optimal Control Applications and Methods* vol. 35, no. 5 (2014), pp. 559–574. DOI: 10.1002/oca.2087.

[118] C. L. Mouser et al. 'A model of hematopoietic stem cell proliferation under the influence of a chemotherapeutic agent in combination with a hematopoietic inducing agent'. In: *Theoretical Biology and Medical Modelling* vol. 11, no. 1 (2014), p. 4. DOI: 10.1186/1742-4682-11-4.

[119] S. Chulián et al. 'Mathematical models of Leukaemia and its treatment: A review.' arXiv, submitted to SEMA Journal. 2020. eprint: https://arxiv.org/abs/2011.05881.

[120] S. Doulatov et al. 'Hematopoiesis: a human perspective'. In: *Cell stem cell* vol. 10, no. 2 (2012), pp. 120–136. DOI: 10.1016/j.stem.2012.01.006.

[121] B. Alberts et al. *Molecular Biology of the Cell*. 4th. 2002. DOI: 10.1016/0307-4412(94)90059-0.

[122] T. W. LeBien and T. F. Tedder. 'B lymphocytes: how they develop and function'. In: *Blood* vol. 112, no. 5 (2008), pp. 1570–1580. DOI: 10.1182/blood-2008-02-078071.

[123] E. Steliarova-Foucher et al. 'International incidence of childhood cancer, 2001–10: a population-based registry study'. In: *The Lancet Oncology* vol. 18, no. 6 (2017), pp. 719–731. DOI: 10.1016/S1470-2045(17)30186-9.

[124] Pujo-Menjouet, L. 'Blood Cell Dynamics: Half of a Century of Modelling'. In: *Math. Model. Nat. Phenom.* vol. 11, no. 1 (2016), pp. 92–115. DOI: 10.1051/mmnp/201611106.

[125] A. Marciniak-Czochra, T. Stiehl and W. Wagner. 'Modeling of replicative senescence in hematopoietic development'. In: *Aging* vol. 1, no. 8 (2009), pp. 723–732. DOI: 10.18632/aging.100072.

[126] I. Roeder. 'Quantitative stem cell biology: computational studies in the hematopoietic system'. In: *Current Opinion in Hematology* vol. 13, no. 4 (2006), pp. 222–228. DOI: 10.1097/01.moh.0000231418.08031.48.

[127] S. Viswanathan and P. W. Zandstra. 'Towards predictive models of stem cell fate'. In: *Cytotechnology* vol. 41, no. 2/3 (2003), pp. 75–92. DOI: 10.1023/a:1024866504538.

[128] V. V. Ganusov and J. Auerbach. 'Mathematical Modeling Reveals Kinetics of Lymphocyte Recirculation in the Whole Organism'. In: *PLoS Computational Biology* vol. 10, no. 5 (2014). Ed. by R. Antia, e1003586. DOI: 10.1371/journal.pcbi.1003586.

[129] T. Stiehl and A. Marciniak-Czochra. 'Characterization of stem cells using mathematical models of multistage cell lineages'. In: *Mathematical and Computer Modelling* vol. 53, no. 7-8 (2011), pp. 1505–1517. DOI: 10.1016/j.mcm.2010.03.057.

[130] M. C. Mackey and R. Rudnicki. 'Global stability in a delayed partial differential equation describing cellular replication'. In: *Journal of Mathematical Biology* vol. 33, no. 1 (1994), pp. 89–109. DOI: 10.1007/bf00160175.

[131] D. Dingli and J. M. Pacheco. 'Modeling the architecture and dynamics of hematopoiesis'. In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* vol. 2, no. 2 (2009), pp. 235–244. DOI: 10.1002/wsbm.56.

[132] G. Shahaf et al. 'B Cell Development in the Bone Marrow Is Regulated by Homeostatic Feedback Exerted by Mature B Cells'. In: *Frontiers in immunology* vol. 7 (2016), p. 77. DOI: 10.3389/fimmu.2016.00077.

[133] S. Hu, O. A. Smirnova and F. A. Cucinotta. 'A biomathematical model of lymphopoiesis following severe radiation accidents-potential use for dose assessment'. In: *Health physics* vol. 102, no. 4 (2012), pp. 425–436. DOI: 10.1097/HP.0b013e318240593d.

[134] N. L. Komarova. 'Principles of regulation of self-renewing cell lineages'. In: *PloS one* vol. 8, no. 9 (2013). DOI: 10.1371/journal.pone.0072847.

[135] E. Manesso et al. 'Dynamical modelling of haematopoiesis: an integrated view over the system in homeostasis and under perturbation'. In: *Journal of the Royal Society Interface* vol. 10, no. 80 (2013), p. 20120817. DOI: 10.1098/rsif.2012.0817.

[136] D. S. Jones, M. Plank and B. D. Sleeman. *Differential equations and mathematical biology*. 2009. DOI: 10.1201/9781420083583.

[137] K. Murphy and C. Weaver. *Janeway's immunobiology*. 2016. DOI: 10.1201/9781315533247.

[138] T. W. LeBien. 'Fates of human B-cell precursors'. In: *Blood, The Journal of the American Society of Hematology* vol. 96, no. 1 (2000), pp. 9–23. DOI: 10.1182/blood.V96.1.9.

[139] D. H. Fuertinger et al. 'A model of erythropoiesis in adults with sufficient iron availability'. In: *Journal of mathematical biology* vol. 66, no. 6 (2013), pp. 1209–1240. DOI: 10.1007/s00285-012-0530-0.

[140] T. Walenda et al. 'Feedback signals in myelodysplastic syndromes: increased self-renewal of the malignant clone suppresses normal hematopoiesis'. In: *PLoS computational biology* vol. 10, no. 4 (2014). DOI: 10.1371/journal.pcbi.1003599.

[141] W. Wang et al. 'Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia'. In: *Haematologica* vol. 102, no. 9 (2017), pp. 1567–1577. DOI: 10.3324/haematol.2016.163584.

[142] F. Knauer, T. Stiehl and A. Marciniak-Czochra. 'Oscillations in a white blood cell production model with multiple differentiation stages'. In: *Journal of Mathematical Biology* vol. 80, no. 3 (2020), pp. 575–600. DOI: 10.1007/s00285-019-01432-6.

[143] H. Jumaa, R. W. Hendriks and M. Reth. 'B cell signaling and tumorigenesis'. In: *Annu. Rev. Immunol.* vol. 23 (2005), pp. 415–445. DOI: 10.1146/annurev.immunol.23.021704.115606.

[144] R. Mehr et al. 'Asynchronous differentiation models explain bone marrow labeling kinetics and predict reflux between the pre-and immature B cell pools'. In: *International immunology* vol. 15, no. 3 (2003), pp. 301–312. DOI: 10.1093/intimm/dxg025.

[145] O. A. Smirnova, S. Hu and F. A. Cucinotta. 'Analysis of the lymphocytopoiesis dynamics in nonirradiated and irradiated humans: a modeling approach'. In: *Radiation Research* vol. 181, no. 3 (2014), pp. 240–250. DOI: 10.1667/RR13256.1.

[146] R. Mostolizadeh, Z. Afsharnezhad and A. Marciniak-Czochra. 'Mathematical model of chimeric anti-gene receptor (CAR) T cell therapy with presence of cytokine'. In: *Numerical Algebra, Control & Optimization* vol. 8, no. 1 (2018), p. 63. DOI: 10.3934/naco.2018004.

[147] J. V. Dongen et al. 'EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes'. In: *Leukemia* vol. 26, no. 9 (2012), pp. 1908–1975. DOI: 10.1038/leu.2012.120.

[148] P. Lúcio et al. 'Flow cytometric analysis of normal B cell differentiation: a frame of reference for the detection of minimal residual disease in precursor-B-ALL'. In: *Leukemia* vol. 13, no. 3 (1999), pp. 419–427. DOI: 10.1038/sj.leu.2401279.

[149] H. Kraus et al. 'A Feeder-Free Differentiation System Identifies Autonomously Proliferating B Cell Precursors in Human Bone Marrow'. In: *The Journal of Immunology* vol. 192, no. 3 (2014), pp. 1044–1054. DOI: 10.4049/jimmunol.1301815.

[150] R. Maddaly et al. 'Receptors and signaling mechanisms for B-lymphocyte activation, proliferation and differentiation–insights from both in vivo and in vitro approaches'. In: *FEBS letters* vol. 584, no. 24 (2010), pp. 4883–4894. DOI: 10.1016/j.febslet.2010.08.022.

[151] J. J. O'Shea, M. Gadina and R. M. Siegel. 'Cytokines and cytokine receptors'. In: *Clinical immunology*. 2019, pp. 127–155. DOI: 10.1016/B978-0-7020-6896-6.00009-0.

[152] G. Petkau and M. Turner. 'Signalling circuits that direct early B-cell development'. In: *Biochemical Journal* vol. 476, no. 5 (2019), pp. 769–778. DOI: 10.1042/BCJ20180565.

[153] A. Marciniak-Czochra, A. Mikelić and T. Stiehl. 'Renormalization group second-order approximation for singularly perturbed nonlinear ordinary differential equations'. In: *Mathematical Methods in the Applied Sciences* vol. 41, no. 14 (2018), pp. 5691–5710. DOI: 10.1002/mma.5107.

[154] Z. Ahmad, N.-U.-R. Durrani and T. Hazir. 'Bone marrow examination in ITP in children: is it mandatory?' In: *Journal of the College of Physicians and Surgeons–Pakistan: JCPSP* vol. 17, no. 6 (2007), pp. 347–349. DOI: 06.2007/jcpsp.347349.

[155] H. Zafar et al. 'Clinical features and outcome in paediatric newly diagnosed immune thrombocytopenic purpura in a tertiary care centre'. In: *Pakistan journal of medical sciences* vol. 34, no. 5 (2018), p. 1195. DOI: doi:10.12669/pjms.345.15687.

[156] G. Finak et al. 'Standardizing flow cytometry immunophenotyping analysis from the Human ImmunoPhenotyping Consortium'. In: *Scientific reports* vol. 6, no. 1 (2016), pp. 1–11. DOI: 10.1038/srep20686.

[157] F. Hahne et al. 'flowCore: a Bioconductor package for high throughput flow cytometry'. In: *BMC bioinformatics* vol. 10, no. 1 (2009), p. 106. DOI: 10.1186/1471-2105-10-106.

[158] Y. Ge and S. C. Sealfon. 'flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding'. In: *Bioinformatics* vol. 28, no. 15 (2012), pp. 2052–2058. DOI: 10.1093/bioinformatics/bts300.

[159] T. Guillaume, D. B. Rubinstein and M. Symann. 'Immune reconstitution and immunotherapy after autologous hematopoietic stem cell transplantation'. In: *Blood, The Journal of the American Society of Hematology* vol. 92, no. 5 (1998), pp. 1471–1490. DOI: 10.1182/blood.V92.5.1471.

[160] D. Leitenberg, J. M. Rappeport and B. R. Smith. 'B-cell precursor bone marrow reconstitution after bone marrow transplantation'. In: *American journal of clinical pathology* vol. 102, no. 2 (1994), pp. 231–236. DOI: 10.1093/ajcp/102.2.231.

[161] J. E. Talmadge et al. 'Rapid immunologic reconstitution following transplantation with mobilized peripheral blood stem cells as compared to bone marrow'. In: *Bone marrow transplantation* vol. 19, no. 2 (1997), pp. 161–172. DOI: 10.1038/sj.bmt.1700626.

[162] A. Parrado et al. 'Repopulation of circulating T, B and NK lymphocytes following bone marrow and blood stem cell transplantation'. In: *Hematology and cell therapy* vol. 39, no. 6 (1997), pp. 301–306. DOI: 10.1007/s00282-997-0301-3.

[163] G. J. Deenen, I. Van Balen and D. Opstelten. 'In rat B lymphocyte genesis sixty percent is lost from the bone marrow at the transition of nondividing pre-B cell to sIgM+ B lymphocyte, the stage of Ig light chain gene expression'. In: *European journal of immunology* vol. 20, no. 3 (1990), pp. 557–564. DOI: 10.1002/eji.1830200315.

[164] Y.-H. Park and D. G. Osmond. 'Dynamics of early B lymphocyte precursor cells in mouse bone marrow: proliferation of cells containing terminal deoxynucleotidyl transferase'. In: *European journal of immunology* vol. 19, no. 11 (1989), pp. 2139–2144. DOI: 10.1002/eji.1830191125.

[165] H. E. Skipper and S. Perry. 'Kinetics of normal and leukemic leukocyte populations and relevance to chemotherapy'. In: *Cancer Research* vol. 30, no. 6 (1970), pp. 1883–1897.

[166] A. K. Abbas, A. H. Lichtman and S. Pillai. *Cellular and molecular immunology*. 1994.

[167] S. . Bendall et al. 'Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development'. In: *Cell* vol. 157, no. 3 (2014), pp. 714–725. DOI: 10.1016/j.cell.2014.04.005.

[168] J. G. Monroe et al. 'Positive and negative selection during B lymphocyte development'. In: *Immunologic research* vol. 27, no. 2-3 (2003), pp. 427–442. DOI: 10.1385/IR:27:2-3:427.

[169] D. G. Osmond. 'Population dynamics of bone marrow B lymphocytes'. In: *Immunological reviews* vol. 93, no. 1 (1986), pp. 103–124. DOI: 10.1111/j.1600-065x.1986.tb01504.x.

[170] P. Clark et al. 'Lymphocyte subsets in normal bone marrow'. In: *Blood* vol. 67, no. 6 (1986), pp. 1600–1606. DOI: 10.1182/blood.V67.6.1600.1600.

[171] C. Andreoni et al. 'Phenotypic analysis of a large number of normal human bone marrow sample by flow cytometry'. In: *Blut* vol. 61, no. 5 (1990), pp. 271–277. DOI: 10.1007/BF01732876.

[172] C. W. Caldwell, E. Poje and M. A. Helikson. 'B-cell precursors in normal pediatric bone marrow'. In: *American journal of clinical pathology* vol. 95, no. 6 (1991), pp. 816–823. DOI: 10.1093/ajcp/95.6.816.

[173] E. M. Rego et al. 'Age-related changes of lymphocyte subsets in normal bone marrow biopsies'. In: *Cytometry: The Journal of the International Society for Analytical Cytology* vol. 34, no. 1 (1998), pp. 22–29. DOI: 10.1002/(SICI)1097-0320(19980215)34:1<22::AID-CYTO4>3.0.CO;2-G.

[174] C. Nombela-Arrieta and M. G. Manz. 'Quantification and three-dimensional microanatomical organization of the bone marrow'. In: *Blood Advances* vol. 1, no. 6 (2017), pp. 407–416. DOI: 10.1182/bloodadvances.2016003194.

[175] Charlotte R. Kleiveland. 'Peripheral Blood Mononuclear Cells'. In: *The Impact of Food Bioactives on Health: in vitro and ex vivo models*. Ed. by K. Verhoeckx et al. Cham, 2015, pp. 161–167. DOI: 10.1007/978-3-319-16104-4_15.

[176] G. Caocci, M. Greco and G. La Nasa. 'Bone marrow homing and engraftment defects of human hematopoietic stem and progenitor cells'. In: *Mediterranean journal of hematology and infectious diseases* vol. 9, no. 1 (2017). DOI: 10.4084/MJHID.2017.032.

[177] H. Kawamoto, H. Wada and Y. Katsura. 'A revised scheme for developmental pathways of hematopoietic cells: the myeloid-based model'. In: *International immunology* vol. 22, no. 2 (2010), pp. 65–70. DOI: 10.1093/intimm/dxp125.

[178] A. Wilson et al. 'Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair'. In: *Cell* vol. 135, no. 6 (2008), pp. 1118–1129. DOI: 10.1016/j.cell.2008.10.048.

[179] L. Biasco et al. 'In vivo tracking of human hematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases'. In: *Cell stem cell* vol. 19, no. 1 (2016), pp. 107–119. DOI: 10.1016/j.stem.2016.04.016.

[180] K. Busch et al. 'Fundamental properties of unperturbed haematopoiesis from stem cells in vivo'. In: *Nature* vol. 518, no. 7540 (2015), pp. 542–546. DOI: 10.1038/nature14242.

[181] R. R. Hardy, P. W. Kincade and K. Dorshkind. 'The protean nature of cells in the B lymphocyte lineage'. In: *Immunity* vol. 26, no. 6 (2007), pp. 703–714. DOI: 10.1016/j.immuni.2007.05.013.

[182] S. Chulián et al. 'Dynamical properties of feedback signalling in B lymphopoiesis: A mathematical modelling approach'. arXiv, submitted to *Journal of Theoretical Biology*. eprint: https://arxiv.org/abs/2007.13526.

[183] A. T. Meadows. 'Book Review Principles and Practice of Pediatric Oncology Fourth edition.' In: *New England Journal of Medicine* vol. 346, no. 24 (2002), pp. 1921–1922. DOI: 10.1056/nejm200206133462423.

[184] T. Terwilliger and M. Abdul-Hay. 'Acute lymphoblastic leukemia: a comprehensive review and 2017 update'. In: *Blood Cancer Journal* vol. 7, no. 6 (2017), e577–e577. DOI: 10.1038/bcj.2017.53.

[185] C.-H. Pui et al. 'Childhood Acute Lymphoblastic Leukemia: Progress Through Collaboration'. In: *Journal of Clinical Oncology* vol. 33, no. 27 (2015), pp. 2938–2948. DOI: 10.1200/jco.2014.59.1636.

[186] D. Bhojwani and C.-H. Pui. 'Relapsed childhood acute lymphoblastic leukaemia'. In: *The Lancet Oncology* vol. 14, no. 6 (2013), e205–e217. DOI: 10.1016/s1470-2045(12)70580-6.

[187] E. G. Weir and M. J. Borowitz. 'Flow cytometry in the diagnosis of acute leukemia'. In: *Seminars in Hematology* vol. 38, no. 2 (2001), pp. 124–138. DOI: 10.1016/s0037-1963(01)90046-0.

[188] E. Lugli, M. Roederer and A. Cossarizza. 'Data analysis in flow cytometry: The future just started'. In: *Cytometry Part A* vol. 77A, no. 7 (2010), pp. 705–713. DOI: 10.1002/cyto.a.20901.

[189] C. E. Pedreira et al. 'Overview of clinical flow cytometry data analysis: recent advances and future challenges'. In: *Trends in Biotechnology* vol. 31, no. 7 (2013), pp. 415–425. DOI: 10.1016/j.tibtech.2013.04.008.

[190] H. Zare et al. 'Automated Analysis of Multidimensional Flow Cytometry Data Improves Diagnostic Accuracy Between Mantle Cell Lymphoma and Small Lymphocytic Lymphoma'. In: *American Journal of Clinical Pathology* vol. 137, no. 1 (2012), pp. 75–85. DOI: 10.1309/ajcpmmlq67yomgew.

[191] S. Chevrier et al. 'An Immune Atlas of Clear Cell Renal Cell Carcinoma'. In: *Cell* vol. 169, no. 4 (2017), 736–749.e18. DOI: 10.1016/j.cell.2017.04.016.

[192] C. Krieg et al. 'High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy'. In: *Nature Medicine* vol. 24, no. 2 (2018), pp. 144–153. DOI: 10.1038/nm.4466.

[193] Y. Lavin et al. 'Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses'. In: *Cell* vol. 169, no. 4 (2017), 750–765.e17. DOI: 10.1016/j.cell.2017.04.014.

[194] M. Reiter et al. 'Automated Flow Cytometric MRD Assessment in Childhood Acute B- Lymphoblastic Leukemia Using Supervised Machine Learning'. In: *Cytometry Part A* vol. 95, no. 9 (2019), pp. 966–975. DOI: 10.1002/cyto.a.23852.

[195] L. Pan et al. 'Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia'. In: *Scientific Reports* vol. 7, no. 1 (2017). DOI: 10.1038/s41598-017-07408-0.

[196] Z. Good et al. 'Single-cell developmental classification of B cell precursor acute lymphoblastic leukemia at diagnosis reveals predictors of relapse'. In: *Nature Medicine* vol. 24, no. 4 (2018), pp. 474–483. DOI: 10.1038/nm.4505.

[197] D. R. Parks, M. Roederer and W. A. Moore. 'A new Logicle display method avoids deceptive effects of logarithmic scaling for low signals and compensated data'. In: *Cytometry Part A* vol. 69A, no. 6 (2006), pp. 541–551. DOI: 10.1002/cyto.a.20258.

[198] C. E. Pedreira et al. 'A Multidimensional Classification Approach for the Automated Analysis of Flow Cytometry Data'. In: *IEEE Transactions on Biomedical Engineering* vol. 55, no. 3 (2008), pp. 1155–1162. DOI: 10.1109/tbme.2008.915729.

[199] G. Lee, W. Finn and C. Scott. 'Statistical file matching of flow cytometry data'. In: *Journal of Biomedical Informatics* vol. 44, no. 4 (2011), pp. 663–676. DOI: 10.1016/j.jbi.2011.03.004.

[200] K. O'Neill et al. 'Deep profiling of multitube flow cytometry data'. In: *Bioinformatics* vol. 31, no. 10 (2015), pp. 1623–1631. DOI: 10.1093/bioinformatics/btv008.

[201] A. L. Pereira et al. 'CytoBackBone: an algorithm for merging of phenotypic information from different cytometric profiles'. In: *Bioinformatics* vol. 35, no. 20 (2019). Ed. by J. Wren, pp. 4187–4189. DOI: 10.1093/bioinformatics/btz212.

[202] L. Balkay. *fca_readfcs*. File Exchange, Mathworks. 2020.

[203] S. Wang et al. 'A feature selection method based on improved fisher's discriminant ratio for text sentiment classification'. In: *Expert Systems with Applications* vol. 38, no. 7 (2011), pp. 8696–8702. DOI: 10.1016/j.eswa.2011.01.077.

[204] N. Radakovich, M. Nagy and A. Nazha. 'Artificial Intelligence in Hematology: Current Challenges and Opportunities'. In: *Current Hematologic Malignancy Reports* vol. 15, no. 3 (2020), pp. 203–210. DOI: 10.1007/s11899-020-00575-4.

[205] N. Mahmood et al. 'Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach'. In: *Medical & Biological Engineering & Computing* vol. 58, no. 11 (2020), pp. 2631–2640. DOI: 10.1007/s11517-020-02245-2.

[206] N. Radakovich, M. Nagy and A. Nazha. 'Machine learning in haematological malignancies'. In: *The Lancet Haematology* vol. 7, no. 7 (2020), e541–e550. DOI: 10.1016/s2352-3026(20)30121-6.

[207] F. Malavasi et al. 'Evolution and Function of the ADP Ribosyl Cyclase/CD38 Gene Family in Physiology and Pathology'. In: *Physiological Reviews* vol. 88, no. 3 (2008), pp. 841–886. DOI: 10.1152/physrev.00035.2007.

[208] S. Deaglio, K. Mehta and F. Malavasi. 'Human CD38: a (r)evolutionary story of enzymes and receptors'. In: *Leukemia Research* vol. 25, no. 1 (2001), pp. 1–12. DOI: 10.1016/s0145-2126(00)00093-x.

[209] S. Ibrahim et al. 'CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia'. In: *Blood* vol. 98, no. 1 (2001), pp. 181–186. DOI: 10.1182/blood.v98.1.181.

[210] A. Keyhani et al. 'Increased CD38 expression is associated with favorable prognosis in adult acute leukemia'. In: *Leukemia Research* vol. 24, no. 2 (2000), pp. 153–159. DOI: 10.1016/s0145-2126(99)00147-2.

[211] J. Naik et al. 'CD38 as a therapeutic target for adult acute myeloid leukemia and T-cell acute lymphoblastic leukemia'. In: *Haematologica* vol. 104, no. 3 (2018), e100–e103. DOI: 10.3324/haematol.2018.192757.

[212] A. A. George et al. 'Detection of leukemic cells in the $CD34^+$ CD38- bone marrow progenitor population in children with acute lymphoblastic leukemia'. In: *Blood* vol. 97, no. 12 (2001), pp. 3925–3930. DOI: 10.1182/blood.v97.12.3925.

[213] Y. Kong et al. '$CD34^+$ $CD38^+$ $CD19^+$ as well as $CD34^+$ CD38-$CD19^+$ cells are leukemia-initiating cells with self-renewal capacity in human B-precursor ALL'. In: *Leukemia* vol. 22, no. 6 (2008), pp. 1207–1213. DOI: 10.1038/leu.2008.83.

[214] D. C. Taussig et al. 'Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells'. In: *Blood* vol. 112, no. 3 (2008), pp. 568–575. DOI: 10.1182/blood-2007-10-118331.

[215] F. Lang et al. 'Plastic CD34 and CD38 expression in adult B-cell precursor acute lymphoblastic leukemia explains ambiguity of leukemia-initiating stem cell populations'. In: *Leukemia* vol. 31, no. 3 (2016), pp. 731–734. DOI: 10.1038/leu.2016.315.

[216] M. Ebinger et al. 'High frequency of immature cells at diagnosis predicts high minimal residual disease level in childhood acute lymphoblastic leukemia'. In: *Leukemia Research* vol. 34, no. 9 (2010), pp. 1139–1142. DOI: 10.1016/j.leukres.2010.03.023.

[217] T. V. Shman, L. V. Movchan and O. V. Aleinikova. 'Frequencies of immature $CD34^+$ CD38 - and $CD34^+$ CD38-$CD19^+$ blasts correlate with minimal residual disease level in pediatric B-cell precursor acute lymphoblastic leukemia'. In: *Leukemia & Lymphoma* vol. 54, no. 11 (2013), pp. 2560–2562. DOI: 10.3109/10428194.2013.778404.

[218] J. Long et al. 'High proportion of $CD34^+$ /CD38-cells is positively correlated with poor prognosis in newly diagnosed childhood acute lymphoblastic leukemia'. In: *Leukemia & Lymphoma* vol. 55, no. 3 (2013), pp. 611–617. DOI: 10.3109/10428194.2013.807924.

[219] Z. Jiang et al. 'CD34 and CD38 are prognostic biomarkers for acute B lymphoblastic leukemia'. In: *Biomarker Research* vol. 4, no. 1 (2016). DOI: 10.1186/s40364-016-0080-5.

[220] T. M. Owaidah et al. 'Expression of CD66c and CD25 in acute lymphoblastic leukemia as a predictor of the presence of BCR/ABL rearrangement'. In: *Hematology/Oncology and Stem Cell Therapy* vol. 1, no. 1 (2008), pp. 34–37. DOI: 10.1016/s1658-3876(08)50058-6.

[221] N. Guillaume et al. 'CD66c expression in B-cell acute lymphoblastic leukemia: strength and weakness'. In: *International Journal of Laboratory Hematology* vol. 33, no. 1 (2010), pp. 92–96. DOI: 10.1111/j.1751-553x.2010.01254.x.

[222] N. Kiyokawa et al. 'Significance of CD66c expression in childhood acute lymphoblastic leukemia'. In: *Leukemia Research* vol. 38, no. 1 (2014), pp. 42–48. DOI: 10.1016/j.leukres.2013.10.008.

[223] E. Mejstříková et al. 'Correlation of CD33 with poorer prognosis in childhood ALL implicates a potential of anti-CD33 frontline therapy'. In: *Leukemia* vol. 19, no. 6 (2005), pp. 1092–1094. DOI: 10.1038/sj.leu.2403737.

[224] U. Testa, E. Pelosi and A. Frankel. 'CD 123 is a membrane biomarker and a therapeutic target in hematologic malignancies'. In: *Biomarker Research* vol. 2, no. 1 (2014). DOI: 10.1186/2050-7771-2-4.

[225] K. Liu et al. 'CD123 and its potential clinical application in leukemias'. In: *Life Sciences* vol. 122 (2015), pp. 59–64. DOI: 10.1016/j.lfs.2014.10.013.

[226] M. Ruella et al. 'Dual CD19 and CD123 targeting prevents antigen-loss relapses after CD19-directed immunotherapies'. In: *Journal of Clinical Investigation* vol. 126, no. 10 (2016), pp. 3814–3826. DOI: 10.1172/jci87366.

[227] E. Angelova et al. 'CD123 expression patterns and selective targeting with a CD123-targeted antibody-drug conjugate (IMGN632) in acute lymphoblastic leukemia'. In: *Haematologica* vol. 104, no. 4 (2018), pp. 749–755. DOI: 10.3324/haematol.2018.205252.

[228] M. Djokic et al. 'Overexpression of CD123 correlates with the hyperdiploid genotype in acute lymphoblastic leukemia'. In: *Haematologica* vol. 94, no. 7 (2009), pp. 1016–1019. DOI: 10.3324/haematol.2008.000299.

[229] S. Chulián et al. 'High-dimensional analysis of single-cell flow cytometry data predicts relapse in Childhood Acute Lymphoblastic Leukemia'. Preprints, submitted to Cancers. eprint: https://www.preprints.org/manuscript/202010.0557/v1.

[230] I. Dagogo-Jack and A. T. Shaw. 'Tumour heterogeneity and resistance to cancer therapies'. In: *Nature Reviews Clinical Oncology* vol. 15, no. 2 (2017), pp. 81–94. DOI: 10.1038/nrclinonc.2017.166.

[231] P. Charoentong et al. 'Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade'. In: *Cell Reports* vol. 18, no. 1 (2017), pp. 248–262. DOI: 10.1016/j.celrep.2016.12.019.

[232] M. Ghandi et al. 'Next-generation characterization of the Cancer Cell Line Encyclopedia'. In: *Nature* vol. 569, no. 7757 (2019), pp. 503–508. DOI: 10.1038/s41586-019-1186-3.

[233] L. Wasserman. 'Topological Data Analysis'. In: *Annual Review of Statistics and Its Application* vol. 5, no. 1 (2018), pp. 501–532. DOI: 10.1146/annurev-statistics-031017-100045.

[234] N. Otter et al. 'A roadmap for the computation of persistent homology'. In: *EPJ Data Science* vol. 6, no. 1 (2017). DOI: 10.1140/epjds/s13688-017-0109-5.

[235] B. J. Stolz et al. 'Geometric anomaly detection in data'. In: *Proceedings of the National Academy of Sciences* vol. 117, no. 33 (2020), pp. 19664–19669. DOI: 10.1073/pnas.2001741117.

[236] C. Curto. 'What can topology tell us about the neural code?' In: *Bulletin of the American Mathematical Society* vol. 54, no. 1 (2016), pp. 63–78. DOI: 10.1090/bull/1554.

[237] Z. Meng et al. 'Weighted persistent homology for biomolecular data analysis'. In: *Scientific Reports* vol. 10, no. 1 (2020). DOI: 10.1038/s41598-019-55660-3.

[238] J. Gamble and G. Heo. 'Exploring uses of persistent homology for statistical analysis of landmark-based shape data'. In: *Journal of Multivariate Analysis* vol. 101, no. 9 (2010), pp. 2184–2199. DOI: 10.1016/j.jmva.2010.04.016.

[239] L. Kanari et al. 'A Topological Representation of Branching Neuronal Morphologies'. In: *Neuroinformatics* vol. 16, no. 1 (2017), pp. 3–13. DOI: 10.1007/s12021-017-9341-1.

[240] M. E. Aktas, E. Akbas and A. E. Fatmaoui. 'Persistence homology of networks: methods and applications'. In: *Applied Network Science* vol. 4, no. 1 (2019). DOI: 10.1007/s41109-019-0179-3.

[241] K. Xia. 'Persistent homology analysis of ion aggregations and hydrogen-bonding networks'. In: *Physical Chemistry Chemical Physics* vol. 20, no. 19 (2018), pp. 13448–13460. DOI: 10.1039/c8cp01552j.

[242] M. Nicolau, A. J. Levine and G. Carlsson. 'Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival'. In: *Proceedings of the National Academy of Sciences* vol. 108, no. 17 (2011), pp. 7265–7270. DOI: 10.1073/pnas.1102826108.

[243] T. Lakshmikanth et al. 'Mass Cytometry and Topological Data Analysis Reveal Immune Parameters Associated with Complications after Allogeneic Stem Cell Transplantation'. In: *Cell Reports* vol. 20, no. 9 (2017), pp. 2238–2250. DOI: 10.1016/j.celrep.2017.08.021.

[244] L. Crawford et al. 'Predicting Clinical Outcomes in Glioblastoma: An Application of Topological and Functional Data Analysis'. In: *Journal of the American Statistical Association* vol. 115, no. 531 (2019), pp. 1139–1150. DOI: 10.1080/01621459.2019.1671198.

[245] P. Lawson et al. 'Persistent Homology for the Quantitative Evaluation of Architectural Features in Prostate Cancer Histology'. In: *Scientific Reports* vol. 9, no. 1 (2019). DOI: 10.1038/s41598-018-36798-y.

[246] R. Rabadán et al. 'Identification of relevant genetic alterations in cancer using topological data analysis'. In: *Nature Communications* vol. 11, no. 1 (2020). DOI: 10.1038/s41467-020-17659-7.

[247] Z. J. Ward et al. 'Estimating the total incidence of global childhood cancer: a simulation-based analysis'. In: *The Lancet Oncology* vol. 20, no. 4 (2019), pp. 483–493. DOI: 10.1016/s1470-2045(18)30909-4.

[248] A. Kakaje et al. 'Rates and trends of childhood acute lymphoblastic leukaemia: an epidemiology study'. In: *Scientific Reports* vol. 10, no. 1 (2020). DOI: 10.1038/s41598-020-63528-0.

[249] J. Stentoft. 'Leukaemia Diagnosis, fifth edition. Barbara J.Bain, Wiley Blackwell,2017. ISBN 9781119210542 (cloth, 559 pages). Also available as e-book.' In: *British Journal of Haematology* vol. 188, no. 1 (2019), pp. 169–169. DOI: 10.1111/bjh.16351.

[250] F. E. Craig and K. A. Foon. 'Flow cytometric immunophenotyping for hematologic neoplasms'. In: *Blood* vol. 111, no. 8 (2008), pp. 3941–3967. DOI: 10.1182/blood-2007-11-120535.

[251] V. D. Silva and G. Carlsson. *Topological estimation using witness complexes*. en. 2004. DOI: 10.2312/SPBG/SPBG04/157-166.

[252] H. Adams et al. 'Persistence Images: A Stable Vector Representation of Persistent Homology'. In: *Journal of Machine Learning Research* vol. 18, no. 8 (2017), pp. 1–35.

[253] L. Breiman. 'Random Forests'. In: *Machine Learning* vol. 45, no. 1 (2001), pp. 5–32. DOI: 10.1023/a:1010933404324.

[254] W. S. Noble. 'What is a support vector machine?' In: *Nature Biotechnology* vol. 24, no. 12 (2006), pp. 1565–1567. DOI: 10.1038/nbt1206-1565.

[255] C. Tralie, N. Saul and R. Bar-On. 'Ripser.py: A Lean Persistent Homology Library for Python'. In: *Journal of Open Source Software* vol. 3, no. 29 (2018), p. 925. DOI: 10.21105/joss.00925.

[256] P. Theunissen et al. 'Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia'. In: *Blood* vol. 129, no. 3 (2017), pp. 347–357. DOI: 10.1182/blood-2016-07-726307.

[257] S. Burgler. 'Role of CD38 Expression in Diagnosis and Pathogenesis of Chronic Lymphocytic Leukemia and Its Potential as Therapeutic Target'. In: *Critical Reviews in Immunology* vol. 35, no. 5 (2015), pp. 417–432. DOI: 10.1615/critrevimmunol.v35.i5.50.

[258] V. F. Azad et al. 'CD7 aberrant expression led to a lineage switch at relapsed childhood acute pre-B lymphoblastic leukemia'. In: *Medical Molecular Morphology* vol. 49, no. 1 (2015), pp. 53–56. DOI: 10.1007/s00795-015-0117-0.

[259] R. A. Gatenby, J. Zhang and J. S. Brown. 'First Strike-Second Strike Strategies in Metastatic Cancer: Lessons from the Evolutionary Dynamics of Extinction'. In: *Cancer Research* vol. 79, no. 13 (2019), pp. 3174–3177. DOI: 10.1158/0008-5472.can-19-0807.

[260] R. A. Gatenby et al. 'Eradicating Metastatic Cancer and the Eco-Evolutionary Dynamics of Anthropocene Extinctions'. In: *Cancer Research* vol. 80, no. 3 (2019), pp. 613–623. DOI: 10.1158/0008-5472.can-19-1941.

[261] J. K. Lin et al. 'Cost Effectiveness of Chimeric Antigen Receptor T-Cell Therapy in Relapsed or Refractory Pediatric B-Cell Acute Lymphoblastic Leukemia'. In: *Journal of Clinical Oncology* vol. 36, no. 32 (2018), pp. 3192–3202. DOI: 10.1200/jco.2018.79.0642.

[262] A. V. Moorman et al. 'A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia'. In: *Blood* vol. 124, no. 9 (2014), pp. 1434–1444. DOI: 10.1182/blood-2014-03-562918.

[263] A. T. Sahlol, A. M. Abdeldaim and A. E. Hassanien. 'Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm'. In: *Soft Computing* (2018). DOI: 10.1007/s00500-018-3288-5.

[264] S. Chulián et al. 'The shape of cancer relapse: Topological data analysis predicts recurrence in paediatric acute lymphoblastic leukaemia'. Preprint.

[265] R. A. Fisher. 'The wave of advance of advantageous genes'. In: *Annals of Eugenics* vol. 7, no. 4 (1937), pp. 355–369. DOI: 10.1111/j.1469-1809.1937.tb02153.x.

[266] N. P. A.N. Kolmogorov I.G. Petrovskii. 'A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem'. In: *Bulletin of Moscow University, Series International* vol. A, no. 1 (1937), pp. 1–25.

[267] J. D. Murray, ed. *Mathematical Biology*. Springer New York, 2002. DOI: 10.1007/b98868.

[268] V. Volpert and S. Petrovskii. 'Reaction-diffusion waves in biology'. In: *Physics of Life Reviews* vol. 6, no. 4 (2009), pp. 267–310. DOI: 10.1016/j.plrev.2009.10.002.

[269] N. Britton. 'Aggregation and the competitive exclusion principle'. In: *Journal of Theoretical Biology* vol. 136, no. 1 (1989), pp. 57–66. DOI: 10.1016/s0022-5193(89)80189-4.

[270] E. E. Holmes et al. 'Partial Differential Equations in Ecology: Spatial Interactions and Population Dynamics'. In: *Ecology* vol. 75, no. 1 (1994), pp. 17–29. DOI: 10.2307/1939378.

[271] H. Malchow, S. V. Petrovskii and E. Venturino. *Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulation [With CD (Audio)]*. Champan & Hall, 2007.

[272] M. A. Lewis, S. V. Petrovskii and J. R. Potts. 'Dynamics of Biological Invasions'. In: *Interdisciplinary Applied Mathematics*. Springer International Publishing, 2016, pp. 19–68. DOI: 10.1007/978-3-319-32043-4_2.

[273] K. R. Swanson et al. 'Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion'. In: *Journal of the Neurological Sciences* vol. 216, no. 1 (2003), pp. 1–10. DOI: 10.1016/j.jns.2003.06.001.

[274] V. M. Pérez-García et al. 'Bright solitary waves in malignant gliomas'. In: *Physical Review E* vol. 84, no. 2 (2011). DOI: 10.1103/physreve.84.021921.

[275] J. Belmonte-Beitia, G. F. Calvo and V. M. Pérez-García. 'Effective particle methods for Fisher-Kolmogorov equations: Theory and applications to brain tumor dynamics'. In: *Communications in Nonlinear Science and Numerical Simulation* vol. 19, no. 9 (2014), pp. 3267–3283. DOI: 10.1016/j.cnsns.2014.02.004.

[276] J. F. Douglas et al. 'Propagating waves of self-assembly in organosilane monolayers'. In: *Proceedings of the National Academy of Sciences* vol. 104, no. 25 (2007), pp. 10324–10329. DOI: 10.1073/pnas.0703620104.

[277] I. R. Epstein, J. A. Pojman and G. Nicolis. 'An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos'. In: *Physics Today* vol. 52, no. 11 (1999), pp. 68–68. DOI: 10.1063/1.882734.

[278] P. Grindrod. *Patterns and waves : the theory and applications of reaction-diffusion equations*. Oxford New York: Clarendon Press Oxford University Press, 1991.

[279] G. de Vries, M. Lewis and T. Hillen. *A Course in Mathematical Biology: A Quantitative Modeling with Mathematical and Computational Methods*. Cambridge, 2006.

[280] N. Shigesada. *Biological Invasions*. OUP Oxford, 1997.

[281] J. Belmonte-Beitia et al. 'Modelling biological invasions: Individual to population scales at interfaces'. In: *Journal of Theoretical Biology* vol. 334 (2013), pp. 1–12. DOI: 10.1016/j.jtbi.2013.05.033.

[282] P. Olver. *Applications of Lie Groups to Differential Equations*. New York, NY: Springer US, 1986.

[283] P. A. Clarksonz and E. L. Mansfield. 'Symmetry reductions and exact solutions of a class of nonlinear heat equations'. In: *Physica D: Nonlinear Phenomena* vol. 70, no. 3 (1994), pp. 250–288. DOI: 10.1016/0167-2789(94)90017-5.

[284] T. E. Mogorosi et al. 'Group analysis of a hyperbolic Lane–Emden system'. In: *Applied Mathematics and Computation* vol. 292 (2017), pp. 156–164. DOI: 10.1016/j.amc.2016.07.033.

[285] W. Chao-Zhong. 'From additional symmetries to linearization of Virasoro symmetries'. In: *Physica D: Nonlinear Phenomena* vol. 249 (2013), pp. 25–37. DOI: 10.1016/j.physd.2013.01.005.

[286] J. Romero, M. Gandarias and E. Medina. 'Symmetries, periodic plane waves and blow-up of lambda-omega systems'. In: *Physica D: Nonlinear Phenomena* vol. 147, no. 3-4 (2000), pp. 259–272. DOI: 10.1016/s0167-2789(00)00161-5.

[287] R. de la Rosa and M. Bruzón. 'On the classical and nonclassical symmetries of a generalized Gardner equation'. In: *Applied Mathematics and Nonlinear Sciences* vol. 1, no. 1 (2016), pp. 263–272. DOI: 10.21042/amns.2016.1.00021.

[288] M. Rosa and M. L. Gandarias. 'Multiplier method and exact solutions for a density dependent reaction-diffusion equation'. In: *Applied Mathematics and Nonlinear Sciences* vol. 1, no. 2 (2016), pp. 311–320. DOI: 10.21042/amns.2016.2.00026.

[289] M. Gandarias, M. Bruzón and M. Rosa. 'Nonlinear self-adjointness and conservation laws for a generalized Fisher equation'. In: *Communications in Nonlinear Science and Numerical Simulation* vol. 18, no. 7 (2013), pp. 1600–1606. DOI: 10.1016/j.cnsns.2012.11.023.

[290] R. Cherniha and J. R. King. 'Lie symmetries and conservation laws of non-linear multidimensional reaction–diffusion systems with variable diffusivities'. In: *IMA Journal of Applied Mathematics* vol. 71, no. 3 (2006), pp. 391–408. DOI: 10.1093/imamat/hxh103.

[291] S. Lie. 'Lie group analysis : classical heritage'. In: ed. by N. Ibragimov. ALGA Publications, 2004. Chap. Integration of a class of linearpartial differential equations by means of definite integrals, pp. 65–100.

[292] L. V. Ovsiannikov. *Group analysis of differential equations*. Academic Press, 1982.

[293] V. Dorodnitsyn. 'On invariant solutions of the equation of non-linear heat conduction with a source'. In: *USSR Computational Mathematics and Mathematical Physics* vol. 22, no. 6 (1982), pp. 115–122. DOI: 10.1016/0041-5553(82)90102-1.

[294] R. Cherniha and M. Serov. 'Lie and non-Lie symmetries of nonlinear diffusion equations with convection term'. In: *Symmetry in Nonlinear Mathematical Physics* vol. 2 (1997), pp. 444–449.

[295] T. Antal and P. L. Krapivsky. 'Exact solution of a two-type branching process: models of tumor progression'. In: *Journal of Statistical Mechanics: Theory and Experiment* vol. 2011, no. 08 (2011), P08018. DOI: 10.1088/1742-5468/2011/08/p08018.

[296] H. Greenspan. 'On the growth and stability of cell cultures and solid tumors'. In: *Journal of Theoretical Biology* vol. 56, no. 1 (1976), pp. 229–242. DOI: 10.1016/s0022-5193(76)80054-9.

[297] A. Griffiths. *Modern genetic analysis*. New York: W.H. Freeman, 1999.

[298] M. Rosa, M. S. Bruzón and M. L. Gandarias. 'A conservation law for a generalized chemical Fisher equation'. In: *Journal of Mathematical Chemistry* vol. 53, no. 3 (2014), pp. 941–948. DOI: 10.1007/s10910-014-0451-9.

[299] M. Rosa, M. S. Bruzón and M. L. Gandarias. 'Lie symmetry analysis and conservation laws for a Fisher equation with variable coefficients'. In: *Applied Mathematics & Information Sciences* vol. 9, no. 6 (2015), p. 2783.

[300] M. Rosa et al. 'Classical and potential symmetries for a generalized Fisher equation'. In: *Journal of Computational and Applied Mathematics* vol. 318 (2017), pp. 181–188. DOI: 10.1016/j.cam.2016.10.028.

[301] K. R. Swanson, E. C. Alvord and J. D. Murray. 'A quantitative model for differential motility of gliomas in grey and white matter'. In: *Cell Proliferation* vol. 33, no. 5 (2000), pp. 317–329. DOI: 10.1046/j.1365-2184.2000.00177.x.

[302] P. Turchin and W. T. Thoeny. 'Quantifying Dispersal of Southern Pine Beetles with Mark-Recapture Experiments and a Diffusion Model'. In: *Ecological Applications* vol. 3, no. 1 (1993), pp. 187–198. DOI: 10.2307/1941801.

[303] M. Costabel and M. Dauge. 'A Singularly mixed boundary value problem'. In: *Communications in Partial Differential Equations* vol. 21, no. 11-12 (1996), pp. 1919–1949. DOI: 10.1080/03605309608821249.

[304] R. J. LeVeque and Z. Li. 'The Immersed Interface Method for Elliptic Equations with Discontinuous Coefficients and Singular Sources'. In: *SIAM Journal on Numerical Analysis* vol. 31, no. 4 (1994), pp. 1019–1044. DOI: 10.1137/0731054.

[305] A. Gasanov and I. Kaporin. 'Use of the exclusion method in solving strictly elliptic systems using the finite element method'. In: *USSR Computational Mathematics and Mathematical Physics* vol. 26, no. 3 (1986), pp. 120–130. DOI: 10.1016/0041-5553(86)90125-4.

[306] M. Kumar and P. Joshi. 'A Mathematical Model and Numerical Solution of a One Dimensional Steady State Heat Conduction Problem by Using High Order Immersed Interface Method on Non-Uniform Mesh'. In: *International Journal of Nonlinear Science* vol. 14, no. 1 (2012), pp. 11–22.

[307] J. Popovic and O. Runborg. 'Adaptive fast interface tracking methods'. In: *Journal of Computational Physics* vol. 337 (2017), pp. 42–61. DOI: 10.1016/j.jcp.2017.02.017.

[308] Z. Muradoglu-Seyidmamedov and E. Ozbilge. 'A mathematical model and numerical solution of interface problems for steady state heat conduction'. In: *Mathematical Problems in Engineering* vol. 2006 (2006), pp. 1–18. DOI: 10.1155/mpe/2006/20898.

[309] M. J. Ablowitz and A. Zeppetella. 'Explicit solutions of Fisher's equation for a special wave speed'. In: *Bulletin of Mathematical Biology* vol. 41, no. 6 (1979), pp. 835–840. DOI: 10.1007/bf02462380.

[310] K. Louw and R. J. Moitsheki. 'Group-Invariant Solutions for the Generalised Fisher Type Equation'. In: *Natural Science* vol. 07, no. 13 (2015), pp. 613–624. DOI: 10.4236/ns.2015.713061.

[311] K. R. Swanson, J. Ellsworth C. Alvord and J. D. Murray. 'Quantifying efficacy of chemother- apy of brain tumors with homogeneous and heterogeneous drug delivery,' in: *Acta Biotheoretica* vol. 50, no. 4 (2002), pp. 223–237. DOI: 10.1023/a:1022644031905.

[312] F. Sanchezgarduno and P. Maini. 'Traveling Wave Phenomena in Some Degenerate Reaction-Diffusion Equations'. In: *Journal of Differential Equations* vol. 117, no. 2 (1995), pp. 281–319. DOI: 10.1006/jdeq.1995.1055.

[313] S. Lie. 'Klassification und Integration von gewöhnlichen Differentialgleichungen zwischen x y, die eine Gruppe von Transformationen gestatten'. In: *Mathematische Annalen* vol. 32, no. 2 (1888), pp. 213–281. DOI: 10.1007/bf01444068.

[314] N. H. Ibragimov. 'Equivalence groups and invariants of linear and nonlinear equations'. In: *Archives of ALGA* vol. 1 (2004), pp. 9–69.

[315] Ibragimov. *Elementary Lie Group Analysis Ordinary*. John Wiley & Sons, 1999.

[316] M. L. Gandarias, M. Rosa and R. Tracinà. 'Symmetry analysis for a Fisher equation with exponential diffusion'. In: *Mathematical Methods in the Applied Sciences* vol. 41, no. 17 (2018), pp. 7214–7226. DOI: 10.1002/mma.4803.

[317] J. Belmonte-Beitia. 'On the existence of traveling wave solutions and upper and lower bounds for some Fisher–Kolmogorov type equations'. In: *International Journal of Biomathematics* vol. 07, no. 05 (2014), p. 1450050. DOI: 10.1142/s1793524514500508.

[318] K. R. Swanson et al. 'Quantifying the Role of Angiogenesis in Malignant Progression of Gliomas: In Silico Modeling Integrates Imaging and Histology'. In: *Cancer Research* vol. 71, no. 24 (2011), pp. 7366–7375. DOI: 10.1158/0008-5472.can-11-1399.

[319] A. Giesexs and M. Westphal. 'Glioma Invasion in the Central Nervous System'. In: *Neurosurgery* vol. 39, no. 2 (1996), pp. 235–252. DOI: 10.1097/00006123-199608000-00001.

[320] K. Swanson. 'Mathematical modeling of the growth and control of tumors'. PhD thesis. University of Washington, Seattle, Washington, 1999.

[321] E. Konukoglu et al. 'Image Guided Personalization of Reaction-Diffusion Type Tumor Growth Models Using Modified Anisotropic Eikonal Equations'. In: *IEEE Transactions on Medical Imaging* vol. 29, no. 1 (2010), pp. 77–95. DOI: 10.1109/tmi.2009.2026413.

[322] S. Chulián, M. Rosa and M. Gandarias. 'Reductions and symmetries for a generalized Fisher equation with a diffusion term dependent on density and space'. In: *Journal of Computational and Applied Mathematics* vol. 354 (2019), pp. 689–698. DOI: 10.1016/j.cam.2018.11.018.

[323] S. Chulián, M. Rosa and M. L. Gandarias. 'Symmetries and solutions for a Fisher equation with a proliferation term involving tumor development'. In: *Mathematical Methods in the Applied Sciences* vol. 43, no. 4 (2020), pp. 2076–2084. DOI: 10.1002/mma.6105.

[324] M. Rosa et al. 'Application of Lie point symmetries to the resolution of an interface problem in a generalized Fisher equation'. In: *Physica D: Nonlinear Phenomena* vol. 405 (2020), p. 132411. DOI: 10.1016/j.physd.2020.132411.

[325] J. E. Herrera, A. Minzoni and R. Ondarza. 'Reaction-diffusion equations in one dimension: particular solutions and relaxation'. In: *Physica D: Nonlinear Phenomena* vol. 57, no. 3-4 (1992), pp. 249–266. DOI: 10.1016/0167-2789(92)90002-5.