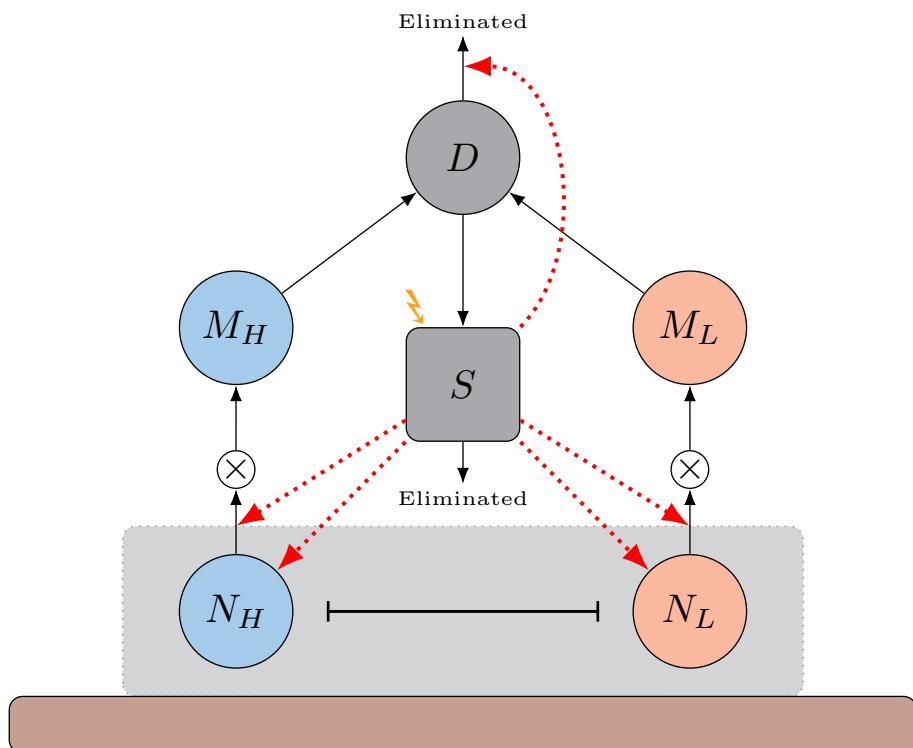


Mathematical Modelling of Myeloproliferative Neoplasms and Hematopoietic Stem Cells

Ph.D. thesis
by Rasmus Kristoffer Pedersen



Supervisors:

Professor Johnny T. Ottesen
Associate Professor Morten Andersen
Professor Dr. med. Hans C. Hasselbalch



This thesis has been submitted to the Doctoral School of Science and Environment, Roskilde University.

August 31st, 2020

Ph.D. thesis
Mathematical Modelling of Myeloproliferative Neoplasms and Hematopoietic Stem Cells

by Rasmus Kristoffer Pedersen

This thesis has been submitted to
the Doctoral School of Science and Environment
Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark
August 31st, 2020

Supervised by:

Professor Johnny T. Ottesen
Department of Science and Environment, Roskilde University

Associate Professor Morten Andersen
Department of Science and Environment, Roskilde University

and

Professor Chief Physician Dr. med. Hans C. Hasselbalch
Department of Haematology, Zealand University Hospital, Roskilde

Frontpage illustration:

The illustration depicts the “Combined Cancitis-Niche model” proposed in chapter 10. The model combines the features of the “Cancitis model” with additional features concerning the behaviour of hematopoietic stem cells, resulting in a model connecting hematopoietic stem cells and the immune system.
For details about the figure, see figure 10.2, page 96.

Abstract

The Philadelphia-negative myeloproliferative neoplasms (MPNs) is a group of blood cancers which include the diseases essential thrombocythemia (ET), polychytemia vera (PV) and primary myelofibrosis (PMF). Evidence suggests that ET, PV and PMF are closely connected and represent different stages along a biological continuum. In recent years, chronic inflammation has been found to be an important driver of cancers in general and of the Philadelphia-negative MPNs in particular. The concept of inflammation driving a biologically continuum of disease makes MPNs apt for investigations through mathematical modelling, with promise of improved patient prognosis and advancements in treatment.

In this thesis, we describe and analyse three mechanism-based mathematical models. All three models are described using systems of ordinary differential equations (ODE). This is done in an effort to gain insight into blood cancers in general and about a cohort of patients enrolled in the clinical trial “DALIAH” in particular (EudraCT number: 2011-001919-31). The patients considered here were all diagnosed with MPNs and treated with pegylated interferon- α (IFN) over a five year period.

The first model was proposed by Andersen et al. (2017) and describes the connection between chronic inflammation and the progression of MPN. A model extension allows us to relate patient-specific IFN-dose scheduling with patient data of leukocyte-counts, thrombocyte-counts and measurements of the $JAK2^{V617F}$ allele burden, showing good agreement between the dynamics of the mathematical model and the behaviour observed in data.

A novel mathematical model of the hematopoietic stem cells (HSC) is proposed. Mathematical analysis of the model suggests a notion of HSC fitness, found to determine long-term competition between HSC clones within the bone-marrow microenvironment. A model reduction is deemed appropriate, and HSC fitness is found to be principal in the behaviour of the reduced model as well.

We combine the two models into a single model describing both HSC-mechanisms, blood-cell production and the effect of chronic inflammation on MPN progression. This combined model can be used to investigate and simulate a wide range of scenarios, allowing us to make novel hypotheses about the behaviour of HSC and the entire hematopoietic system. In particular, we are able to interpret the effect IFN-treatment has on MPN-diagnosed patients, by relating the model to data of individual patients. For the IFN-treated MPN patients of the DALIAH trial, we observed a difference in the time-scale of the response of blood-cell counts and of the $JAK2^{V617F}$ allele burden. Based on our model-based investigations, we hypothesize that this difference in response is due to a two-fold effect of IFN: Production of blood-cells is inhibited for both healthy and malignant clones on a short time-scale, while malignant stem cells are substantially inhibited on longer time-scales.

Finally, we present a proof-of-concept of how mathematical modelling calibrated to patient-measurements at diagnosis can make predictions on the level of individual patients. We hypothesize that, with sufficient collection of patient-data and model-calibration, mathematical modelling could be an important prognostic tool in the clinic in the near future, allowing for improved treatment of MPNs using IFN.

Resumé (Abstract in danish)

De Philadelphia-negative myeloproliferative neoplasier (MPN) dækker over en gruppe af blodcancere der inkluderer sygdommene essentiel trombocytose (ET), polycytæmi vera (PV) og primær myelofibrose (PMF). Forskning tyder på at ET, PV og PMF er tæt forbundet og repræsenterer forskellige stadier langs et biologisk kontinuum. I de seneste år er kronisk inflammation blevet foreslæbt som afgørende for udviklingen af cancersygdomme generelt og for Philadelphia-negative MPN i særdeleshed. Konceptet om at kronisk inflammation styrer et biologisk kontinuum af sygdomme gør MPN til en oplagt gruppe af sygdomme at undersøge ved hjælp af matematisk modellering, med håb om forbedret patientprognose og videreudvikling af behandling.

I denne afhandling beskrives og analyseres tre mekanisme-baserede matematiske modeller. Alle tre modeller beskrives ved brug af ordinære differentialligninger (ODE). Dette gøres i et forsøg på at opnå indsigt om blodcancere generelt og om en cohorte af patienter fra det kliniske forsøg "DALIAH" specifikt (EudraCT ID: 2011-001919-31). Patienterne der her tages i betragtning, blev alle diagnosticeret med MPN og blev behandlet med pegyleret interferon- α (IFN) over en fem-årig periode.

Den første model blev oprindeligt beskrevet af Andersen et al. (2017) og beskriver forbindelsen mellem kronisk inflammation og udviklingen af MPN. En modeludvidelse gives os muligheden for at relatere patient-specifik timing af IFN-dosis til patient data for leukocyt-tal, thrombocyt-tal samt målinger af $JAK2^{V617F}$ allelebyrde, hvilket viser god overenstemmelse mellem dynamikken af den matematiske model og opførslen der observeres i data.

En ny matematisk model af hematopoitiske stamceller (HSC) præsenteres. Matematisk analyse af modellen leder til et udtryk for HSC fitness, der er afgørende for den lang-sigtede konkurrence mellem HSC-kloner i knoglemarven. Det vurderes at en model reduktion er passende, og HSC fitness viser sig også at være afgørende for opførslen af den reducerede model.

Vi kombinerer de to modeller til en samlet model der beskriver både HSC-mekanismer, blodcelle produktion og effekten af kronisk inflammation på udviklingen af MPN. Denne kombinerede model kan bruges til at undersøge og simulere en bred vifte af scenarier, hvilket gives os mulighed for at komme med nye hypoteser om opførslen af HSC og hele det hematopoitiske system. Dette gøres os i stand til at fortolke den effekt IFN-behandling har på MPN-diagnosticerede patienter, ved at relatere modellen til data for individuelle patienter. For de IFN-behandlede patienter fra DALIAH observerede vi en forskel i tidsskalaerne for respons af blodcelletal og $JAK2$ allelebyrde. Baseret på vores model-baserede undersøgelser, vurderer vi at forskellen på respons skyldes en to-delt effekt af IFN: Produktionen af blodceller begrænses på den korte tidsskala for både raske og maligne celler, hvorimod de maligne stamceller bliver særligt begrænset på længere tidsskalaer.

Afslutningsvis præsenterer vi et "proof-of-concept" for hvordan matematisk modellering kalibreret med patient-målinger fra diagnosetidspunkt kan lede til forudsigelser for individuelle patienter. Vi vurderer at matematisk modellering, kombineret med tilpas indsamling af patient-data og model-kalibration, kan blive et vigtigt værktøj for patient-prognoser i klinikken i den nære fremtid, og vil give anledning til forbedret behandling af MPN ved brug af IFN.

Preface

This Ph.D. thesis has been submitted to the Doctoral School of Science and Environment, Roskilde University, as part of the requirements for obtaining the Ph.D. degree. The work presented in the thesis was initiated in September 2017, and submitted in August 2020. During this period, I worked at the Department of Science and Environment, Roskilde University, Denmark, under the supervision of Professor Johnny T. Ottesen (Roskilde University), Associate professor Morten Andersen (Roskilde University), and Dr. Med. Hans C. Hasselbalch (Zealand University Hospital). Part of the work was carried out during a five month research stay from February 2019 to June 2019 at the Applied Analysis group at Heidelberg University, Germany, collaborating with Dr. rer. nat. Thomas Stiehl, MD. The research was supported by funding from Roskilde University and Region Sjælland.

Throughout the thesis, I use “we” to refer both to work and thoughts that are my own, as well as work and thoughts that are attributed to my supervisors and collaborators. This is done to acknowledge the great work that others have played in bringing the present work to fruition. As such, the results, findings and novel ideas presented here is the result of hard work of many great people, but the errors are mine alone.

Acknowledgements

I could not have written this Ph.D. thesis without the support and help of others. For this I am very thankful, and despite the saying “No-one mentioned — No-one forgotten”, I wish to thank some people in particular.

First and foremost, I wish to thank my main supervisor Professor Johnny T. Ottesen for his excellent help and supervision through these three years. In addition to being a great mathematician and offering his knowledge in all matters mathematical, Johnny has also been a support personally. With sincere interest and curiosity, Johnny has listened to, and helped with, any question or challenge that arose these three years, either through supervisor-meetings or informal “knock-on-the-door-during-stressful-times” talks. No-one else could have given me a better introduction into the field of mathematical modelling. In addition, Morten Andersen deserves my thanks, for not only his supervision and mathematical guidance, helpful suggestions for corrections (both mathematical and linguistic) and other academic things, but also for showing how the world of academia can be both fun and exciting. Hans C. Hasselbalch has similarly played an important role in supervising and directing my Ph.D.-project, for which I am very grateful. Helping a mathematician with no background in biology understand the vast world that is cancer, stem cells and “fire in the bone-marrow” is no easy task, but Hans managed to guide me with great skill. Through his numerous (and humorous) metaphors and vivid imagery of biological mechanisms, my understanding of cancer biology was formed with great pleasure.

I have had the pleasure to be part of the newly formed “Cancitis” research group. It has been fantastic to be able to work alongside fellow Ph.D.-students Johanne, Zamra and Marc through these three years, and I am thankful to have had them to collaborate and discuss with in this collaborative effort of trying to connect mathematics and cancer research. I hope, and expect, that the Cancitis group will see much expansion and many publications in the years to come, and feel privileged to have been part of the early beginnings of the group.

It can be difficult for mathematicians and medical professionals to successfully collaborate. However, thanks to the great people at Zealand University Hospital/Roskilde Hospital, it has been an absolute joy. In addition to my thanks to Hans for initiating the collaboration, I would also like to thank Lasse, Vibe and Trine for their company and insight in the many (and

sometimes long) meetings in the greater Roskilde MPN collaboration. You have illustrated that cross-discipline collaboration is not only an important and necessary part of research, but also a fun endeavour.

Part of this project was done in collaboration with Thomas Stiehl during my five-month stay at Heidelberg University. Thomas was incredible welcoming in regards to my stay and I am very grateful for having had the opportunity to work with him and learn from him. People with such expertise in both mathematics and medicine are rare to come by, and I appreciate the chance to take out time of Thomas' busy schedule to have his inputs for my work. I would also like to thank professor Anna Marciniak-Czochra for making my stay in Heidelberg a possibility and welcoming me into the research group during my stay. Similarly, I wish to thank the Ph.D.-students, post-docs and visiting students of the group whom I had the great pleasure of spending time with during my time in Heidelberg. You all contributed to making my five months in the city feel like home and made it an unforgettable and positive experience. Vielen Dank dafür.

Through my studies, including those preceding the Ph.D., I've been happy to be part of the IMFUFA environment at Roskilde University. The people of IMFUFA are some of the smartest mathematicians and physicists that I will ever know, and it has been a pleasure to be able to call them my colleagues these past three years. A special thanks goes out to the IMFUFA youngsters, that is, the Ph.D.-students, post-docs and other "young" members of staff. In particular, I would like to thank Natasja for being the best office-mate that one could wish for. Sharing (and decorating) the office these years has been such a pleasure and has made even the most stressful and depressing parts of the project very enjoyable.

Doing a Ph.D. is not always easy, but I'm lucky to have had great people around me that support me both personally and emotionally. I particularly wish to thank Stefan, Jakob and Caro for all their help and all they have given me.

Finally, I wish to thank my family, and especially my mother, for love and support throughout the entirety of my education. This has been a long journey from the first day of school to now doing a Ph.D., and I could not have done it without a loving and caring family to help me through.

- Rasmus Kristoffer Pedersen
August 2020

Contents

Abstract	i
Resumé (Danish Abstract)	ii
Preface	iii
Acknowledgements	iii
I Introduction	1
1 Introductory remarks	3
1.1 Overview and introduction to the field	3
1.2 Related publications	4
1.3 Outline of thesis and reading guide	5
1.4 Overview of supplementary material	6
II Myeloproliferative Neoplasms	7
2 Biological background	9
2.1 General introduction to MPNs and related blood cancers	9
2.2 HSC and the bone-marrow microenvironment	10
2.3 Interferon- α	12
3 The DALIAH trial	15
3.1 Presentation of data	15
3.2 Empirical modelling of JAK2 $V617F$ dynamics	17
3.3 Summarizing discussion	21
III Mathematical modelling of MPNs	23
4 Brief review of mathematical models of blood cancers	25
5 The Cancitis model	29
5.1 Presentation of the Cancitis model and overview of previous results.	29
5.2 The reduced Cancitis model	36
5.3 Summarizing discussion	37

6 Modelling MPN-patient responses with the Cancitis model	43
6.1 Response to IFN treatment in the Cancitis model	44
6.2 Interpretation of $JAK2^{V617F}$ allele burden in the Cancitis model	44
6.3 Fitting procedure for $JAK2^{V617F}$ allele burden	45
6.4 Results of fitting to available patient data	46
6.5 Predicting $JAK2^{V617F}$ kinetics	49
6.6 Fitting procedure for blood-cell counts	51
6.7 Considering multiple parameters	51
6.8 Summarizing discussion	51
IV Modelling HSCs and the bone-marrow microenvironment	55
7 Brief review of mathematical models of HSC	57
8 The HSC niche model	59
8.1 Constructing the HSC niche model	59
8.2 Asymmetric division	63
8.3 Dynamics and behaviour of the HSC niche model	64
8.3.1 Existence of an attracting trapping region	64
8.3.2 Steady states of the HSC niche model	65
8.3.3 Local stability of steady states	68
8.3.4 Estimates for parameters and numerical simulations	71
8.4 Summarizing discussion	72
8.4.1 Model formulation	72
8.4.2 Fitness and structure of steady states	74
8.4.3 Model parametrization and numerical findings	75
9 Model reduction	77
9.1 Reducing the HSC niche model	77
9.2 Production of progenitors in the reduced niche model	81
9.3 The transformed HSC niche model	82
9.4 The logistic approximation in the two-clone case	84
9.5 Summarizing discussion	85
V Combining the Cancitis model and the bone-marrow model	89
10 Formulation of a combined Cancitis-Niche model	91
10.1 Connecting HSC and mature cells	91
10.2 The combined model	94
10.2.1 Formulation of the combined model	94
10.2.2 Brief comments on the formulation of the combined model	97
10.2.3 Feasibility	98
10.2.4 Steady states	98
10.2.5 Parametrization and numerical investigations	102
10.3 Summarizing discussion	112

11 Modelling MPN-patient responses with the combined model	115
11.1 Mapping blood-concentration to parameter perturbation	115
11.2 Methodology for fitting the Combined model to data	117
11.3 Examples of patient fits	118
11.4 Population-level interpretation of fits to data	122
11.5 Proof-of-concept of population-level responses	125
11.6 Summarizing discussion	128
VI Discussion and concluding remarks	133
12 General discussion of findings, limitations and future work	135
12.1 Implications of the proposed models	135
12.2 Limitations of data	137
12.3 Relating data and models	139
12.4 Future aims of mathematical modelling of MPN	141
Bibliography	144
VII Appendix	153
A Reprints of articles	155
A “Data-driven analysis of JAK2V617F kinetics during interferonalpha2 treatment of patients with polycythemia vera and related neoplasms”	155
B “HSC niche dynamics in regeneration, pre-malignancy and cancer: Insights from mathematical modeling”	169
C “Mathematical modelling of the hematopoietic stem cell-niche system: Clonal dominance based on stem cell fitness.”	197
B Poster presentations	231
A “Treatment of Blood Cancers and the Importance of Quiescence”	231
B “Modelling the Dynamics of Hematopoietic Stem Cells”	233

Part I

Introduction

Chapter 1

Introductory remarks

1.1 Overview and introduction to the field

Mathematical modelling is becoming an increasingly important part of civilization. Whether it is in the exploration of the big or the small, mathematical modelling is an all-encompassing tool in strengthening our understanding about any system. Mathematical models can make the invisible visible and make predictions about parts of the world that are otherwise inconceivably complex. There is no clearer example of the significance of mathematical modelling in recent memory than the incredible work done to mathematically model and predict the spread of the novel Corona-virus *COVID-19*, spreading through the world at the time of writing. Rarely has the importance and effectiveness of mathematical modelling been actively disseminated and used to influence political decisions about public health on the scale that we currently see.

Biological systems are incredibly complex and observing a biological system without disturbing its natural behaviour is a challenge. As vast quantitative measurements are becoming more readily available, the use of mathematical modelling and data-driven tools are expanding at an incredible rate. The interest in mathematical biology is clear, with new educations being introduced, as well as journals and conferences on mathematical biology increasing in popularity. It is therefore no surprise that the intersection between mathematics and the field of cancer research is growing into a significant sub-field by itself. Daily, many lives are lost to cancer, and as a consequence cancer research and prevention is one of the largest and most important fields of research in the modern era. The important cross-disciplinary field of mathematical oncology, that is, the combination of mathematics and cancer research, entails the use of many disciplines, both mathematical, medical and biological. From the mathematics of evolutionary biology, to hypothesis-generating mechanism-based mathematical models of particular cancers, mathematics provides many of the tools that can benefit cancer research. The increasing focus on mathematical oncology is clear, from both review-articles on mathematical modelling in medical journals (Altrock et al., 2015), to the recent creation of a “Mathematical Oncology”-subgroup of the Society of Mathematical Biology (SMB). Mathematics is likely to play a role in future treatment of different types of cancers, both solid tumours and hematologic cancers. Executable mathematical models have already been used successfully to improve treatment of cancer patients (Clarke and Fisher, 2020), and personalized models of the cancer of an individual person can help clinicians give patients the best possible treatment.

The work of this thesis positions itself within this field of mathematical oncology. Focusing primarily on a particular group of hematologic malignancies, the myeloproliferative neoplasms (MPNs), the work presented here represent a small fraction of the general mathematical work on hematologic malignancies and cancer in general. No two cancers are the same, and neither are any two patients. Thus, to attain meaningful results that can help patients, it is necessary to

restrict oneself to a particular cancer and consider patients on an individual level. This thesis, and the considerations described herein, was initiated in an effort to relate mathematical modelling to a particular patient cohort; the MPN-diagnosed patients treated in the “DALIAH” clinical trial. As such, the overarching question of this work was simple: With a general mathematical model of blood-cell production in the human body, is it possible, using the DALIAH trial data, to gain insight about MPNs in general and the patients of the trial in particular? Investigating this question led us from an empirical study of the data itself, through the formulation and investigation of a mechanism-based model of hematopoietic stem cells and finally to relating multiple proposed models to patient-data on the level of individual patients. Whether these investigations tie together and provides an answer to the question sufficiently well is left to the reader to decide. Regardless, we believe that the work presented in this thesis demonstrates that mathematical modelling can indeed be an important tool in the treatment of MPNs. While the road toward clinical use of mathematical models to guide treatment of MPN-diagnosed patients is still long, this thesis describes our work toward that goal.

1.2 Related publications

Through my Ph.D. I have been part of the Cancitis group at Roskilde University, Denmark. Current and future publications by the group are related and relevant to the topics discussed in this thesis. I participated in the work and preparation of the following publications during my Ph.D.-studies:

- Pedersen, R. K., Andersen, M., Knudsen, T. A., Sajid, Z., Gudmand-Hoeyer, J., Dam, M. J. B., Skov, V., Kjær, L., Ellervik, C., Larsen, T. S., Hansen, D., Pallisgaard, N., Hasselbalch, H. C., and Ottesen, J. T. (2020). Data-driven analysis of JAK2 V617F kinetics during interferon-alpha2 treatment of patients with polycythemia vera and related neoplasms. *Cancer Medicine*, 9(6):2039–2051
- Ottesen, J. T., Pedersen, R. K., Dam, M. J. B., Knudsen, T. A., Skov, V., Kjær, L., and Andersen, M. (2020). Mathematical Modeling of MPNs Offers Understanding and Decision Support for Personalized Treatment. *Cancers*, 12(8):2119
- Pedersen, R. K., Andersen, M., Skov, V., Kjær, L., Hasselbalch, H. C., Ottesen, J. T., and Stiehl, T. (n.d.,1). Combining stem cell mobilization with preconditioning: Insights about competition in the stem cell niche from mathematical modelling. *Submitted*
- Pedersen, R. K., Andersen, M., Stiehl, T., and Ottesen, J. T. (n.d.,2). Mathematical Modelling of the Hematopoietic Stem Cell–Niche System: Clonal Dominance based on Stem Cell Fitness. *Submitted*
- Dam, M. J. B., Andersen, M., Pedersen, R. K., Kjær, L., Skov, V., Hasselbalch, H. C., and Ottesen, J. T. (n.d.). Data-Driven Analysis of the Kinetics of the JAK2V617F Allele Burden and Blood Cell Counts During Hydroxyurea Treatment of Patients with Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis. *Submitted*
- Ottesen, J. T., Pedersen, R. K., Sajid, Z., Gudmand-Hoeyer, J., Bangsgaard, K. O., Skov, V., Kjær, L., Knudsen, T. A., Pallisgaard, N., Hasselbalch, H. C., and Andersen, M. (2019). Bridging blood cancers and inflammation: The reduced Cancitis model. *Journal of Theoretical Biology*, 465:90–108

- Andersen, M., Sajid, Z., Pedersen, R. K., Gudmand-Hoeyer, J., Ellervik, C., Skov, V., Kjær, L., Pallisgaard, N., Kruse, T. A., Thomassen, M., Troelsen, J., Hasselbalch, H. C., and Ottesen, J. T. (2017). Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. *PLOS ONE*, 12(8)

For additional publications by the Cancitis group, see dirac.ruc.dk/cancitis.

1.3 Outline of thesis and reading guide

Most of the details of the publications mentioned above are described in this thesis. To distinguish between sections that describe novel work and work that have previously been presented, we here provide an overview of the thesis.

The thesis has been structured such that three main “stories” are told: The first describes the application of the *Cancitis model* to data, the second goes through the formulation and analysis of a novel model of HSC, while the third and final story combines the two models and applies the combined model to the same data. To explain the background behind these stories, an overview of the biology of blood-production in the human body is given, as well as a brief overview of previous work on mathematical modelling of blood cancers.

Because of the cross-discipline nature of the work, this thesis does not follow the classic structure of describing theory and methodology separate from results and discussion. Instead the thesis has been structured such that information is presented when they are deemed useful to the reader.

The outline of the thesis is as follows:

- Part II describes blood-production and Myeloproliferative Neoplasms (MPNs) without explicitly considering the primary mathematical models that are the focus of this thesis. An overview of blood cancers, MPNs, hematopoietic stem cells and the drug Interferon- α is given in chapter 2, to provide the most significant parts of the biological background on which the present work is based. The data-material from the DALIAH trial is described in section 3, and the empirical analysis from (Pedersen et al., 2020) is summarized.
- In part III, focus is on mathematical modelling of MPNs. First, previous work on mathematical modelling of blood cancers is discussed in section 4, followed by the formulation of the *Cancitis model* in section 5, including some of the most important results that the model gives rise to. This includes a short review of some of the results of Ottesen et al. (2019). The work of Ottesen et al. (2020) is described in detail in section 6. The work consists of a comparison and fitting of the Cancitis model to the data from the DALIAH trial, in which model parameters are perturbed in accordance with the real data from patient-specific treatment protocols.
- Part IV presents a novel model of HSC and its background. First, important related work on modelling HSC dynamics is presented in section 7. The novel model is then presented in section 8. The analysis of the model described in (Pedersen et al., nd1) and (Pedersen et al., nd2) is summarized, and the model reduction of (Pedersen et al., nd2) is described in section 9.
- The focus of part V is to combine the HSC model with the Cancitis model and investigate the resulting combined model. The formulation of the combined model is given in section 10, along with some immediate results. In section 11 the model is compared to the data from the DALIAH trial in a way similar to how the Cancitis model was compared to the data in section 6. The entirety of part V is novel, and has not yet been published.

- Finally, in part VI we discuss and comment on the results and investigations described throughout the thesis. This includes both comments on specific findings of some parts as well as a zoomed-out discussion of the work as a whole.

The three stories mentioned above provides three separate ways that this thesis can be read:

- *Application of the Cancitis model to data:* After the biological background of section 2, and the presentation of the data-material in section 3, the reader can skip to the Cancitis model and the application to data in sections 5 and 6 respectively.
- *The novel HSC model:* Section 2.2 gives a background about the biology of HSC. This leads into the mathematical background, model-formulation, -analysis and -reduction described in part IV.
- *A combined model and its application:* While the entire thesis builds around this story, reading parts II and V should be sufficient to understand how the combined model relates to the data of the DALIAH trial and MPN in general.

1.4 Overview of supplementary material

Full text of three scientific articles are included in supplementary material A at the end of this thesis. Two of these were not yet published at the time of writing, and we refer to them as (Pedersen et al., nd1) and (Pedersen et al., nd2). The final article, (Pedersen et al., 2020), is included for completeness sake. Similarly, two posters are included, in supplementary material B. In addition to the supplementary publications, a large number of figures of data from individual patients and model fits to the data play an important role in this thesis. To reduce the number of pages that this thesis contains, these are only available electronically. Select examples are presented and discussed throughout the thesis, but we refer the reader to the electronic supplementary whenever available.

On both the pages rasmuspedersen.com/phd and dirac.ruc.dk/~rakrpe/phd links to the electronic supplementary can be found.

Feel free to contact mail@rasmuspedersen.com if any problems arise with accessing the supplementary material.

The online supplementary material consists of PDF-files and ZIP-archives with images in PNG-format. Each ZIP-archive also contains a `readme.txt` textfile, explaining the figures in the archive. The supplementary files are:

Electronic Supplementary Material 1 *Supplementary material of (Pedersen et al., 2020).*

Detailed methodology and patient-specific fits. Discussed and referred to in section 3. PDF-file.

Electronic Supplementary Material 2 *Supplementary material of (Ottesen et al., 2020).*

Detailed methodology and patient-specific fits. Discussed and referred to in section 6. PDF-file.

Electronic Supplementary Material 3 *Additional fits of the Cancitis model to data.*

Further investigations that were not discussed in (Ottesen et al., 2020). Discussed and referred to in section 6. ZIP-archive of PNG-files.

Electronic Supplementary Material 4 *Fits of the Combined Cancitis-Niche model to data.*

Discussed and referred to in section 11. ZIP-archive of PNG-files.

Electronic Supplementary Material 5 *Virtual patient simulations, Combined Cancitis-Niche model.*

Discussed and referred to in section 11.5. ZIP-archive of PNG-files.

Part II

Myeloproliferative Neoplasms

Chapter 2

Biological background

This thesis describes the development, analysis and application of three mathematical models related to the hematopoietic (blood-cell-producing) system of the human body. A common goal for the development of the models was to understand and describe the behaviour of cancers of the blood. While the models presented later do not necessarily relate to any particular blood cancer, special focus was given to the long-term effects observed in slowly developing blood cancers and in particular the group of Myeloproliferative Neoplasms (MPNs). In this section, we briefly describe some of the background about blood cancers and MPNs that sets the stage for the remainder of the work described in this thesis. Focus of this thesis is on mathematical modelling rather than cell biology or clinical practice, and hence this section should be considered an introduction to hematopoietic malignancies instead of a thorough review of the intricacies of malignancies.

2.1 General introduction to MPNs and related blood cancers

The production of blood cells in the human body, commonly referred to as *hematopoiesis*, is a complex and intricate system of cells of different cell-lines and in all stages of cell-maturity. A branching tree of cell-types have at one end terminally differentiated mature cells such as e.g. *thrombocytes*, *leukocytes* or *erythrocytes*, and at the other end a common root of pluripotent *hematopoietic stem cells* (HSC). HSC exist within the bone-marrow microenvironment and maintain a steady production of differentiated progenitor cells without exhaustion of HSC counts under normal circumstances. Cell-division of HSC can give rise to mutations that can ultimately lead to the emergence of a malignant stem cell clone and in turn hematologic malignancy. Further details about HSC and the bone-marrow microenvironment necessary to maintain hematopoiesis is discussed in section 2.2 below.

Acute myeloid leukemia (AML), is the most common myeloid leukemia (Estey and Döhner, 2006), and is believed to agree with the cancer stem cell hypothesis, that is, the hypothesis that a population of malignant or leukemic stem cells are the root cause of disease (Dick, 2008; Tan et al., 2006). Similarly, chronic myeloid leukemia (CML) also derives from malignant stem cells. CML is included in the group of myeloproliferative neoplasms (MPNs), and arises from a stem cell clone acquiring the Philadelphia-chromosome expressing the BCR-ABL gene (Goldman and Melo, 2003; Holyoake and Vetrie, 2017). To distinguish other MPNs from CML, the term Philadelphia-negative MPN is used to refer to diseases where the Philadelphia-chromosome is not found. Common for MPNs is an excessive production of mature blood-cells, included in

Part II Myeloproliferative Neoplasms

the diagnostic criteria (Arber et al., 2016). In this thesis, focus is given to the Philadelphia-negative MPNs, in particular the blood cancers essential thrombocythemia (ET), polycythemia vera (PV) and primary myelofibrosis (PMF), which were the investigated in the “DALIAH” clinical trial described in section 3. A diagnosis of prefibrotic myelofibrosis (prePMF) is also considered in the trial, however, for our purposes we consider PMF and prePMF the same group. Patients diagnosed with ET and PV are at risk of major thrombosis (Tefferi and Elliott, 2007). In addition, MPN patients have been found to be at higher risk of secondary cancers, both hematologic and nonhematologic (Frederiksen et al., 2011). A meta-analysis of literature on incidence rates determined estimates of 1.03, 0.84 and 0.47 per 100,000 for ET, PV and PMF respectively (Titmarsh et al., 2014), suggesting that the Philadelphia-negative MPNs are rare compared to other cancers. However, a recent Danish study has shown that the MPNs are significantly under-diagnosed in the population, and accordingly a large number of people are at constant risk of suffering potentially life-threatening blood clots (Cordua et al., 2019). It is therefore appropriate that the Philadelphia-negative MPNs recently has been the focus of a special issue of the journal “*Cancers*”, in which our own work was included (Ottesen et al., 2020). In the 2016 revision of the WHO classification of MPNs, the presence of the JAK2^{V617F} mutation is considered a major criteria for the diagnosis of both ET, PV and PMF (Arber et al., 2016). Other criteria of relevance for this work is the increased thrombocyte count characteristic to ET and the minor criteria of increased leukocyte counts for PMF. As a measure of the stage of disease, the JAK2^{V617F} allele burden is typically considered. The allele burden is given by the number of alleles with mutation divided by the total number of alleles. For heterozygous cell where the JAK2^{V617F} mutation appears only on one allele, the allele burden cannot be above 50%. However, homozygosity of the JAK2^{V617F} mutation can arise, leading the JAK2^{V617F} allele burdens that are above 50%. Allele burdens above 50% are hence indicative of cells with homozygous JAK2^{V617F} mutations. For newly diagnosed patients, Larsen et al. (2007) found that ET patients had a median JAK2^{V617F} allele burden of 7%, PV patients a median of 33% and PMF patients had a median of 67%. Note that while presence of the JAK2^{V617F} mutation is a criteria for ET, the presence of the CALR and MPL mutations are also sufficient for ET diagnosis. In fact, while almost all PV patient were positive for the JAK2^{V617F} mutation, only about half of ET patients have been found positive (Larsen et al., 2007). A concept of a biological continuum of disease that ET, PV and PMF are part of, date as far back as the 1950’s (Dameshek, 1951), and modern gene expression profiling supports this concept (Skov et al., 2010). Hence, the increasing JAK2^{V617F} allele burden from ET through PV to PMF suggests that the JAK2^{V617F} allele burden could be used to identify where along the continuum the patients are. While the ranges of JAK2^{V617F} allele burdens can vary much between individual patients, the concept of ET to PV to PMF being characterized by increasing JAK2^{V617F} allele burden is a fundamental assumption throughout this thesis. In a general population study, 0.1% of the population were found positive for the JAK2^{V617F} mutation, of which the majority later developed MPN-symptoms (Nielsen et al., 2014). This could suggest that the early stages of MPN-development is symptom-free and that a slow increase in the JAK2^{V617F} allele burden can be used as a measure of the disease, even in the early stages before symptoms arise. Understanding how the disease develops in this early stages is hence important for clinical decisions. For brevity, we write JAK2 allele burden when referring to the JAK2^{V617F} allele burden for the remainder of this thesis.

2.2 Hematopoietic stem cells and the bone-marrow microenvironment

HSC maintain hematopoiesis and mutations of HSC are believed the root cause for most hematologic malignancies, including MPNs. Wilson and Trumpp (2006) gives a thorough review of the nature of HSC and the HSC niches within the bone-marrow microenvironment. The text in this section is based on this work, unless otherwise noted.

Stem cells are traditionally classified by their capability of multi-potent differentiation and self-renewal. Multi-potent differentiation allows similar stem cell to give rise to different cell-lines, while self-renewing division allows the stem cells to multiply and hence maintain cell-counts, even when stem cells are lost due to differentiation. The hematopoietic stem cells are, as the name suggests, the stem cells that maintain the hematopoietic system. Although it can be difficult to define distinct cell-types explicitly, a number of cell-sorting assays exist to identify and distinguish HSC. We assume the classification of “HSC” to be well-defined for the purposes of this thesis. After differentiation, the resulting cell is committed toward a particular cell-line. While there exists a hypothesis of de-differentiation where a cell returns to a less committed state (Jilkine and Gutenkunst, 2014), it will not be considered in this thesis and all differentiated HSC will be referred to as *progenitors*. Progenitors can differentiate into later-stage cell-types which ultimately, after multiple steps of progenitors, give rise to the mature cells that make up the blood. Progenitors are also capable of self-renewal, however, only to a limited degree (Oguro et al., 2013). Hence, without newly differentiated HSC, a population of progenitors will eventually die out. As mature cells undergo apoptosis (programmed cell-death), a steady production of mature cells is necessary. Hence a steady output of progenitors must be maintained by the HSC population.

The potential of HSC has been demonstrated by transplanting a single HSC into lethally irradiated mice (i.e. mice that had undergone irradiation such that the entire HSC population was eradicated, which leads to death of the mouse due to the lack of blood-production.). In such experiments, the single HSC is capable of reconstituting the entire hematopoietic system in the long term, saving the mouse.

A concept of asymmetric division exists in which HSC division leads to two distinct cells, one which is a progenitor cell and one which maintains stemness. How frequently this occurs for HSC is unclear, and while it has been observed *in vitro*, it is possible that it does not occur naturally *in vivo*. Wilson and Trumpp (2006) also hypothesizes that purely environmental differentiation can occur where a HSC spontaneously differentiates due to environmental changes. Hence HSC are hypothesized to divide, either symmetrically, resulting in two progenitors or two HSC, or asymmetrically resulting in one of each, but individual HSC are also hypothesized to differentiate directly into a single progenitor cells.

While HSC can be found throughout the body (e.g. in the bloodstream or in the spleen), they are primarily found within the bone-marrow microenvironment (BM). While direct observation of the BM is difficult *in vivo*, HSC are typically found in the vicinity of certain cellular groupings, and a concept exists of particular *BM niches*. Defining what cells constitute these niches is not trivial and is under great debate (Pinho and Frenette, 2019). Indeed multiple types of niches are hypothesized to exist, such as a niche for long-term storage of HSC and a separate niche for actively dividing HSC. Differences in niches could explain the environmental differentiation mentioned above. A significant proportion of HSC are dormant or quiescent for extended periods of time, referred to as the G_0 phase of the cell-cycle. Particular BM niches are believed to contribute to this maintenance of long-term cellular quiescence of HSC. For simplicity, we limit the definition of BM niche to refer to just this quiescence-inducing type of niche for the rest of

Part II Myeloproliferative Neoplasms

this thesis, and hence distinct from BM niches with different cellular composition and purposes, such as the vascular niches.

Long-term quiescence of HSC is considered critical for maintaining blood-production. Feedback-signalling is believed to allow for activation of quiescent HSC which can regulate the production of progenitors and hence the subsequent production of different mature cells. Such feedback could occur due to e.g. blood-loss or disease giving rise to a scarcity of immune cells such as T-cells that requires replenishing. Hence a vast and complex system of signalling is involved in ensuring a life-long production of blood-cells that is robust to external changes, and a large reservoir of quiescent HSC plays an important role in this system. One possible evolutionary benefit of HSC quiescence could be that it implies less dependence on single cells and hence increase the chance that a single erroneous division or mutation has significant impact on the survival of the organism. In addition, it is possible that there are some undetermined cellular differences between HSC before division and the resulting post-division HSC, and that quiescence and cellular maintenance through the BM niches is necessary to revert the HSC back to the original state that allowed for self-renewing division. This hypothesis explains the lack of self-renewal of HSC observed in cell-culture; outside the body, HSC do not expand in numbers (Kumar and Geiger, 2017). Such *ex vivo* expansion of HSC would be very beneficial for use in e.g. stem cell transplantation, and methods and protocols for inducing HSC expansion without the BM niche remains an important field of research (Zhang and Gao, 2016).

As discussed in section 2.1 above, rare mutations can occur during HSC division. The cell-biological details of such mutations are insignificant for our present purposes, and will not be discussed in this thesis. Instead cells are considered mutated when their cell-properties are distinct from the pre-mutation population of HSC. Mutations can influence a myriad of cell-properties, leading to subpopulations of HSC that behave differently or react to signalling in a different way. To distinguish between a subpopulation of mutated cells, *HSC clone* is used throughout this thesis to refer to a population of HSC that are functionally similar. Some mutations lead to a HSC clone that outcompete the healthy wild-type (unmutated) HSC population. This can give rise to clonal hematopoiesis of indeterminate potential (CHIP), which is characterized by a highly homogeneous population of HSC. While not necessarily a health risk by itself, if the clone is similar to wild-type HSC, subsequent mutations is believed to be a risk-factor for individuals in which CHIP has been identified (Steensma et al., 2015; Jaiswal and Ebert, 2019). Many hematologic malignancies are believed to arise from mutations in HSC, in agreement with the cancer stem cell hypothesis, a commonly accepted theory of cancer development (Reya et al., 2001). While some malignancies are hypothesized to arise from mutations of progenitors, either by themselves or supplementary to HSC mutations, (Dingli and Pacheco, 2010), these will not be considered in this thesis. A mutated clone that lead to hematologic malignancy is referred to as a *malignant* or *leukemic stem cell clone*. A range of different defects of the malignant stem cells lead to the symptoms observed in the given malignancies. For some hematologic malignancies, specific mutations have been identified to give rise to the malignant stem cells that are the root of the diseases. The $JAK2^{V617F}$ mutation is indicative of MPNs, and a clone of $JAK2$ mutated malignant stem cells are believed to be at fault for the harmful increase in blood-cell production observed in MPNs. Malignant stem cell clones have been found to engraft in the BM at sites where healthy HSC are typically found (Ishikawa et al., 2007), suggesting that the malignant cells interact with the same BM niches as the wild-type cells. Mathematical modelling work suggests that the decreasing population-size of healthy HSC observed in AML-patients can be explained by the malignant clone having a high affinity niche-interaction compared to the healthy clone, and possibly even an ability to cause the healthy clone to dislodge from the BM niche (Wang et al., 2017). The binding of malignant stem cells with the BM niches lead to a hypothesis for why treatment of hematologic malignancies can be difficult: If the niches induce quiescence of

malignant cells as they do for wild-type HSC, this could explain how malignant cells persist after chemotherapy that targets actively cycling cells (Ishikawa et al., 2007).

2.3 Interferon- α

While it is not entirely clear which kinds of signalling is involved in the activation of quiescent HSC and their return to cycling, interferon- α (IFN) is known to play a significant role. IFN induces cell-cycling for both quiescent HSC and early stage progenitor cells (Trumpp et al., 2010). For this reason, it has been suggested that treatment with IFN can be beneficial to awaken the dormant HSC e.g. to allow for clearing of malignant HSC that have entered quiescence. Based on this idea, Trumpp et al. (2010) suggests CML patients could benefit from treatment in which IFN is used to activate malignant HSC followed by subsequent eradication the active cells using *Imatinib*, a popular and successful drug for CML treatment. Combination therapy in which both IFN and *Imatinib* is used concurrently has already shown great promise for successful treatment of CML patients (Talpaz et al., 2015; Simonsson et al., 2011), and the concept of IFN-induced activation of quiescent HSC thus provides an explanation for this success.

Various forms of IFN are used to treat patients diagnosed with MPN. In particular, pegylated interferon alfa-2a (*Pegasys*) and pegylated interferon alfa-2b (*PegIntron*) is the focus of the present work. In general, we refer to all types of interferon- α as simply IFN throughout this thesis and assume their effect to be similar enough to be considered identical.

Treatment with IFN has been associated with a significant decrease in risk of thrombosis (Stauffer Larsen et al., 2013), and even induce “minimal residual disease” (MRD) in some patients, classified as a sustained decrease of risk-factors for the patient (Hasselbalch, 2011). This suggests that IFN can be beneficial in the long term for MPN-patients and perhaps even be part of a possible cure (Hasselbalch and Holmström, 2018). IFN has been found to lead to a sustained decrease of the JAK2 allele burden in a subset of patients, even bringing it to undetectable low level (Quintás-Cardama et al., 2013). The reduction of the JAK2 allele burden is achieved through long-term/chronic treatment with IFN, however, induced cell-cycling of both wild-type and $JAK2^{V617F}$ mutated HSC has been found to be induced both in short- and long-term administration of IFN (Austin et al., 2020). In addition, Austin et al. (2020) found that long-term (eight weeks) treatment with IFN reduced leukocyte-counts significantly in mice, as well as reduced the enlarged spleen commonly observed for $JAK2^{V617F}$ positive mice and humans. While many of the effects of IFN were found to apply to both wild-type and JAK2 mutated HSC, IFN has been found to preferentially target mutated cells (Mullally et al., 2013). Activation of HSC has been shown to cause DNA damage of the HSC in mice (Walter et al., 2015), suggesting that the additional activation of JAK2 mutated HSC could lead to additional accumulation of DNA damage for the malignant clone compared to the wild-type HSC. This provides a hypothetical explanation for why mono-therapy with IFN is sufficient for successful treatment of some MPN-diagnosed patients.

The high risk of thrombotic events associated with untreated MPNs and the evidence that IFN can induce long-term benefits is suggestive of the idea of early treatment with IFN, to reduce the MPN-related risks and halt the disease progression already in the early stages (Hasselbalch and Bjørn, 2015).

There is an urgent unmet need for insight about how the blood system changes from early pre-MPN-diagnosis stages with low JAK2 allele burden to later stages related to high risk of thrombosis. Additionally, it is of great importance to understand and identify quantitatively how IFN works, so future treatment methodologies with improved patient prognosis can be developed. It is from these purposes that the mathematical work detailed in this thesis takes form.

Chapter 3

The DALIAH trial

In this section, we describe the data material considered throughout the thesis, as well as the clinical trial “DALIAH” from which it arose. As the design and details of the trial are not the focus of this work, we only briefly discuss some of the relevant details. In our published work, Pedersen et al. (2020), the data was analysed. The analysis was based on an empirical modelling approach, rather than a mechanism-based approach. These preliminary investigations about the data on its own are relevant for the mathematical modelling presented later in this thesis.

The DALIAH trial (EudraCT number: 2011-001919-31) was a prospective randomized open-label phase III clinical trial comparing Pegylated r-IFN α (IFN) monotherapy with Hydroxyurea monotherapy in MPN patients. While two types of IFN were considered (Interferon alfa-2a “*Pegasys*” and Interferon alfa-2b “*PegIntron*”), we do not distinguish between these in the present work, but rather assume the effect of treatment to be identical. Patients were followed for five years with blood samples taken at regular pre-determined intervals. Focus for the study presented by Pedersen et al. (2020) was the observations of the JAK2 allele burden for the patients receiving IFN monotherapy. Such measurements were made at inclusion (baseline) and approximately 4, 8, 12, 18, 24, 36, 48 and 60 months following inclusion. Blood-cell counts of e.g. thrombocytes and leukocytes were measured on a more frequent basis. While about 200 patients were enrolled in the study, some patients dropped out due to toxicity, death or other reasons, and hence full 60 month data for all patients was not available. Additionally, a low number of patients had periods where IFN was substituted or used in combination with Hydroxyurea. For these patients, only initial periods of treatment with only IFN was included in this study. With some additional minor exclusions of data from certain patient, due to e.g. clear errors in the data-material, the total data-set consisted of 63 patients. Of these patients, 17 patient were diagnosed with ET, 35 patients were diagnosed with PV, 6 patients with primary myelofibrosis and finally 5 patients with prefibrotic myelofibrosis. Throughout this thesis, anonymized IDs are used, allowing for comparison of patients between the different studies described. The numbering was based on the original number of patients before the various exclusion criteria, and hence while some patients have IDs above 63, only the 63 patients described were considered.

3.1 Presentation of data

At inclusion into the study, blood samples were taken, allowing for a baseline-measurement before initiation of treatment. The distribution of these baseline measurements for all 63 patients are shown in figure 3.1, separated by diagnosis. Patients diagnosed with primary myelofibrosis or prefibrotic myelofibrosis were combined under the label “PMF”. Two-sample t -tests reveal that

ET patients had baseline thrombocyte counts and JAK2 allele burden that were statistically significantly different from the baseline values of the other diagnoses, PV and PMF (See caption of figure 3.1 for details). Note that a heightened thrombocyte count is a diagnostic criteria for ET, but not for PV and PMF diagnosis (Arber et al., 2016).

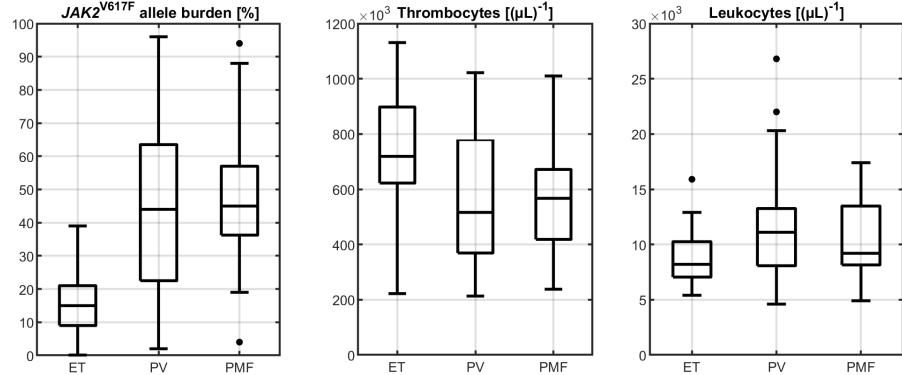


Figure 3.1: Distribution of the JAK2 allele burden, thrombocyte and leukocyte count at baseline, separated by diagnosis. Black dots depict outliers. Two-sample t -test were carried out to compare the baseline values. The baseline values of the JAK2 allele burden for ET patient were significantly different from PV patients and PMF patients with p-values of $7.3 \cdot 10^{-5}$ and $1.3 \cdot 10^{-4}$ respectively. Comparing the baseline count of thrombocytes for ET patients and PV patients gave a p-value of 0.013, while comparing ET and PMF gave a p-value of 0.044. Finally, the p-value of leukocytes baseline counts for ET compared to PV were 0.05, while all other comparisons gave p-values above 0.1.

In figure 3.2 the baseline data is depicted. The linear fit shown in the figure suggests that a linear relation between JAK2 allele burden and thrombocytes cannot be assumed ($r^2 = 0.13$). The same is the case for thrombocytes and leukocytes ($r^2 = 0.006$). A weak positive correlation between the JAK2 allele burden and the leukocytes counts is however observed ($r^2 = 0.26$). We note however, that visual inspection of the treatment-responses of individual patients suggests that thrombocyte and leukocyte-counts have similar kinetics in response to IFN treatment, and that decreasing JAK2 allele burden is also typically associated with decreasing blood-cell counts. Hypotheses about how differences in disease level (as assessed by the JAK2 allele burden) can relate to blood-cell counts, in particular thrombocyte-counts is discussed later in section 11.

For most patients, IFN had a strong immediate effect on the blood-cell counts, leading to a fast decline. Simultaneously, IFN also induced a decline of the JAK2 allele burden, however for most patients this appeared to occur more slowly. Figure 3.3 illustrates a typical patient response. The figure also depicts the average daily IFN dose for the particular patient. For some patients, significant changes in dose-size and timing were made, dependent on certain criteria. One exemplary criteria for dose escalation was the lack of a reduction in the blood-cell counts, specifically thrombocyte-counts that were not below $400 \cdot 10^3 (\mu L)^{-1}$ within the first 4 months. Further details about these criteria are detailed in the supplementary of (Pedersen et al., 2020), available as Electronic Supplementary Material 1. In the mathematical modelling presented later, the doses were averaged such that the daily average dose was used, rather than the wide range of different doses and timings possible. As an example, a weekly dose of $45\mu g$ was simplified as a

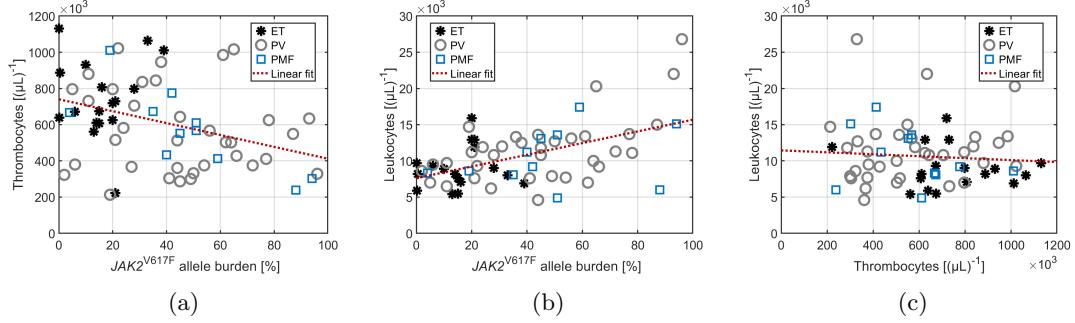


Figure 3.2: Plots of baseline data, with linear fits. All plots display the first data-point for each patient, with asterisks $*$ displaying patients that were diagnosed with ET, circles \circ showing patients diagnosed with PV and squares \square depicting patients with primary or prefibrotic myelofibrosis. A best fit of a linear function is shown in each plot. The linear fit of JAK2 allele burden to thrombocytes shown in panel (a) had $r^2 = 0.13$. For the relation between JAK2 allele burden and leukocytes, panel (b), $r^2 = 0.26$. Finally, the linear fit between thrombocytes and leukocytes shown in panel (c) had $r^2 = 0.006$.

daily dose of $6.43\mu\text{g}$. Initially blood-cell counts were measured after two weeks, four weeks and eight weeks, before the frequency was reduced to every eight to 13 weeks. With a minimum of eight weeks between measurements, we assume the difference between high-frequency low-dose and low-frequency high-dose to be insignificant, and hence that our simplification is appropriate on the time-scale considered. Typically IFN doses were given once a week, allowing for eight cycles of treatment between all measurements except in the initial eight weeks.

3.2 Empirical modelling of JAK2^{V617F} dynamics

Our initial work with the DALIAH trial data consisted of an empirical study of the dynamics of the JAK2 allele burden during standard-of-care IFN monotherapy. The results are described in (Pedersen et al., 2020), on which this section is based. For all details, we refer the reader to the published article and the corresponding supplementary material, available as Electronic Supplementary Material 1. The response of the blood-cell counts or the changes in IFN dosing were not considered.

A small set of data from five untreated MPN-patients were also considered in the study. These patients were all previously treated for MPN and the data were obtained in an extended period before treatment was re-initiated. The exact details are available in the article (Pedersen et al., 2020). Although the data-set was limited, (in particular, two patient had very few measurements taken, and were only used for validation), it provided a rough estimate for the general behaviour of the JAK2 allele burden in untreated patients. By pooling data together into a single data-set and fitting a functional expression for exponential growth to the data, an estimate of the growth was determined as $J(t) = J(0)e^{0.49t}$ where J is the JAK2 allele burden, and t is the time in years. This corresponds to a period of doubling of approximately 1.4 years. This growth estimate is included in figures 3.4 and 3.5 as a dotted black line. As the JAK2 allele burden is a measure of the ratio of alleles with the mutation to total number of alleles, JAK2 allele burdens greater than 100% is meaningless. Because of this, and since our estimate of the growth was determined for patients with allele burdens $< 50\%$, we only consider the expression to be valid for low JAK2

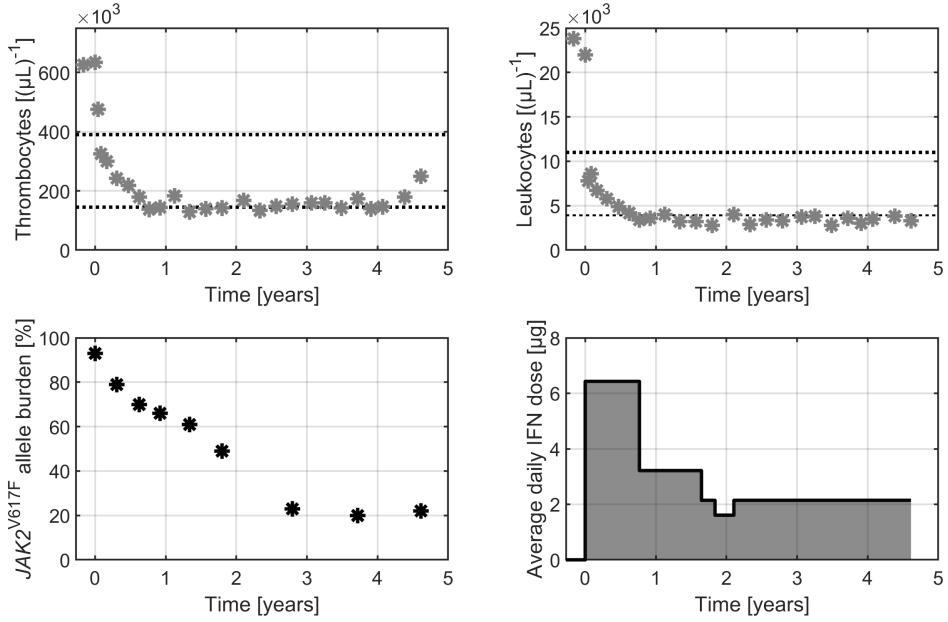


Figure 3.3: Illustration of the immediate decay in blood-cell-counts of IFN treated patients. In the top two panels, data for thrombocyte and leukocyte counts are shown as grey asterisks *, along with the healthy intervals defined as between $145 \cdot 10^3$ and $390 \cdot 10^3$ for thrombocytes and $4 \cdot 10^3$ to $11 \cdot 10^3$ for leukocytes (in cells per microliter). Note that a single “pre-baseline” measurement was available for this patient. Both cell-counts show a quick decrease, reaching the lower part of the interval within one year of treatment. In contrast, the $JAK2$ allele burden displayed in the bottom left panel shows a slower decrease, remaining above 50% for almost two years. The average daily IFN dose is depicted in the bottom right panel. For this particular patient (Patient “P075”), the dose was not changed, but the timing was varied, from 45 μg once a week to every 14 days, every 21 days, and briefly to once every 28 days.

allele burdens.

Investigating the treatment-data of the IFN-treated patients, we found that some patients had a response to treatment such that the JAK2 allele burden decayed exponentially during treatment, while the JAK2 allele burden showed an initial increase before decaying exponentially in other patients. A smaller group of patients had responses that were neither of the two response-types. To obtain estimates of responses on a patient-specific levels, we fitted a function of exponential decay to data from each of the patients. Similarly, a function of bi-exponential decay was also fitted to the patient-data. The bi-exponential decay was formulated such that continuity of the slope at $t = 0$ is ensured, with an initial slope related to the growth-expression described above. For this reason, we used the functional form:

$$J_b(t) = J_0 \cdot \left(\frac{\beta^2 + c^2 + \nu}{c^2} \cdot e^{-\beta^2 t} - \frac{\beta^2 + \nu}{c^2} \cdot e^{-(\beta^2 + c^2)t} \right) \quad (3.1)$$

where ν is the growth-rate before treatment is initiated, J_0 is the JAK2 allele burden at treatment initiation (baseline measurement) and β and c the fitting parameters. This form ensures positivity of β and c . From the growth-expression above, $\nu = 0.49$ was used.

Considering the distribution of the parameters fitted in the patient-specific responses, we determined a population-level set of parameters that describe a significant proportion of the patient-responses to a satisfying degree. When choices relevant to the determination of the population-level response has to be made, the conservative estimate was chosen, that is, the estimate with the slowest decay, to avoid over-estimating the effect of treatment. We refer the reader to the supplementary material of (Pedersen et al., 2020) for further details, available as Electronic Supplementary Material 1. The resulting population-level responses for IFN-treatment were estimated as

$$J_m(t) = J_0 e^{-0.46t} \quad (3.2)$$

for patients exhibiting a mono-exponential response, and

$$J_b(t) = J_0 (1.32e^{-0.74t} - 0.32e^{-4.61t}) \quad (3.3)$$

for the bi-exponential response-type. The population-level growth and the two population-level responses are illustrated in figure 3.4, together with exemplary data.

Having both an estimate of the development of the JAK2 allele burden for MPN-patients during IFN mono-therapy and without therapy, we were able to consider the consequences of initializing treatment at different points of disease progression as well as that of halting treatment. Figure 3.5 displays the growth of the JAK2 allele burden starting at 1%, as well as the estimated patient responses if treatment was initiated 7, 8 and 9 years later respectively. Both the mono-exponential response of equation (3.2) and the bi-exponential response of equation (3.3) are shown. The figure illustrates how postponing treatment can lead to a longer course of therapy than when it is initiated early. Additionally, it illustrates how the patients for which the JAK2 allele burden develops following the bi-exponential response show only little sign of improvement in the first half-year of treatment, but after about a year of treatment has a decreased JAK2 allele burden compared to patients with a mono-exponential response. While the response of the individual patient can differ greatly from the population-level responses, the results suggest that the potential for successful treatment with IFN can be difficult to determine if one only considers measurement of the JAK2 allele burden taken within the first year of treatment.

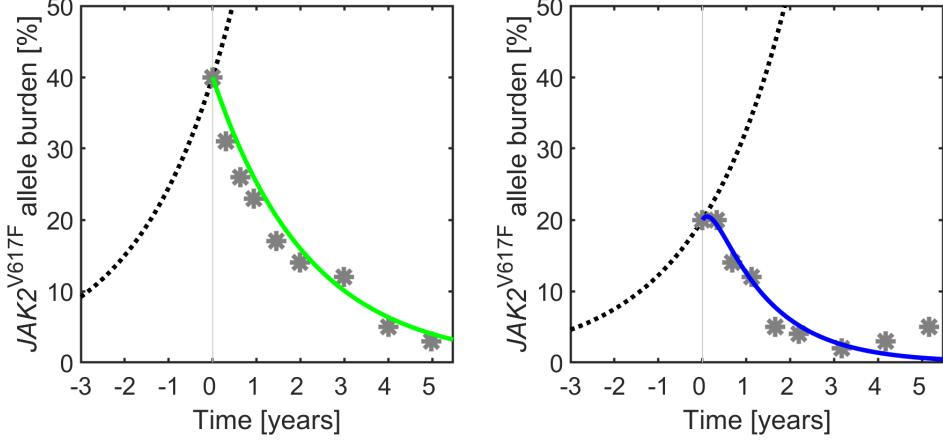


Figure 3.4: **Patient-data example with population-level expression.** The left-hand panel shows the mono-exponential decay of equation (3.2) shown in green, with exemplary patient-data shown as grey * (Patient “P016”), while the right-hand panel shows the bi-exponential expression of equation (3.3) with exemplary patient-data shown as grey * (Patient “P028”). In both panels, the JAK2 growth expression is depicted as a dotted black line, time-shifted to coincide with the baseline measurement. Note that the depicted curves are population-level expressions and not fitted to the particular patient-data. The patients chosen are however particularly exemplary of the given response types.

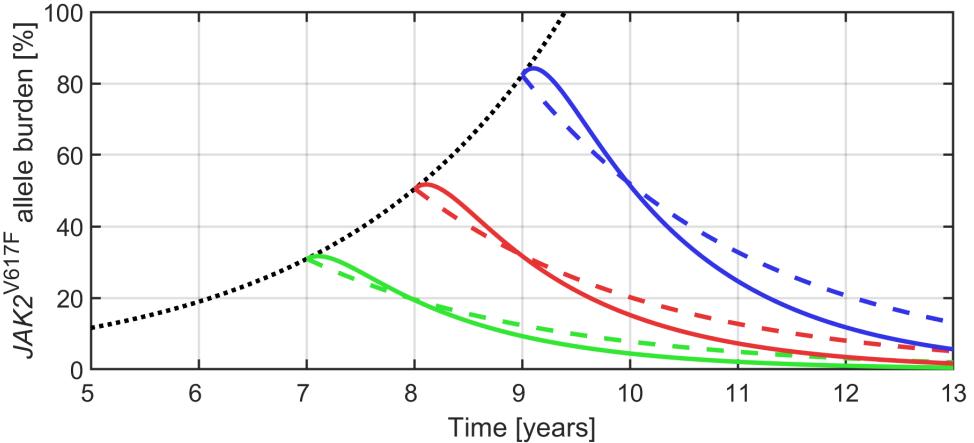


Figure 3.5: **Summarizing figure of the population-level responses for MPN-patients receiving IFN monotherapy.** The untreated growth is shown as a dotted black line, starting from a JAK2 allele burden of 1% at time $t = 0$, not shown. The two IFN monotherapy population-level responses, mono- and bi-exponential, are shown in dashed and full lines respectively. The three colors depict treatment initiated at different points of disease, namely 7, 8 and 9 years after an JAK2 allele burden of 1%, shown in green, red and blue respectively. The figure is similar to Figure 4 of (Pedersen et al., 2020).

3.3 Summarizing discussion

The findings described above represent a first estimate for the dynamics of the JAK2 allele burden in IFN-treated patients. Longitudinal studies such as the DALIAH trial are rare, and hence similar descriptions of the JAK2 allele burden kinetics are typically based on data with very few measurements of the same patients. Any prediction about a “good response” or the expected time-course of treatment has clinical importance, even a coarse estimate as the one described above. There are however a few short-comings of our approach that we briefly state here. The models described above (exponential growth, exponential decay and bi-exponential decay) were all chosen ad-hoc based on visual inspection of data. The models provide statistically reasonable descriptions about the dynamics of JAK2 allele burden measurements, but fails to provide any explanation or reasoning behind the dynamics observed. Thus they elucidate the question of “how” the JAK2 allele burden develops but not the question of “why” it develops in the way it does. Developing more sophisticated and complex mechanism-based models can be a method for answering such questions. Such models is the focus of the rest of this thesis. The way IFN-treatment is considered in the presented approach is oversimplified. All patients from the DALIAH trial randomized for IFN were receiving standard-of-care with IFN. However, as described above, standard-of-care allows for changes in dosage of IFN, and hence patients did not receive the same dose of IFN throughout the entire study. These changes and the conditions required for a change were standardized and included in the formal description of the DALIAH trial. Because of these changes to IFN-dose, the statistical models described above do not directly describe the response of the JAK2 allele burden dynamics to a given dose of IFN, but rather describe the response to “IFN-monotherapy standard-of-care”. Hence the effects observed could be caused by changes in treatment-timing and dosing (in agreement with the standard-of-care) rather than solely due a biological response to the drug. As the exact details of timing and dosing was recorded in great detail, a more sophisticated study of the JAK2 allele burden dynamics ought to consider the dosing and resulting concentration of the drug within the body. Our choice of how to do this is discussed in section 6.1.

While it was possible to describe the dynamics of the JAK2 allele burden on a population level for a group of the patients, the individual patients showed different types of responses. Differences in patients responses are normal for many disease. In particular, Lewin et al. (2020) considers solid oropharyngeal cancer tumors, and illustrates four different types of response to therapy: “fast responders” with an immediate decay in tumor size, “poor responders” showing sustained tumor size, “plateaued response” where the tumor size decays toward a particular non-zero level and “pseudo progression” where the tumor initially increases in size, before decreasing to a size smaller than the original tumor size. Each of these types of responses also appear in the JAK2 allele burden for some of the patients in the DALIAH trial. These are shown in figure 3.6. While the causes for these responses in MPN patients are unclear, the existence of different types of patient responses suggests that a grouping of patients based on response-types is possible. Such grouping could perhaps reveal genetic differences between patients, and could possibly lead to improved stratification and precision of diagnoses.

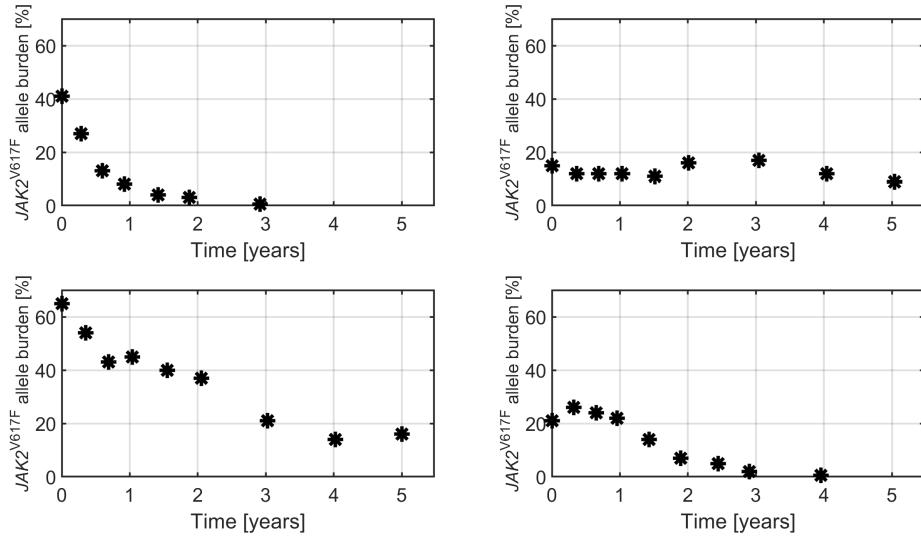


Figure 3.6: Different types of response for DALIAH patients. Four exemplary patient of the DALIAH trial that show responses in the JAK2 allele burden similar to the types considered by Lewin et al. (2020). Top left depicts a fast respond (Patient “P184”), top right a poor responder (Patient “P106”), bottom left a plateaued response (Patient “P140”) and bottom right depicts a response of the pseudo progression type, showing an initial increase in JAK2 allele burden (Patient “P113”). Note that all four patients experienced blood-cell counts that decreased and stayed within the healthy interval (not shown here), and hence treatment improved patient outcome in all four cases.

Part III

Mathematical modelling of MPNs

Chapter 4

Brief review of mathematical models of blood cancers

Using mathematics to understand and interpret biological systems has lead to important findings in the past (Mackey and Maini, 2015) and the field of mathematical biology shows great promise for the future. Mathematical modelling of cancer provides a tool for not only understanding the malignancies, but also for improving treatment of patients and hence potentially save lives. The field of mathematical oncology entails many different disciplines such as e.g. optimization of treatment protocols (Poleszczuk and Enderling, 2018; Enderling et al., 2018; Jarrett et al., 2020), modelling the spatial composition of a tumour (Lewin et al., 2020), modelling the evolution of a tumour (Gerlee and Anderson, 2015) or models of evolutionary dynamics of cancer cells (Gatenby and Brown, 2020). An overview of many of these disciplines and suggestions for the path forward for research was described in great detail in the 2019 Mathematical Oncology Roadmap (Rockne et al., 2019).

In this section, we describe some of the important work within the field of mathematical modelling of blood cancers. The scientific literature on this topic is vast, and hence we only highlight some of the work that is particularly relevant in regards to this thesis. Although a significant part of the work described below relates to blood cancers different from MPNs, the structure of the hematopoietic system suggests that findings about other blood cancers can be relevant for MPNs as well.

In a recent review, Clapp and Levy gave an overview of mathematical models of blood cancers and lymphoma (Clapp and Levy, 2015). For a review of more general cancer modelling, Altrock et al. (2015) provides a thorough review of mathematical models of different cancers, touching on e.g. solid tumors, scheduling of treatment-doses and game theoretical considerations. Similarly, the mathematical oncology roadmap mentioned above also provides a detailed review of the state of the art of mathematical oncology (Rockne et al., 2019).

Mathematical modelling of blood cancers traces its roots back to at least the 70's with the seminal work of (Mackey, 1978). The proposed model consists of a system of delay differential equations describing the HSC in either the resting G_0 phase or a proliferating phase. Through the years, the simple model and related model extensions have shown great versatility, with applicability in relation to e.g. periodic CML (Colijn and Mackey, 2005b), cyclic neutropenia (Colijn and Mackey, 2005a) and characterizing the kinetics of HSC numerically (Mackey, 2001). Dale and Mackey (2015) discuss the significance that this particular mathematical model has had within the research of cyclic neutropenia and in regards to building bridges between clinical and mathematical understanding.

A stochastic model of hematopoiesis and the mechanisms of HSC was proposed by (Abkowitz et al., 1996). In the model, stochastic processes of HSC were simulated and found to agree with experimental data of the production of progenitors for different HSC clones in cats. The model was later applied to show that HSC numbers were similar for both mice and cats and argue that, by extension, similar counts of HSC are expected in the human hematopoietic system (Abkowitz et al., 2002). In later work, the authors estimated the replication rate of HSC using the same model (Catlin et al., 2011). While these results are not directly related to blood cancers, they provide invaluable insight into the mechanisms and processes within the human hematopoietic system in a disease-free setting and give oft-cited numerical estimates for the rates of HSC-related processes.

The hierarchical structure of the blood system, with several distinct levels of cell-maturity, lends itself to mathematical models in which the distinct levels of differentiation are considered separately. This can lead to simple frameworks which allows for both estimating cell counts and properties based on the limited experimental data available, and even for comparison between different mammals (Dingli and Pacheco, 2010). Michor et al. (2005) proposes a mathematical model of CML in which four levels of maturity of blood cells are described by a system of ODEs; Stem cells, progenitor cells, differentiated cells and terminally differentiated cells. Stem cells are modelled to self-renew, while more mature cells have just an influx from the previous step of maturity and a rate of apoptosis. By modelling malignant cells in a similar way, the model displays a biphasic decline of BCR-ABL disease burden during treatment, one of the primary features of treatment of CML with the drug Imatinib. In addition, Michor et al. (2005) are able to both capture the effect of treatment discontinuation and to investigate how cellular resistance to therapy can influence the treatment outcome. A refined model is proposed by Dingli and Michor (2006), in which just two levels of maturity are considered; HSC and mature (terminally differentiated) cells. From this simple model it is possible to determine that therapy targeting mature cells cannot lead to long-lasting cure, and neither can therapy which decreases the production of malignant mature cells. Rather, therapy must inhibit replication or increase the apoptosis-rate of malignant stem cells, or as aptly put by the authors in the title of the article: “Successful therapy must eradicate cancer stem cells” (Dingli and Michor, 2006). While this conclusion appears simple, the mathematical model of (Dingli and Michor, 2006) provides a supporting argument for certain types of treatment and against other forms of therapy or treatment of symptoms. Considering cellular quiescence of HSC as well as immune system feedback, Clapp et al. (2015) introduces a similar model of the response to Imatinib in CML patients. The added complexity allows Clapp et al. (2015) to model a phenomenon of oscillating disease burden resulting from sustained treatment with Imatinib. These oscillations can be modelled and interpreted on the level of individual patients. The investigations of the model and comparison to data suggests that carefully timed immunotherapy in combination with Imatinib, can help patients avoid relapse.

Other approaches of mathematical modelling have also been considered to describe the dynamics of CML, in particular during Imatinib treatment. Roeder et al. (2006) presents an agent-based model in which healthy and malignant stem cells switch between two states; actively cycling and in the G_0 cell-phase, dependent on their cellular microenvironment. The actively proliferating cells give rise to mature cells following an amplification in numbers as progenitors. The model has previously been used to investigate clonal competition (Roeder et al., 2005). In the CML setting, the model is used to investigate the specific effect Imatinib has on malignant stem cells, suggesting that complete cure of the malignancy could be attained through Imatinib therapy. By also modelling the effects of IFN treatment in the model, Glauche et al. (2012) are able to consider combination therapy which suggests that IFN-induced stem cell activation can be beneficial for Imatinib-treated CML patients. Apart from other important findings, the

model has also been reformulated as both a system of deterministic difference equations (Kim et al., 2008b) and as a PDE model (Kim et al., 2008a). Further simplifications were proposed and analysed by (Besse et al., 2017), suggesting that optimal treatment in the long term might be different from optimal treatment in the short term. Hence the model suggested by Roeder et al. (2006) continues to have great importance on the field of mathematical modelling of blood cancer, in particular in relation to CML and treatment with Imatinib.

Stiehl et al. (2014) demonstrates how a mathematical model can be related to clinical data of stem cell transplantation and suggests that increased doses of transplanted cells can be beneficial for some patients. Using the model, a lower bound of the size of the dose required for successful transplantation can be computed on the level of individual patients. The model, originally proposed by the same authors in (Marciniak-Czochra et al., 2009), describes the different levels of maturity, where cells replicate on a faster time-scale the more differentiated and mature they are. Feedback signalling from the blood-cells affects the rate of self-renewal along all steps of maturity, which can regulate the increasing cell-numbers to obtain a stable production of blood with slowly dividing HSC at the root of the system. In later work, the model was extended to consider the healthy HSC clone and the leukemic clone separately, allowing for more direct investigation of how the resulting cell-lines interact and respond to feedback-signalling (Stiehl and Marciniak-Czochra, 2012). Using the model to identify and estimate the properties of leukemic stem cells on a patient-specific basis revealed that both frequent self-renewal and proliferation leads to poor patient outcome (Stiehl et al., 2015). This was found to not only predict patient survival but also to determine if the malignancy has worsened following relapse. By investigating a modified model where malignant cells do not respond to cytokine feedback, Stiehl et al. (2018) were able to distinguish between two hypothetical sub-types of AML. Fitting both a version of the model with cytokine-dependence and a version without to patient-specific data, the authors found that overall patient survival was poorer when only one model agreed with data. In a recent paper, an extended model was applied to the challenge of risk-stratification of patients. The model was used to predict patient survival and disease-free survival accurately, demonstrating that mathematical models can be beneficial in the clinical assessment of patient prognosis (Stiehl et al., 2020).

Mathematical modelling has been used to investigate how resistance to different therapies can be avoided through combination therapy, for different malignancies, such as e.g. CML (Komarova and Wodarz, 2005) or chronic lymphocytic leukemia (Komarova et al., 2014). Modelling resistance of CML therapy, Komarova and Wodarz (2007) finds that for treatment that fails due to drug-resistance, the resistant clone must have appeared prior to treatment initiation. The model suggests that failure of single-drug treatment is independent of quiescence. However, by considering combination therapy consisting of cell activation followed by clearing of cells, Komarova and Wodarz (2007) find that it can be an indication of an increased tendency to enter quiescence if a population of stem cells survive such combination therapy.

HSC quiescence is considered in many different ways in the literature. Examples include a resting phase before further proliferation (Mackey, 1978), a stochastic cell-change (Komarova and Wodarz, 2007) or a microenvironmental effect on HSC processes (Roeder et al., 2006). Other authors consider HSC-specific bone-marrow niches that induce quiescence of HSC (Becker et al., 2019; Stiehl et al., 2020; Ashcroft et al., 2017). This notion highlights the significance of the properties of leukemic stem cells in relapse of patients with AML (Stiehl et al., 2020). In section 7 we discuss further details of these models of interaction between HSC and the bone-marrow niches, as the formulations and findings directly relate to the novel niche model presented in part IV of this thesis.

Mathematical models have been developed to investigate how mutations of HSC arise and lead to different malignancies. Traulsen et al. (2010) show that even though the malignant stem

Part III Mathematical modelling of MPNs

cells that give rise to CML do not have a fitness advantage compared to healthy HSC, mutational differences could provide an advantage for their progeny. This, and other related work, provides a hypothesis for why some malignancies such as CML could arise solely due to neutral drift of a mutated HSC population even without a competitive advantage of the malignant stem cells (Traulsen et al., 2013).

Competition between healthy and malignant HSC clones has been investigated by different authors. In particular, Park et al. (2019) propose and investigate a mathematical model of interacting clones with a common feedback mechanism and delayed production of mature cells. The model is applied to scenarios of bone marrow transplantation, allowing the authors to investigate how the outcome of the transplant depends on clonal composition, in particular for patients with CHIP. Importantly, (Park et al., 2019) find that competition between the clones, modelled by Lotka-Volterra competition of HSC, is a key factor in determining the outcome of transplantation.

Understanding how the interaction between HSC within the bone-marrow induces clonal competition is important, as competition has consequences for the entire hematopoietic system, both preceding and during disease. Hence, successful treatment of patients with hematologic malignancies requires insight about clonal competition of HSC, and thorough investigation of HSC dynamics is necessary.

Chapter 5

The Cancitis model

In this chapter, we present a mathematical model of MPN, referred to as the “*Cancitis model*”. Although analysis of the Cancitis model is not the focus of this thesis, its formulation and some of the results of the model analysis forms the base for the modelling work discussed in later chapters. The model was initially presented in (Andersen et al., 2017). A model extension was described in (Ottesen et al., 2019) along with a reduced form of the model. We here present the latter form, but refer the reader to (Andersen et al., 2017) for the biological details about the model formulation.

5.1 Presentation of the Cancitis model and overview of previous results.

The “Cancitis model” is a mechanism-based mathematical model of the full hematopoietic system, described through a six dimensional system of ordinary differential equations (ODEs). It aims to connect the self-renewing nature of HSC with the mature blood-cell count and feedback from the blood to the production of HSC. The base biological hypothesis is that increased blood cell counts result in an increased level of debris from cells that have undergone apoptosis. This triggers an immune response to remove the debris and upregulates the HSC self-renewal. The positive feedback on HSC production is assumed to be limited by an unspecified mechanism, possibly due to physical constraint on the bone-marrow microenvironment. This provides multiple possible interpretations of how malignant stem cells can grow in greater numbers than healthy HSC and thus a possible understanding for the rise of MPN or other hematologic malignancies.

The model describes the healthy HSC x_0 , the healthy mature cells x_1 , the malignant stem cells y_0 , the mature cells arising from the malignant stem cells y_1 , as well as the current debris of mature cells that have undergone apoptosis a and an abstract quantity of the immune system activity level s .

The model equations are:

$$\dot{x}_0 = (r_x \phi_x(x_0, y_0) s - d_{x_0} - a_x) x_0 - r_m s x_0 \quad (5.1a)$$

$$\dot{x}_1 = a_x A_x x_0 - d_{x_1} x_1 \quad (5.1b)$$

$$\dot{y}_0 = \left(r_y \phi_y(x_0, y_0) s - \hat{d}_{y_0} - \tilde{d}_{y_0} y_0 - a_y \right) y_0 + r_m s x_0 \quad (5.1c)$$

$$\dot{y}_1 = a_y A_y y_0 - d_{y_1} y_1 \quad (5.1d)$$

$$\dot{a} = d_{x_0} x_0 + d_{x_1} x_1 + \left(\hat{d}_{y_0} + \tilde{d}_{y_0} y_0 \right) y_0 + d_{y_1} y_1 - e_a a s \quad (5.1e)$$

$$\dot{s} = r_s a - e_s s + I \quad (5.1f)$$

where \cdot denotes the time-derivative. The functions $\phi_x(x_0, y_0)$ and $\phi_y(x_0, y_0)$ are defined as:

$$\phi_x(x_0, y_0) = \frac{1}{1 + c_{xx} x_0 + c_{xy} y_0} \quad (5.2)$$

$$\phi_y(x_0, y_0) = \frac{1}{1 + c_{yx} x_0 + c_{yy} y_0} \quad (5.3)$$

An illustrative compartment diagram is shown in figure 5.1, and an exemplary simulation is shown in figure 5.2.

Default parameters given by Andersen et al. (2017) and Ottesen et al. (2019) were based on estimates from the literature. For this thesis and for the work described in (Ottesen et al., 2020) minor modifications were made. The updated default parameters are shown in table 5.1. This entails a reduction of A_x , A_y , d_{x_1} , \hat{d}_{y_1} and e_a , resulting in a slower reaction to changes in the mature cell compartments, without changing the overall behaviour of the model under simulation of disease growth. The death-rate of mature cells of $1.29 \cdot 10^{-2}$ corresponds to an average life-span of mature cells of approximately 77.5 day, which is a slight overestimate compared to e.g. the life-span of thrombocytes, which is around 8 to 12 days (Singh and Singh, 2018). In addition, the growth rate and the differentiation rate of the leukemic clone was also changed slightly, such that the disease development starting from a single mutated HSC leads to a growth of relative frequency of malignant cells that resembles the JAK2 allele burden growth estimate discussed in section 3 for a relative frequency below 33%. For higher relative frequency, the exponential growth described in section 3 is assumed to overestimate the disease level.

While we do not discuss all details of the model, a few comments are warranted before describing previous work on model analysis due to Andersen et al. (2017) and Ottesen et al. (2019).

- The renewal of HSC are due to the terms $r_x \phi_x(x_0, y_0) s x_0$ and $r_y \phi_y(x_0, y_0) s y_0$. The dependence on s is due to the assumption that the immune response upregulates HSC self-renewal, as discussed above.
- HSC do not directly produce mature blood cells. An intricate system of hematopoietic progenitor cells and other differentiated cells are intermediate steps between HSC and the mature cells. These intermediate cells lead to a vast increase in numbers as they divide in a self-renewing manner, albeit to a limited degree compared to the unlimited self-renewal of HSC. In the Cancitis model, the increase in numbers are accounted for through an amplification factor on the differentiation from HSC to mature cells, A_x and A_y . This was briefly presented in the supplementary material of Bangsgaard et al. (nd), and a similar argument is explored in detail in section 10.1.

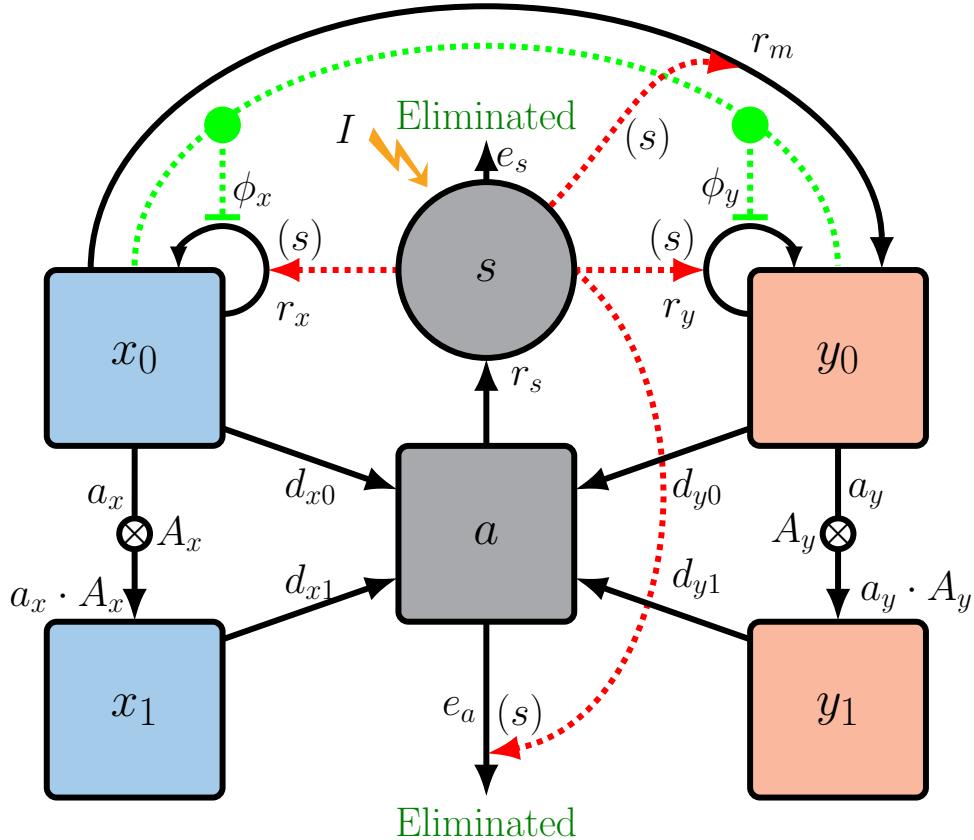


Figure 5.1: **Compartment diagram of the Cancitis model.** The healthy cells counts, x_0 and x_1 , are shown as blue boxes on the left-hand side, while the malignant cells, y_0 and y_1 , are depicted as red boxes on the right-hand side. The grey box in the middle illustrate the dead cells, a , while the grey circle illustrate the abstract notion of inflammatory level s . The arrows represents the flows between compartments, with the small circle with the \times symbol depicting a multiplication factor. The red arrows from s depict how the inflammatory level influences different rates in the system, while the green limiting arrows connecting x_0 and y_0 illustrates the limit on self-renewal (r_x and r_y) due to the expressions $\phi_x(x_0, y_0)$ and $\phi_y(x_0, y_0)$ from equation (5.2). The figure was based on a figure from (Andersen et al., 2017).

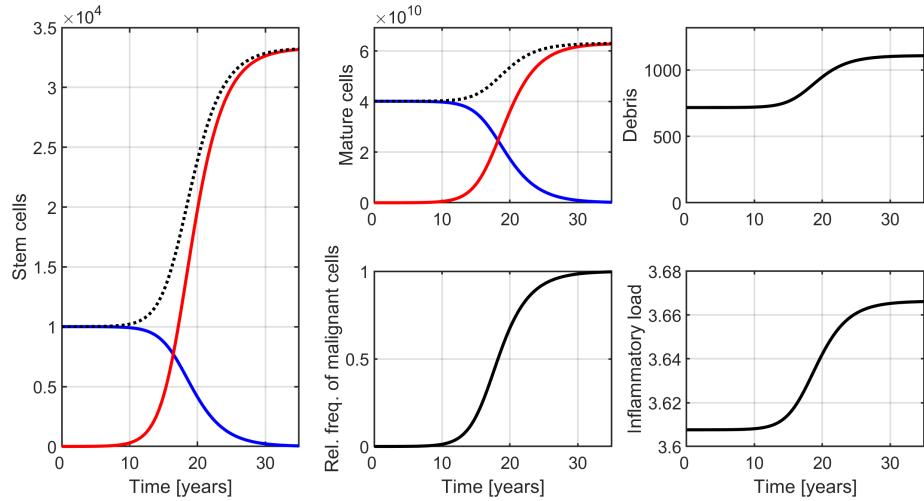


Figure 5.2: **Exemplary simulation of the Cancitis model.** The panels depict the trajectories of the six variables of the Cancitis model, equations (5.1), as well as the relative frequency of malignant mature cells, defined as $\frac{y_1}{x_1+y_1}$, in a simulation with the default parameters shown in table 5.1. For the simulation $r_m = 0$ and with initial conditions defined as the healthy steady state with a single malignant stem cell added. Blue lines depict healthy cells (x_0 and x_1), red lines depict malignant cells (y_0 and y_1), while the dotted black lines depict the corresponding sums of cells. The debris of dead cells, a , is shown in the top-right panel, while the abstract inflammatory load is depicted in the bottom-right panel.

r_x	$8.7 \cdot 10^{-4}$	day $^{-1}$	r_y	$15 \cdot 10^{-4}$	day $^{-1}$	A_x	$4.7 \cdot 10^9$	—
r_m	0	day $^{-1}$	d_{x_0}	$2 \cdot 10^{-3}$	day $^{-1}$	A_y	$4.7 \cdot 10^9$	—
\tilde{d}_{y_0}	$2 \cdot 10^{-9}$	day $^{-1}$	\hat{d}_{y_0}	$2 \cdot 10^{-3}$	day $^{-1}$	c_{xx}	$5.6 \cdot 10^{-5}$	—
a_x	$1.1 \cdot 10^{-5}$	day $^{-1}$	a_y	$0.52 \cdot 10^{-5}$	day $^{-1}$	c_{xy}	$5.4 \cdot 10^{-5}$	—
d_{x_1}	$1.29 \cdot 10^{-2}$	day $^{-1}$	d_{y_1}	$1.29 \cdot 10^{-2}$	day $^{-1}$	c_{yx}	$5.2 \cdot 10^{-5}$	—
e_a	$2 \cdot 10^5$	day $^{-1}$	e_s	2	day $^{-1}$	c_{yy}	$5.0 \cdot 10^{-5}$	—
r_s	$3 \cdot 10^{-4}$	day $^{-1}$	I	7	day $^{-1}$			

Table 5.1: **Default parameters of the Cancitis model.** The parameters as given in the supplementary material of (Ottesen et al., 2020). The values differ slightly from those given in (Andersen et al., 2017) and (Ottesen et al., 2019), see the text.

- The Cancitis model allows for mutation of HSC into malignant stem cells, through the term $r_m s x_0$. Little is known about the rate at which such mutations occur, r_m . For the purposes of this thesis, we maintain $r_m = 0$, and instead include one malignant stem cell for the initial conditions. This corresponds to a single mutational event of one HSC with the resulting population of malignant cells tracing back to the mutation.
- Apoptosis of malignant stem cells is modelled to have a second order term, $-\tilde{d}_{y_0} y_0^2$. This term was not included by Andersen et al. (2017) and was introduced in (Ottesen et al., 2019) based on considerations about immune surveillance through T-cell targeting of malignant cells. Hence, \hat{d}_{y_0} represent the naturally occurring apoptosis rate of malignant stem cells, while \tilde{d}_{y_0} describes a immune-system targeted death-rate.
- The expression of ϕ_x and ϕ_y , equations (5.2), were defined as $\phi_x(x_0, y_0) = \frac{1}{1 + (c_{xx} x_0 + c_{xy} y_0)^2}$ and ϕ_y similarly in (Andersen et al., 2017). Numerical investigation suggested that this change had little to no effect on general model behaviour for parameters close to the default values. It did, however, vastly simplify the mathematical analysis of the model.

Sajid et al. (2019) investigated that steady states of the Cancitis model, and determined criteria for existence of different types of steady states. This includes a *trivial steady state* where $x_0^* = x_1^* = y_0^* = y_1^* = a^* = 0$ and $s^* = \frac{I}{e_s}$, a *healthy steady state* where $y_0^* = y_1^* = 0$ and all other variables are positive, corresponding to a situation without disease, and a *full-blown disease steady state* where $x_0^* = x_1^* = 0$ and all other variable are positive, related to a disease stage where the healthy cells have been completely eradicated. Finally, under specific restrictions on parameters, co-existences steady states also exists, in which all variables are positive. For further details, see (Sajid et al., 2019).

The original proposal of the Cancitis model by Andersen et al. (2017) aimed to investigate the role of inflammatory stimulus on the progression of MPN. The parameter I models an external source of inflammation, assumed to depend on external factors such as e.g. disease unrelated to MPN or from smoking. Inflammation is considered a hallmark of cancer in general (Colotta et al., 2009), and in recent years it has been suggested that chronic inflammation plays an important

role in advancing MPN-progression (Hasselbalch and Bjørn, 2015). Using the Cancitis model, it was possible to illustrate the link between how inflammation and MPN, by simulating the effect of increased inflammation. In figure 5.3 a simulated scenario is shown in which a sustained increase of inflammation results in a higher count of mature blood cells. By considering an arbitrary threshold over which the risk of thrombotic events is increased, e.g. 50% higher cell counts than under disease-free circumstances, this illustrates how patient-risk is higher much earlier, even when considering minor increases in inflammation. Figure 5.3 illustrates how patient outcome is worsened significantly if the external inflammatory stimulus of the patient increase. Such increase could correspond to e.g. the onset of a smoking habit. In the scenario shown in figure 5.3, the relative frequency of the malignant clone at year 5 is below 0.1%. If the relative frequency of the malignant clone is considered indicative of the JAK2 allele burden, this would be below the detection limit of most assays. As such, even if the disease has not yet reached a detectable level, the increased inflammatory stimulus triggers faster progression of disease.

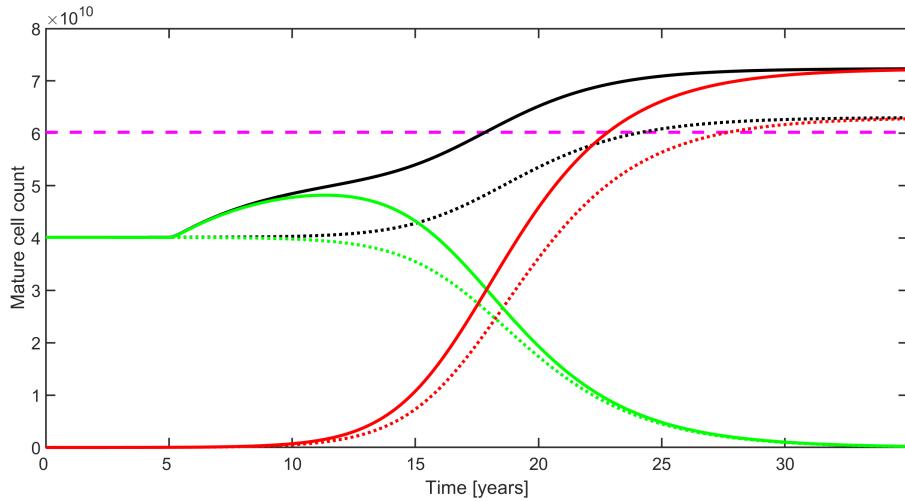


Figure 5.3: Increased inflammation cause complications to occur earlier. At time $t = 0$, the healthy steady state is perturbed by adding a single leukemic stem cells. The dotted curves display the progression of x_1 , y_1 and $x_1 + y_1$ under default parameters, in green, red and black respectively. The full curves show a scenario where external inflammation I is increased by 10% at year 5. This leads to increased cells counts, both healthy and leukemic. A threshold of 50% above the healthy steady state count of x_1 is shown as a dashed magenta line. For comparison, the default scenario crosses this threshold around year 25, while the scenario with increased inflammation crosses the threshold earlier, around year 18.

Hypothetical treatment scenarios can be simulated with the Cancitis model. Figure 5.4 shows a scenario where disease progression is reversed by simulated treatment perturbing the parameter \hat{d}_{y_0} , resulting in a period with an increased death-rate of leukemic stem cells. In the figure, two important measures of disease are illustrated: The mature cell counts and the relative frequency of the leukemic clone. As discussed previously, increased counts of blood-cells is associated with an increase risk of thrombosis and related event. The relative frequency $\frac{y_1}{x_1+y_1}$ in the model is interpreted as the JAK2 allele burden for MPN patients, and hence a clinical aim of treatment

is reduction of this measure.

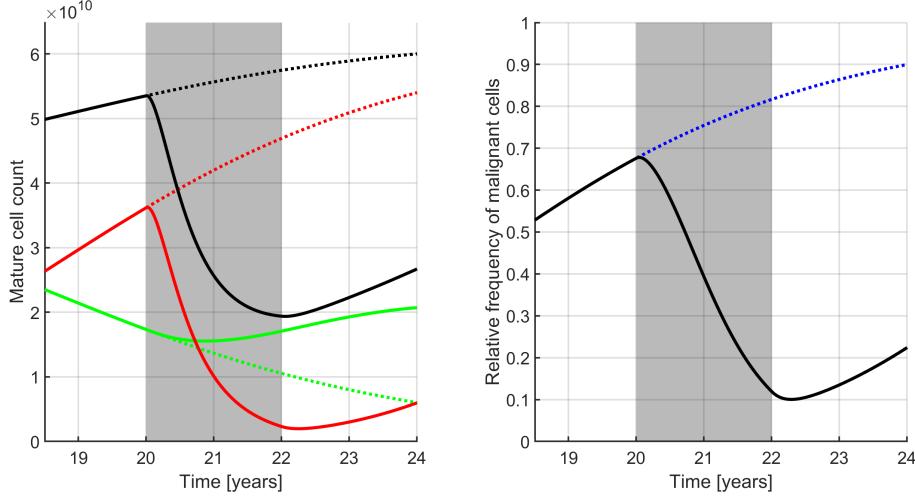


Figure 5.4: **Increasing \hat{d}_{y_0} can reverse disease progression.** A scenario with disease progression is simulated, in which one HSC x_0 is replaced by one leukemic cell y_0 at time $t = 0$. The resulting progression was shown in figure 5.2, and is shown in dotted lines here. Between year 20 and year 22, hypothetical treatment of a 4-fold increase of \hat{d}_{y_0} was simulated. At year 22, parameters were returned to the default values. The grey background in the figure shows the treatment-period. In the left panel, the full red curve shows the leukemic mature cells y_1 , the green curve shows the healthy mature cells, x_1 . The sum of mature cells is shown in black. In the right-hand panel, the relative frequency $\frac{y_1}{x_1+y_1}$ is shown in black, with the scenario without treatment shown in dotted blue. The simulated treatment results in a significant decrease in leukemic cells as well as in the relative frequency. Note that the decline in relative frequency of leukemic cells appears to continue briefly after treatment cessation, before increasing once again.

5.2 The reduced Cancitis model

In the mathematical analysis described in (Ottesen et al., 2019), the Cancitis model was investigated in rigorous detail. Through considerations of time-scales, a slow manifold approximation was identified, reducing the model to a two-dimensional ODE of scaled cell-counts of the healthy and malignant HSC, X_0 and Y_0 , with algebraic equations for the remaining variables. The model, referred to as *the reduced Cancitis model*, is given as:

$$X'_0 = \left(\frac{J + \sqrt{J^2 + 2B_x X_0 + 2B_y X_0}}{1 + X_0 + C_y Y_0} - 1 \right) X_0 \quad (5.4a)$$

$$Y'_0 = \left(R \frac{J + \sqrt{J^2 + 2B_x X_0 + 2B_y Y_0}}{1 + C_x X_0 + Y_0} - D_0 - D_1 Y_0 \right) Y_0 \quad (5.4b)$$

Parameter	Definition	Value	Parameter	Definition	Value
R	$\frac{r_y}{r_x}$	1.49	J	$\frac{I}{2e_s} \frac{r_x}{d_{x_0} + a_x}$	0.76
D_0	$\frac{\hat{d}_{y_0} + a_y}{d_{x_0} + a_x}$	1.00	D_1	$\frac{\hat{d}_{y_0}}{c_{yy}} \frac{1}{d_{x_0} + a_x}$	0.10
C_x	$\frac{c_{yx}}{c_{xx}}$	0.93	C_y	$\frac{c_{xy}}{c_{yy}}$	1.08
B_x	$\frac{a_x A_x}{2c_{xx}} \frac{r_s}{e_s e_a} \frac{r_x^2}{(d_{x_0} + a_x)^2}$	0.06	B_y	$\frac{a_y A_y}{2c_{yy}} \frac{r_s}{e_s e_a} \frac{r_x^2}{(d_{x_0} + a_x)^2}$	0.07

Table 5.2: **Parameters of the reduced Cancitis model.** The reduced parameters are given in terms of the parameter of the Cancitis model. The numerical values used in (Ottesen et al., 2019) is also shown. Note that these values are the numerical values presented in the article, which differ slightly from the values given in table 5.1, as the default parameters of the Cancitis model were modified slightly for this thesis.

where ' denotes the derivative with respect to a scale time-variable defined in (Ottesen et al., 2019). The reduced parameters of the reduced model are dimensionless compound parameters, with dependence on the original parameters of the Cancitis model in equations (5.1). The reduced parameters are shown in table 5.2. The reduced parameters J , D_0 and D_1 were focused on in the analysis described in (Ottesen et al., 2019). As shown in table 5.2, J relates to the external inflammatory stimulus, while the death-rates of the leukemic stem cells are found in the expressions for D_0 and D_1 . We identified the steady states of the reduced model and their stability. In addition, a parameter-dependent value M was determined for which $X_0 + Y_0 > M$ implies that both $X'_0 < 0$ and $Y'_0 < 0$, and as a result an attractive trapping region was determined for non-negative initial conditions. Considering only steady states with non-negative cell-counts, the reduced Cancitis model features a *trivial* steady state with vanishing cell-counts, a *hematopoietic* steady state with $X_0 > 0$ and $Y_0 = 0$, a *malignant* steady state with $X_0 = 0$ and $Y_0 > 0$ and, under certain conditions of the parameters co-existence steady states with both $X_0 > 0$ and $Y_0 > 0$ arise. Through numerical investigation, we determined that a maximum of one co-existence steady state with non-negative cell-counts exists. The resulting structure allowed us to consider the long-term effect of changes to inflammation (by perturbation of J), or the effect of sustained T-cell therapy (by increasing D_1). Since just two variables are considered, phase-diagrams of X_0 and Y_0 can provide a full view of the model dynamics for a given set of parameters. In figure 5.5 examples are shown for a range of values of J and D_0 .

The shown scenarios can be interpreted as the range of dynamics that arise under changes to external inflammation I and changes to the death-rate of leukemic stem cells \hat{d}_{y_0} . In addition, we found that the asymptotic dynamics of the system were effectively described through two expressions, which we referred to as the primary and the secondary reproduction ratios, \mathcal{R} and \mathcal{S} respectively. The local stabilities of the steady states were identified in (Ottesen et al., 2019), and hence, the reduced Cancitis model provides an estimate of the asymptotic behaviour of the full model. As such, when comparing patients responses to parameter perturbations, it is possible to determine whether a long-term treatment is to be expected from the reduced model or if the leukemic steady state remains stable and attracting as is the case without treatment.

As for the Cancitis model, hypothetical treatment can be simulated in the reduced Cancitis model. Figure 5.6 shows the same simulated treatment scenario as figure 5.4, but simulating the reduced model instead. The overall dynamics observed is very similar, as expected since

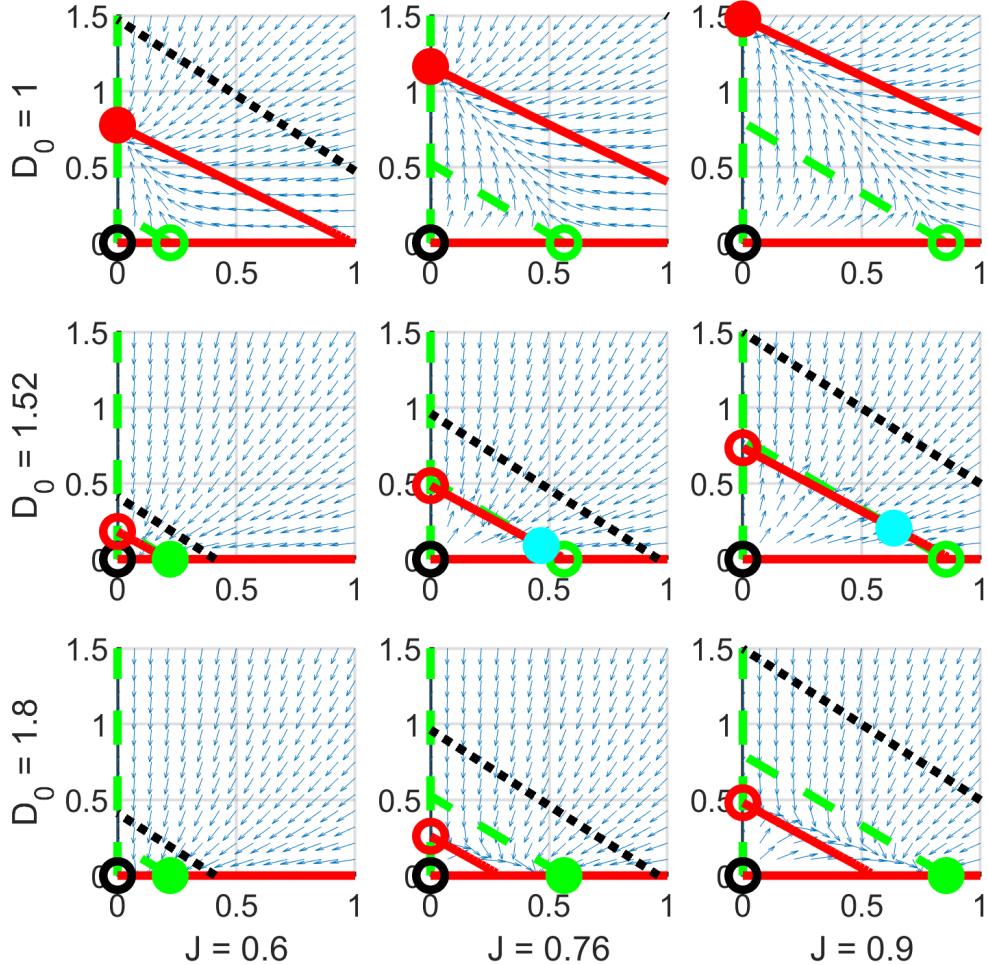


Figure 5.5: Phase-diagrams of the reduced Cancitis model for different values of J and D_0 . All panels display the dynamics of the reduced Cancitis model for a given set of parameter. The variable X_0 is along the first axes while the variable Y_0 is along the second axes. The rows of panels from top to bottom have $D_0 = 1, 1.52, 1.8$ respectively, while the columns from left to right have $J = 0.6, 0.76, 0.9$. From the biological interpretation of the parameters, left to right can be interpreted as increasing external inflammation, while top to bottom can be interpreted as increasing the death-rate of malignant cells. Circles depicts steady states, with the full circle showing the stable steady state and open circles unstable steady state. The red circle shows the leukemic steady state, the green the healthy steady state, the black circle the trivial steady state and, when non-negative, the blue circle depicts a co-existence steady state. Nullclines for X_0 and Y_0 are shown in stipulated green and full red respectively. The trapping region for $X_0 + Y_0$ is shown as a dotted black line. Small blue arrows depict the direction of flow. (Reprinted from (Ottesen et al., 2019).)

the models are closely related. The reduced Cancitis model does however show a transient response to parameter-perturbation, which is highlighted in figure 5.7. This difference is however only significant because the quasi-steady-state approximation of mature cells implies that the dynamics of mature cells in the reduced Cancitis model follows that of the stem cells. When considering the relative frequency of malignant stem cells, the response of the Cancitis model resembles the response of the reduced Cancitis model more closely.

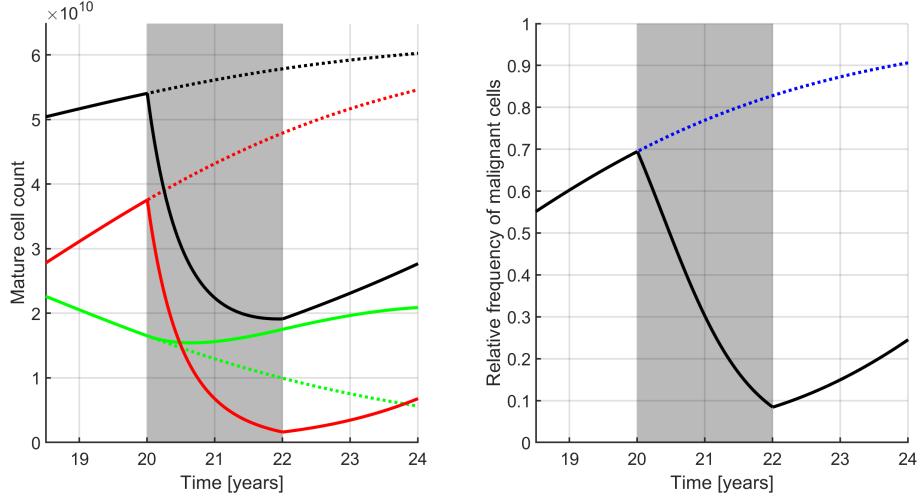


Figure 5.6: Increasing D_0 can reverse disease progression in the reduced Cancitis model. The same scenarios as in figure 5.4 is here simulated for the reduced Cancitis model. Between year 20 and year 22, the D_0 parameter was increased 3.99-fold, equivalent to a 4-fold increase of \hat{d}_{y_0} . In terms of the primary reproduction ratio \mathcal{R} , this is a change from $\mathcal{R} = 1.74$ before treatment to $\mathcal{R} = 0.44$ during treatment. For a description of colors and line types, see the caption of figure 5.4. Note that mature cell counts are calculated subsequently from the solutions of equations (5.4) rather than independently considered in the model. Following the cessation of treatment, both cell-counts and relative frequency of leukemic cells begin to increase immediately.

5.3 Summarizing discussion

The Cancitis model describes a hypothetical link between blood production and the immune system. While many different pathways play a role in feedback from the blood to the HSC, the feedback considered in the Cancitis model is of a simplified form, assumed to represent the most significant part. Considering the immune system abstractly, the relationship between increased inflammation and the progression of MPN-disease could be investigated in the general sense, in agreement with the notion of a connection between chronic inflammation and excessive production of mature cells due to malignant stem cells.

The formulation of the Cancitis model is similar to the model of Dingli and Michor (2006). The additional details about the debris of dead cells and the immune system feedback results in a model that is comparable to previous work, but adds important details about the indirect interaction of the healthy clone with the malignant clone.

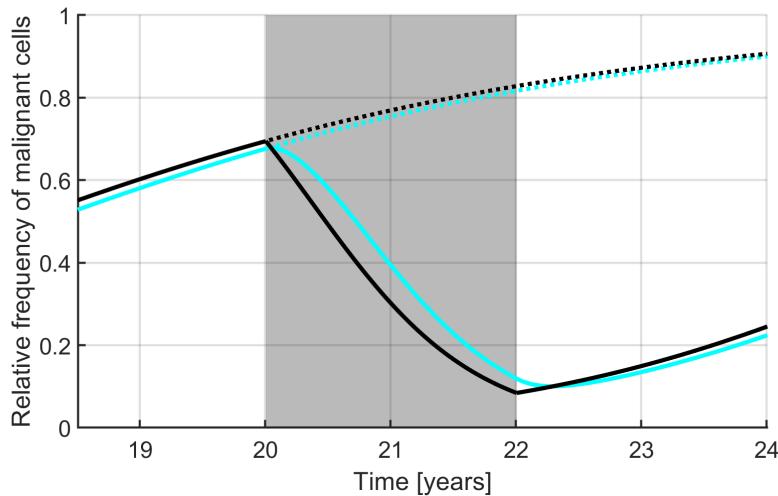


Figure 5.7: **Transient change of the death-rate of leukemic cells has a different effect in the reduced Cancitis model.** The relative frequency of malignant cells in the equivalent simulated treatment scenarios of figures 5.4 and 5.6 are shown together. The Cancitis model is shown in cyan, while the black lines show the reduced Cancitis model. The dotted lines show the scenario with no treatment. The immediate response in the reduced Cancitis model leads results in a larger difference between the two models during treatment, which decreases again after treatment cessation. Note however that in the Cancitis model, an immediate response in the relative frequency of stem cells occurs, similar to the one observed for the reduced Cancitis model. As such, the delayed reaction of the mature cells is a particular feature missing because of the quasi-steady-state approximation that gave rise to the reduced Cancitis model.

Steps of cell maturity acting as intermediate between the stem cells and the mature cells were not directly considered in the Cancitis model. Instead, a linear relation between the production of progenitors and the subsequent production of mature cells was assumed. A simple extension of the model could be considered where multiple intermediate stages of cell maturity are explicitly modelled. However, if the difference between malignant and healthy progenitors cells is insignificant, this has little effect on the system. In section 10.1, we argue for the validity of this assumption, and show that it relates to the rate that progenitors differentiate.

A model reduction was presented in (Ottesen et al., 2019), leading to the reduced Cancitis model, where only the stem cells are modelled dynamically. This allowed for identification and classification of steady states, hence classifying the asymptotic dynamics of the model. Two important conglomerations of parameters, the primary and secondary reproduction ratios, \mathcal{R} and \mathcal{S} , were found to be the primary determinant of disease progression.

By considering a specific hypothetical treatment, it was possible to simulate how the Cancitis model and its reduced counterpart behave during and after a particular biological change, namely an increase in the death-rate of leukemic cells. Both models showed that such change could lead to a substantial decrease in disease burden, both in terms of the mature cell-counts and in the relative frequency of the malignant clone. As such, the models suggest that increasing the death-rate could be one way to improve patient outcome. In particular, the dynamics of the relative frequency observed in figures 5.4 and 5.6 appears visually to agree with the changes in the JAK2 allele burden observed for IFN-treated MPN patients in the DALIAH trial, see section 3.

While the Cancitis model provides an important view into the dynamics of blood cancers and allows for a possible explanation behind the treatment-responses seen in data (as discussed in detail in the next section), some significant features of the hematopoietic system has been omitted from the model for simplicity. The abstract notion of the immune system as solely up-regulating the HSC production omits the role of immuno-surveillance, where malignant cells are actively targeted by the immune system and cleared. Such effects were partly included in the extension of the model presented in (Ottesen et al., 2019), through the term $-\tilde{d}_{y_0} y_0^2$ in the differential equation for \dot{y}_0 . However, more detailed effects of the immune system could be considered, especially since HSC also give rise to the cells of the immune system, and hence, an intricate feedback between the production of e.g. T-cells and the self-renewal and differentiation of HSC is expected. In the Cancitis model, any non-zero number of malignant stem cells will eventually lead to full-blown malignancy, assuming they have an advantage over the healthy clone, which in the reduced model were determined by the reproduction ratios \mathcal{R} and \mathcal{S} . Adding details of the immune system modelling could lead to a system where sufficiently few malignant cells could be completely eradicated.

The cell cycle of HSC play an important role in both hematopoiesis and malignancy, and has been modelled extensively. As discussed in section 4, HSC can remain quiescent for extended periods of time. No such dynamics of HSC are considered in the Cancitis model, and rather all HSC are assumed to behave equally. As a result, the self-renewal rate, the death-rate and the differentiation-rate are all “effective” rates, describing the average rate at which the related processes occur, when all HSC are considered. This is assumed to have minimal effect when considering the general behaviour of the entire system, as differences would average out. However, if feedback or treatment were to interact with the HSC and their tendency to enter, remain or exit quiescence, this could have an important effect on the model dynamics. In particular, this is known to be the case of IFN, which is believed to stimulate quiescent HSC and induce them to enter cell cycling. The lack of this feature of quiescence is a limitation of the Cancitis model when considering IFN-treated patients from the DALIAH trial. However, it is unclear how big the effect is, and whether the perturbing the effective parameters is in fact sufficient to capture the dynamics expected to arise from activation of quiescent HSC.

Chapter 6

Modelling MPN-patient responses with the Cancitis model

The Cancitis model defined in the previous chapter describes the hematopoietic system during MPN development on a conceptual level. In this section the model is related directly to clinical data from the DALIAH trial. The work was previously described in the supplementary material of (Ottesen et al., 2020), available as Electronic Supplementary Material 2. This section describes the work in more detail, and includes some additional results.

Throughout this section, the reduced Cancitis model is used, to reduce model complexity when comparing the model with data. To maintain the simplicity of the reduced Cancitis model but still have the biological interpretation of parameters, an equivalent formulation is used, consisting of the algebraic expressions:

$$x_1 = \frac{a_x A_x}{d_{x_1}} x_0 \quad (6.1a)$$

$$y_1 = \frac{a_y A_y}{d_{y_1}} y_0 \quad (6.1b)$$

$$a = \frac{1}{2} \sqrt{\left(\frac{I}{r_s}\right)^2 + 4\frac{e_s \kappa}{e_a r_s} - \frac{I}{2r_s}} \quad (6.1c)$$

$$s = \frac{r_s}{e_s} a + \frac{I}{e_S} \quad (6.1d)$$

where $\kappa = d_{x_0} x_0 + d_{x_1} x_1 + (\hat{d}_{y_0} + \tilde{d}_{y_0}) y_0 + d_{y_1} y_1$. The differential equations of the reduced Cancitis model can then be written as:

$$\dot{x}_0 = (r_x \phi_x(x_0, y_0) s - d_{x_0} - a_x) x_0 \quad (6.2a)$$

$$\dot{y}_0 = (r_y \phi_y(x_0, y_0) s - \hat{d}_{y_0} - \tilde{d}_{y_0} y_0 - a_y) y_0 \quad (6.2b)$$

For the rest of this section, “the Cancitis model” refers to this particular formulation of the model.

6.1 Response to IFN treatment in the Cancitis model

In section 3, we discussed how changes in IFN dosing and timing could influence the outcome of treatment and hence must be considered in greater detail in more extensive modelling work. Before the patient-data is related to the Cancitis model, we first describe how IFN is administered and how the effect of changes in dosing is considered in the model.

The field of pharmacokinetics and pharmacodynamics is vast and there exists many different methods for modelling the subcutaneous administration of IFN. We consider a simple model of IFN uptake to avoid over-parametrization. As an initial assumption, the different timings (e.g. “ μg per week”, “ μg per 21 days”, etc.) were simplified into simply an average daily dose, as described in section 3. It is possible that there is a difference in the immediate effect when larger doses are administered on a infrequent basis compared to frequent small doses. However, we assume that these effects are insignificant in the long-term and that the subcutaneous administration leads to a slow absorption into the blood system, allowing for a simple model of the blood-concentration of IFN. The uptake and clearance of IFN in the bloodstream is assumed to occur with the same rate, τ , independently of the current dose. This leads to a simple expression of IFN blood-concentration:

$$\dot{B} = \tau (D(t) - B(t)) \quad (6.3)$$

where $D(t)$ is the average dose at time t and $B(t)$ is the blood concentration at the given time. Inspired by previous work on pharmacokinetic modelling of IFN (Saito et al., 2012), we chose $\tau = \frac{1}{7} \frac{\mu\text{g}}{\text{day}}$. Note that $D(t)$ is a stepsize constant function. Considering an interval $t \in \{t_0, t_1\}$ where $D(t)$ is constant, $D(t) = D_0$, the differential equation (6.3) has solution:

$$B(t) = D_0 - (D_0 - B(t_0)) e^{-\tau \cdot t} \quad (6.4)$$

for $t \in \{t_0, t_1\}$. The function $B(t)$ describes the total amount of IFN in the blood of the patient. A continued daily dose of e.g. $5\mu\text{g}$ IFN, leads asymptotically to $B(t) = 5$. For this reason, the unit μg is used for $B(t)$ throughout this thesis to describe the equivalent IFN dose, even though the actual concentration is dependent on the volume of the blood of the patient. The expression in equation 6.4 allows for a very simple estimate of the blood concentration level. In figure 6.1 an example of $B(t)$ for a particular patient is illustrated. The IFN dosing and timing, $D(t)$ for the given patient is shown in gray, while a step-wise calculation of $B(t)$ using equation (6.4) is shown in black.

The effect of the IFN concentration in the bloodstream is assumed to affect certain parameters of the Cancitis model. To model this, we define the IFN-perturbed parameter, $\check{\theta}$, of any given parameter, θ , as:

$$\check{\theta}(t) = (1 + \nu_\theta B(t))\theta \quad (6.5)$$

where ν_θ determines the effect of IFN on the given parameter θ . Note that in general $\nu_\theta \geq -B(t)$ must hold for all t , so $\check{\theta}(t) \geq 0$ is fulfilled. For simplicity, we require $\nu_\theta \geq 0$ when fitting the model to patient data, and hence only consider parameters that are increased by heightened IFN blood-concentration. The relation between IFN blood-concentration and parameter value is hence a simple linear relation, with $\hat{\theta} = \theta$ for a blood-concentration of $B(t) = 0$.

6.2 Interpretation of JAK2^{V617F} allele burden in the Cancitis model

As discussed in section 2.1, the JAK2 allele burden is an important measure for the disease progression of MPN. While measurements of HSC give the most accurate description of a given

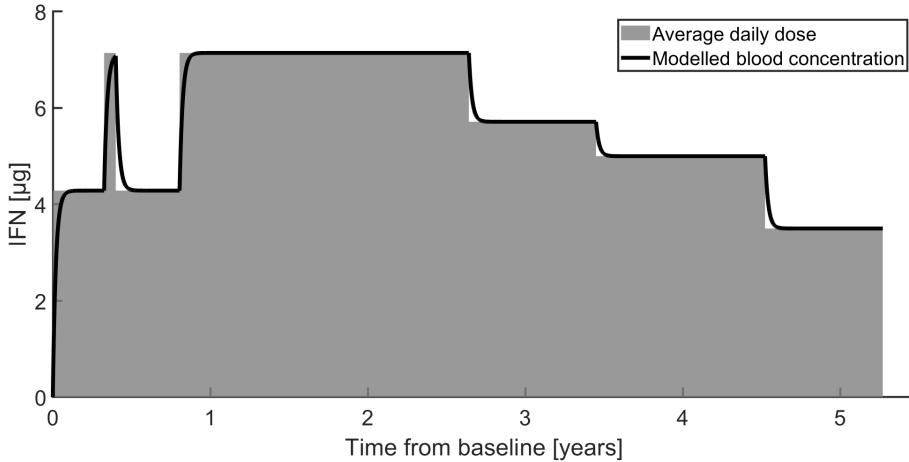


Figure 6.1: **Illustration of estimation of IFN blood-concentration** The figure shows the current daily IFN-dose for an exemplary DALIAH patient as grey rectangles. The step-wise solution of the IFN blood-concentration, equation 6.4 $B(t)$, is shown as a black line overlaid on the figure. This illustrates how changes in IFN blood-concentration can be delayed compared to changes in dosage. Throughout this thesis, similar patient-specific curves of $B(t)$ are shown along with figures of data, instead of the raw data for the dose-timing.

mutational burden, human studies must often rely on measurements of blood samples. Particularly for the JAK2 mutation, a significant correlation between HSC and blood-cell measurements has been shown (Takahashi et al., 2013) and hence for this particular disease burden, measurements of blood samples provide a reliable estimate for the JAK2 allele burden of HSC. For the DALIAH trial data, described in section 3, such blood sample measurements were made. Thus, to compare the Cancitis model to the data from the DALIAH trial, we consider the mature cell-counts. Considering the JAK2 allele burden a measure of the proportion of cells that arose from mutated stem cells, we define:

$$C_{cancitis}(t) = \frac{y_1(t)}{x_1(t) + y_1(t)} \quad (6.6)$$

as the measure of JAK2 allele burden in the Cancitis model, i.e. simply the proportion that mature malignant cells make out of all mature cells. This relative frequency of malignant cells in a typical scenario of disease progression was included in figure 5.2.

6.3 Fitting procedure for $JAK2^{V617F}$ allele burden

For the default parameters given in table 5.1, the Cancitis model agrees with the exponential growth described in section 3. In the Cancitis model, the growth is achieved by replacing a single healthy HSC with one malignant stem cell at time $t_0 = 0$. The disease development was assumed to be the same for all patients. This was necessitated by the lack of data available for patients before treatment initiation. Hence the baseline measurement is the only measurement we used to determine how far the patient is on the population-level curve of disease development. The common disease development was simulated from a single malignant cell added initially until

the JAK2 allele burden approached 100%. For each patient, the disease development curve was time-shifted such that it agreed with the JAK2 allele burden at baseline. This provided an estimate for the behaviour of the model leading up to the baseline measurement and of the value of the model variables at treatment initiation. Note that this does however imply complete agreement between the baseline JAK2 allele burden measurement and the model, and hence assumes complete confidence in the baseline measurement. For most patients, this assumption is found to be appropriate, however, for few patients the JAK2 allele burden at baseline is much different from later measurement. In particular, patient “P089” initially had an JAK2 allele burden of 19%, but all other measurements the first two years gave JAK2 allele burden above 60%. While this could be treatment induced, it appears unlikely.

Having determined the patient-specific variable states at treatment initiation, the model was simulated for the period of time available in the DALIAH trial data for the given patient. By considering perturbation of specific parameters in accordance to the IFN-response given in equation (6.5), the treatment response could be simulated in the model as affecting different parameters to different degrees, by varying the relevant ν parameters. Hence it was possible to consider scenarios where a subset of parameters were perturbed while the remaining parameters remained unchanged.

The model-error was defined as the difference between the JAK2 allele burden measurements and the relative frequency of malignant mature cells in the model, equation (6.6), at the given time:

$$E_{jak,i} = C_{cancitis}(t_i, \nu_\theta) - J_i \quad (6.7)$$

where J_i are the measurements of the JAK2 allele burden and time t_i and $C_{cancitis}(t_i, \nu_\theta)$ are given by evaluating equation (6.6) at time t_i when solving with the specific value of ν , related to parameters θ . The sum of squared errors, $\sum_{i=1}^n E_{jak,i}^2$, could then be minimized for the chosen parameters. This was done using the MATLAB function `fminsearch`. Note that `fminsearch` does not necessarily find the global minimum, but only a local minimum. Figure 6.2 illustrates examples where the parameter \hat{d}_{y_0} was perturbed, and patient-specific values for $\nu_{\hat{d}_{y_0}}$ were determined to minimize model-error. Both patient data and the model simulation using the time-dependent $\check{d}_{y_0}(t)$ are depicted in the figure. For most patient, this resulted in good visual agreement between model and data. Model fits to data for all patients from the DALIAH trial considered are shown in chapter B of Electronic Supplementary Material 2.

Since the degree of perturbation of the model-parameters depend on the IFN-dose, it was possible to analyse the asymptotic behaviour of the model for any hypothetical dose using the value, ν_{opt} , which minimized the sum of squared error. Hence, it could be determined which steady states were locally stable for a given dose, D_0 , by setting $B(t) = D_0$, calculating the perturbed parameters $\check{\theta} = (1 + \nu_{opt}D_0)\theta$ and analysing the local stability of all steady states for the resulting parameters. The stability information is shown in the bottom panel of figure 6.2. We show doses where the combination of dose and ν_{opt} resulted in a scenario where the full-blown leukemic steady state was the only locally stable steady state in red, while doses for which the only locally stable steady state is the healthy state are shown in green. There were no cases with multiple locally stable steady state, nor examples of a co-existence steady state being locally stable.

6.4 Results of fitting to available patient data

Since the Cancitis model was based on biological mechanisms, the relation between the parameter perturbation and the IFN-dose can be interpreted biologically on a patient-specific level. Figure 6.3 shows a bar graph, with a bar for each of the IFN-treated patients of the DALIAH trial that

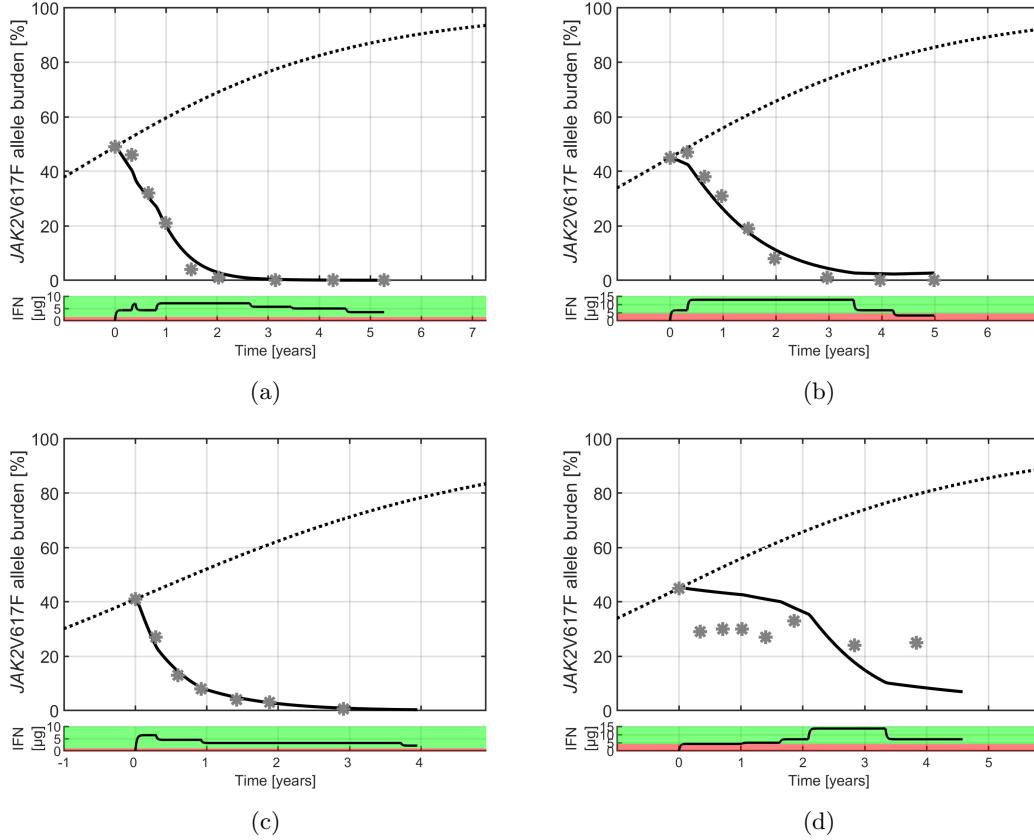


Figure 6.2: Examples of patient-specific fit of Cancitis model. The untreated disease development, shown as a dotted black line, was shifted such that time $t = 0$ coincides with the baseline measurement of the JAK2 allele burden. The patient data is shown as grey asterisks $*$. The parameter \hat{d}_{y_0} was perturbed in accordance with equation (6.5) to minimize the squared error. In the scenario shown, $\nu_{\hat{d}_{y_0}} = 0.52$ minimized the error. The model simulation with the resulting time-dependent $\check{d}_{y_0}(t)$ is shown as a full black line. The bottom part of all figures shows the modelled blood-concentration of IFN in a black curve. The coloured background displays the asymptotic behaviour of the model for the given dose, with green displaying blood-concentrations that would lead to eradication of the leukemic clone and red background leading to full-blown disease. Panel (a), (b) and (c) depicts patients “P002”, “P198” and “P184” respectively, and the model-fits shows good visual agreement with data. For panel (d), depicted patient “P164”, the model does not visually agree with data, yielding a bad fit.

had three or more measurements of the JAK2 allele burden. For the figure, IFN-dose dependent perturbation of the parameter \hat{d}_{y_0} was fitted to the JAK2 allele burden data as described above for all patients with three or more JAK2 allele burden measurement. Since \hat{d}_{y_0} relates to the death-rate of malignant stem cells, the fitted relation between a given dose and the relative increase of the death-rate was determined for each patient. The figure depicts the relative increase for a range of doses, between 0 and 15 μg per day. For patient “P002” depicted in figure 6.2a, the fitted value $\nu_{\hat{d}_{y_0}} = 0.52$ implies that for a dose of 5 μg , the parameter \hat{d}_{y_0} is increased by a factor $1 + 5 \cdot 0.52 = 3.60$. The asymptotic dynamics of the model suggest that a threshold over which treatment will be successful, in the sense that the healthy steady state is locally stable. Considering only \hat{d}_{y_0} , the bifurcation where local stability switches from the leukemic steady state to the healthy steady state occurs when \hat{d}_{y_0} is increased by a factor 1.8. This threshold is shown as a dotted line in figure 6.3. The model-fits suggests that some patients could be successfully treated with low levels of IFN while some patients required higher doses. Certain patients have such a low response to IFN that successful treatment requires an incredibly high dose, at which toxicity could become a concern. Hence, the model-fits suggest that other drugs must be considered for these patients. For the majority of patients however, successful treatment is attained for doses below 10 μg IFN per day.

By relating the IFN-dose to the biological effect we were thus able to give a quantifiable estimate of how much the death rate had to be increased by IFN for the given patients, assuming that other effects of treatment were insignificant.

Since the response is dose-dependent, it was possible to determine the threshold dose for each patient, that is, the dose which would result in a 1.8-fold increase in \hat{d}_{y_0} and hence leads to the change of local stability. Four patient were excluded, as the required doses greatly exceeded 30 μg IFN daily. Excluding these patients, the distribution of the doses required to attain successful treatment for the other patients is shown in figure 6.4. While there are differences between patients, the distribution suggests that the starting dose of 5 μg is sufficient for most patients, with approximately one third of patients requiring doses above 5 μg . The behaviour of the model also suggest that any dose above the given patient-specific threshold will result in faster treatment. Note however that no effect of toxicity is considered in the model, and high-dose IFN can be a concern in the clinical setting.

6.5 Predicting JAK2^{V617F} kinetics

Definitions of molecular response to treatment for ET and PV patients were described by Barosi et al. (2013), updating previous criteria given by Barosi et al. (2009). Complete molecular response was defined a complete eradication of the molecular abnormality, while the criteria for partial molecular response was defined as a decrease in mutant allele burden greater than 50%, for patients with an initial burden above 20%. In the work presented in (Ottesen et al., 2020), we considered the use of the Cancitis model as a tool for predicting molecular response. For simplicity, we focused on partial molecular response, and patients with initial JAK2 allele burden below 20% were also considered. This was done by fitting the model to a subset of the JAK2 allele burden data, and evaluating the model prediction at the time of the next measurement. Possible changes in dose between the fitted data-points and the predicted data-points were used in the prediction. From just the three first data-points; measured at baseline, 4 months and 8 months, the model was able to predict whether partial molecular response would be achieved at the next measurement well. Defining it a positive response when the patient did experience partial molecular response and a negative response when the patient did not, the predictions of the model was split into true positive (TP), false positive (FP), true negative (TN) and false

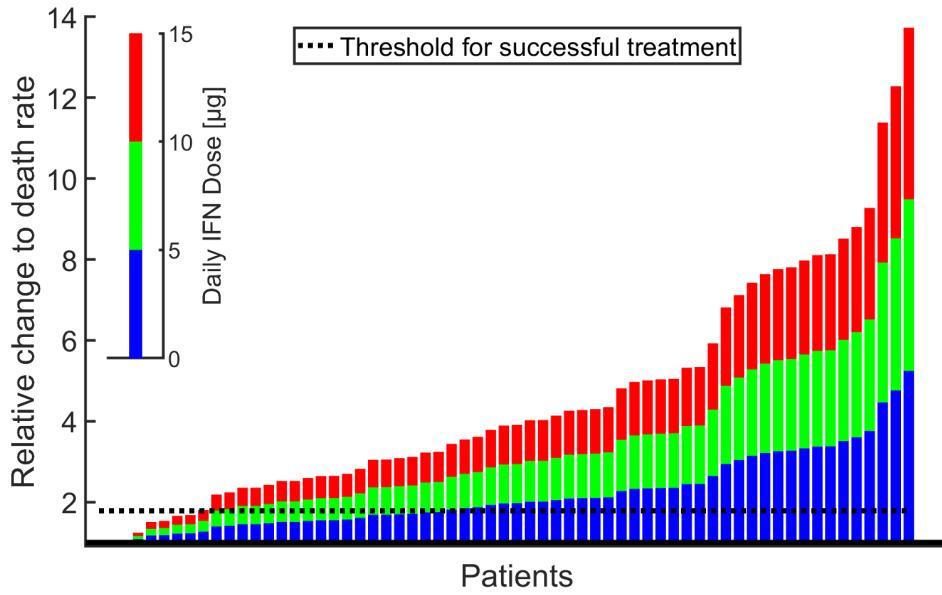


Figure 6.3: Biological interpretation of fitted IFN-responses when fitting perturbations of \hat{d}_{y_0} . Interpreting the perturbation of \hat{d}_{y_0} as the corresponding fold-increase of the death-rate of leukemic stem cells, each bars represent a given patient response. The blue part shows the model-prediction for 5 μg of IFN, the green for 10 μg , and the red for 15 μg . The threshold at 1.8-fold increase required for long-term eradication of the leukemic clone is shown as a black dotted line. The respond of three patients is not visible, as the fitted parameters $\nu_{\hat{d}_{y_0}}$ was too small. Six additional patients are below the threshold for 15 μg . Hence, the model-fits suggest that for nine of the patients, a dose of 15 μg is insufficient. (Reprinted from (Ottesen et al., 2020))

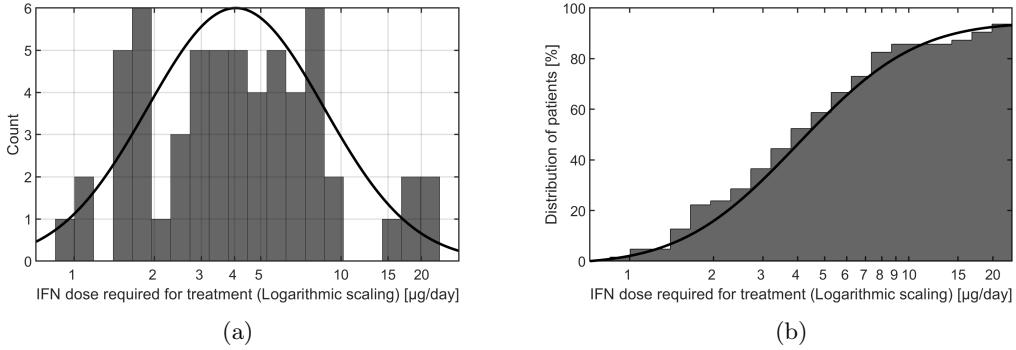


Figure 6.4: Distribution of minimum IFN-dose resulting in successful treatment. Panel (a) displays a histogram of the required doses determined by the model fits. A log-normal distribution fitted to the data is scaled and shown for illustration. In panel (b), a cumulative histogram of the threshold doses is shown, normalised by the number of patients. A cumulative curve of the log-normal distribution from panel (a) is also shown. Note that the x-axis of both panels scales logarithmically. For four patient, the necessary average daily dose exceed $30\mu g$, leading use to exclude them from panel (a). Hence, in panel (b) the maximal possible percentage of patients are 93.7% (59/63).

negative (FN) predictions. This suggested that the model prediction had sensitivity $\left(\frac{TP}{TP+FN}\right)$, specificity $\left(\frac{TN}{TN+FP}\right)$, and accuracy $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$ that were all above 75%. These measures increased when more data-points were considered. Hence, the model and the fitting procedure described above is capable of adequately predicting whether a patient will experience a partial molecular response to treatment or not. Figures showing this is shown in (Ottesen et al., 2020), but are omitted here for brevity.

An example of a more direct prediction of patient response is shown in figure 6.5. The figure illustrates the procedure described above, where a subset of JAK2 allele burden data was used to fit the perturbation of \hat{d}_{y_0} as above. The black line shows the model prediction given the data for changes to IFN dose as they were for the given patient. The dashed magenta and dashed blue curve however illustrate two hypothetical scenarios with either halted treatment or constant-dose treatment (i.e. continuing the dose at the last data-point used in the fit), respectively. These hypothetical model progressions allowed for a visual interpretation of the difference between continuing treatment and stopping treatment which can be useful in a clinical setting. Predictive plots for all patients from the DALIAH trial considered are shown in chapter D of Electronic Supplementary Material 2.

6.6 Fitting procedure for blood-cell counts

Fitting the model to the JAK2 allele burden provides an approximate description of one of the important clinical biomarkers for MPN. However, in the clinic, additional goals are considered in addition to reduction of the JAK2 mutational burden. In particular, normalization of the cell-counts of thrombocytes and leukocytes are important clinical endpoints, and failure to achieve normalized cell-counts were a criteria for increasing IFN dose in the DALIAH trial. Hence, agreement between the model and the data is also important in regards to thrombocyte- and

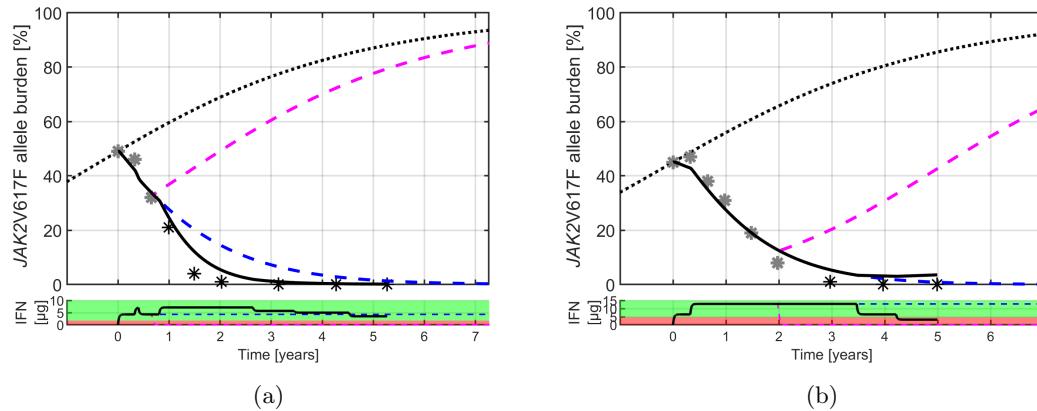


Figure 6.5: **Examples of patient-specific fits with prediction.** Similarly to figure 6.2, patient data is shown as *. The grey data-points are used in the current fit, while the black data-points show of the rest of the patient data. In panel (a) patient “P002” is shown, where only the first three data-points are used for the fitting procedure. Panel (b) depicts patient “P198”, using the first six data-point in the fitting procedure. The black line shows the model behaviour following the specific patient dosing-schedule. The dashed blue line shows a simulation where no further changes are made in dose after the last grey data-point, while the red dashed line shows a scenario where treatment is halted after the last grey data-point. In the bottom part of the figures, the dose information is shown as in figure 6.2, along with the blood-concentrations of the simulated scenarios in the corresponding colors.

leukocyte-counts. To relate the model-fits described above to the cell-counts, an additional subsequent fitting procedure was carried out. Since the model considers a conglomerate expression of mature blood-cells (x_1 and y_1), it is assumed that a certain fraction of the mature cells in the model are cells of a particular cell-line, e.g. thrombocytes. We assume a linear relation between the sum of mature cells in the model, $x_1 + y_1$, and the cell-counts in data. We define the cell-line specific errors:

$$E_{c,i} = (x_1(t_i, \nu_\theta) + y_1(t_i, \nu_\theta)) \mathcal{R}_c - D_{c,i} \quad (6.8)$$

where $D_{c,i}$ are the measurements of the cell line c at time t_i and \mathcal{R}_c is the fraction of mature cells in the model that relate to cell line c . By changing \mathcal{R}_c for each cell line (thrombocyte and leukocyte), it is possible to minimize the sum of squared errors, $\sum_{i=1}^n E_{c,i}^2$ where n is the number of data-points for the given patient. In figure 6.6 an example is illustrated. The remaining patient-specific fits are available in chapter C of Electronic Supplementary Material 2. By scaling the sum of mature cells with the cell-line-specific factor, both the disease development without treatment and the treatment-induced decrease in mature cell-counts can be shown. The treatment-free development illustrates the predicted development if treatment was not initiated for the given patient at time $t = 0$. For many of the patients of the DALIAH trial, the mature cell-counts of the model were found to agree well with the data when scaled linearly this way.

6.7 Considering multiple parameters

In the examples shown above, only the parameter \hat{d}_{y_0} was perturbed. As a final example of this section, figure 6.7 displays a least-square minimizing fit where distinct values for $\nu_{\hat{d}_{y_0}}$ and ν_{a_y} were fitted simultaneously, i.e. resulting in simultaneous changes to both \hat{d}_{y_0} and a_y . As expected, increasing the number of parameters fitted leads to an improved fit for most patients. This is observed in the particular example shown in figure 6.7. Fits to data of all patients considered are shown in Electronic Supplementary Material 3. Note that for some patients, considering a_y as well as \hat{d}_{y_0} visually only makes little difference compared to just considering \hat{d}_{y_0} .

For some patients, such as the examples shown in figure 6.7, fitting with perturbation of a_y in addition to \hat{d}_{y_0} lead a notable increase in JAK2 allele burden when the IFN dose is increased, particularly in the beginning, while for other patients the perturbation of a_y had an insignificant effect. Different a_y responses to IFN could explain why some patients have a short-term increase of the JAK2 allele burden before a decay, that is, a small hump in the JAK2 allele burden, similar to the pseudo-progression discussed by Lewin et al. (2020).

6.8 Summarizing discussion

In this section, we showed that the DALIAH trial data can be accurately modelled by the Cancitis model. The model was fitted to patient data by considering parameter perturbations that depend on the IFN-dosing. A simple interpretation of the effect of IFN lead to patient-specific fits that were able to recreate the dynamics of the JAK2 allele burden well. By considering a particular singular effect of IFN, namely an increase of the death-rate of malignant stem cells, we were able to describe how patients responded to treatment. For a significant proportion of patients, an IFN dose of $5\mu\text{g}$ was determined to lead to a doubling of the death-rate of malignant stem cells, with higher doses further increasing the death-rate.

From the fits of increasing death-rates of malignant stem cells to the JAK2 allele burden data, it was possible to determine patient-specific thresholds of how large a dose of IFN was necessary

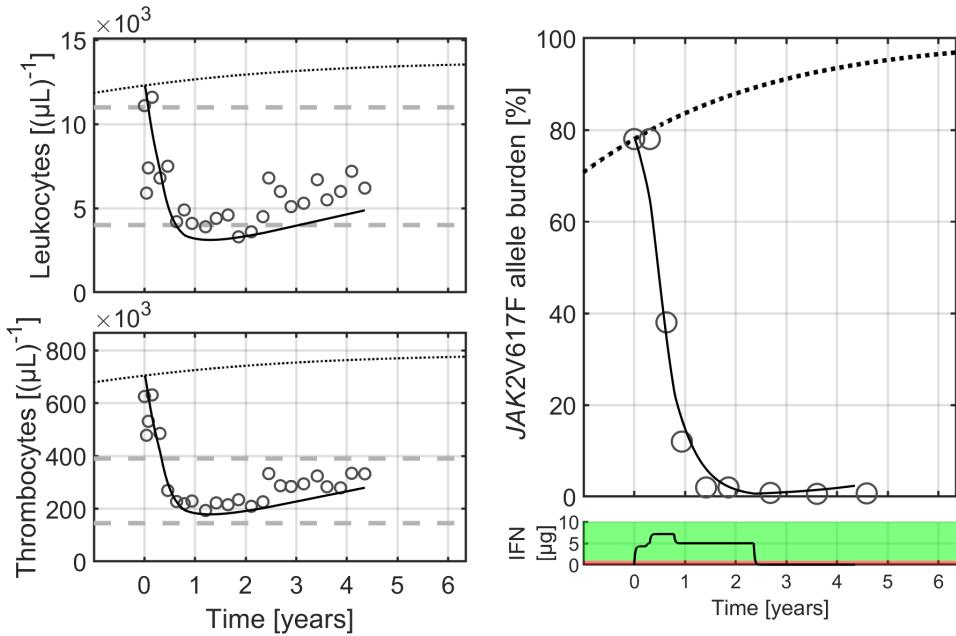


Figure 6.6: **Example of the subsequent fit to blood-cell counts.** Patient-data from patient “P086”. The right-hand panels show the minimizing fit when perturbing \hat{d}_{y_0} , as in figure 6.2. The top-left panel displays the leukocyte-measurement in cells per microliter, while the bottom-left panel displays the thrombocyte measurements, also in cells per microliter. The model behaviour without treatment is shown in dotted black, while the simulated model behaviour for the sum of mature cells is shown as a full black line. The model simulations has been scaled linearly to minimize the sum of squared errors, defined in equation (6.8), see the text. A healthy interval is shown in dashed grey, defined as between $4 \cdot 10^3$ and $11 \cdot 10^3 (\mu L)^{-1}$ for leukocytes and between $145 \cdot 10^3$ and $390 \cdot 10^3 (\mu L)^{-1}$ for thrombocytes.

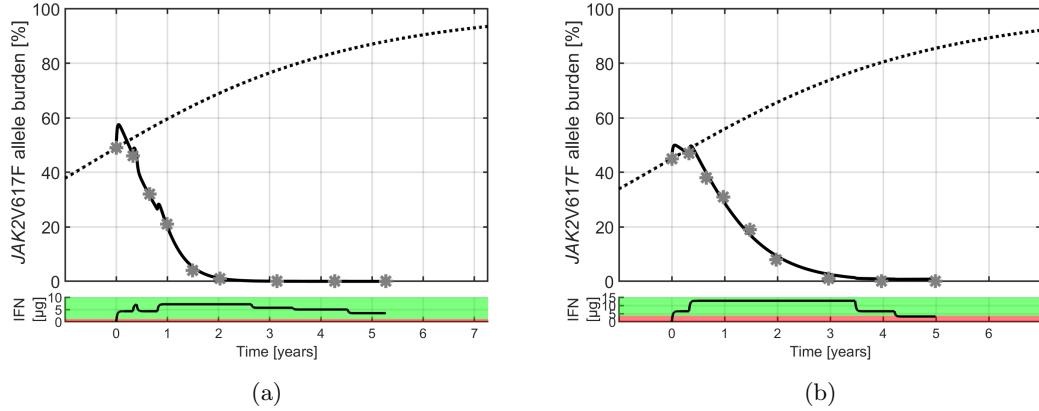


Figure 6.7: **Example of patient-specific fit of parameters \hat{d}_{y_0} and a_y .** Similarly to figure 6.2, model fits to data from patients “P002” and “P198” are shown in panel (a) and (b) respectively. In these figures, both \hat{d}_{y_0} and a_y were perturbed simultaneously. The perturbations depended on the modelled IFN blood-concentration, but not necessarily to an equal degree. All data-points were used in the fit shown. For more details, see the caption of figure 6.2.

to attain successful treatment for the given patients. The results suggest that for some patients, low doses could be sufficient, however larger doses are assumed to lead to faster treatment, even for patients where a low dose is sufficient. We determined that the typical IFN starting dose of approximately $5\mu\text{g}$ should be sufficient for two thirds of patients, while one third needs higher doses.

Based on the best fits to patient-specific JAK2 allele burden measurements, we were able to subsequently fit the model to mature cell counts, through a simple scaling of the mature cell count in the model. These fits demonstrated that the Cancitis model accurately describes the dynamics of cell counts, allowing for interpretation of the model variables as thrombocyte and leukocyte counts. Describing such measurable counts increase the possible clinical endpoints that the model can be used to evaluate, estimate and predict, and opens a path for testable model hypotheses about blood-cell counts.

It is possible that IFN-treatment affects multiple parts of the hematopoietic system, and hence only perturbing one parameter by the IFN-dose is a simplification. Considering multiple model-parameters is expected to improve the agreement with data. To investigate this, we perturbed both the death-rate and the differentiation rate of leukemic stem cell, \hat{d}_{y_0} and a_y respectively. This resulted in an improved fit to the JAK2 allele burden data. However, in addition to longer computational time of the fitting procedure, considering multiple parameters also introduces additional uncertainty and possibly overfitting. In general, one should consider the simplest model that agrees with data, and only introduce additional complexity when the simplest model is insufficient for describing the data. A simple model that agrees with data can provide simple hypothetical biological interpretation, while a complex but overfitted model can give false confidence in the model fit, and over-estimate the significance of the effect when interpreted biologically. Hence, although the improved fits shown in figure 6.7 agrees well with data, additional data must be considered in the future to verify that fitting both \hat{d}_{y_0} and a_y is necessary for accurate model-fits and biological interpretation of the effect of IFN. If more data had been obtained between baseline and the initial JAK2 allele burden measurement at month

4, it would have been possible to determine whether the fit shown in figure 6.7 more accurately describes the JAK2 kinetics than the fit considering only \hat{d}_{y_0} shown in figure 6.2.

Part IV

Modelling hematopoietic stem cells and the bone-marrow microenvironment

Chapter 7

Brief review of mathematical models of HSC

In this part of the thesis, we present a mathematical model of HSC. This entails a presentation of various reduced forms of the model as well as a thorough analysis on the model dynamics. Before discussing the specifics of the model, we reiterate some of the important features of HSC, previously described in section 2.2, and present some of the important mathematical models of HSC dynamics already presented and discussed in the literature. In particular, the models of Stiehl et al. (2020); Ashcroft et al. (2017) and Becker et al. (2019) are discussed. While other important mathematical models of HSC exist, the models presented in these three articles formed the basis of inspiration for the development of the model presented in the next chapter.

We summarize the most significant features of HSC, as discussed in section 2.2. HSC are primarily found within the bone-marrow microenvironment (BM), and typically in close proximity of certain cells referred to as niches. HSC undergo self-renewing division (where one HSC divides into two HSC) or differentiation (where a HSC turns into a more mature cell). Differentiation occurs either spontaneously or during division. The term “asymmetric division” is used to describe a division where one HSC gives rise to one HSC and one differentiated cell. Mutations can occur, resulting in HSC-clones that differ from the rest of the HSC population, and in some cases lead to malignancy. In case of malignancy, the mutated cells are referred to as malignant stem cells or leukemic stem cells. Mathematical modelling has been used as a useful tool to investigate the role that the HSC-specific niches play in maintaining normal HSC behaviour.

Stiehl et al. (2020) present a mechanistic model of HSC as part of a larger model of hematopoiesis. The work is an extension of previous investigations by the same authors (Wang et al., 2017), and is used for risk-stratification of AML-patients with great success. In the model, niche-bound HSC divides in such a way that one daughter-cell remains in the niche while the other cell is free and subsequently attempts to find a separate niche to occupy. When the free HSC encounter an empty niche, it attaches to it. Whe in proximity of an occupied niche, it attempts to dislodge the occupying cell. If the free cell is unsuccessful in attaching to a niche after a particular number of tries, it differentiates into a more mature cell. As a result, differentiation and in turn production of mature blood cells depends directly on the ability for the HSC population to occupy niche-space. By modelling leukemic stem cells that compete for the same niches, the model suggests that if a leukemic clone dislodge healthy HSC, it can out-compete the health clone. Relating the HSC dynamics to a model of progenitor-dynamics due to some of the same authors (Marciniak-Czochra et al., 2009) (previously mentioned in section 4), and comparing with experimental data, Wang et al. (2017) provides evidence that HSC and leukemic stem cells

do indeed compete over a shared resource of niches, rather than leukemic stem cells occupying a separate set of niches. The recent work described by Stiehl et al. (2020) further substantiates that the probability that a leukemic stem cell dislodge a healthy stem cell must be sufficiently high, since otherwise the progression observed in AML-patients cannot occur.

In the model proposed by Ashcroft et al. (2017), niche-bound HSC similarly gives rise to two cells where one maintains the niche space of the original HSC. The other cell either attaches to an empty niche, if any are available, or exits the BM and circulates in the peripheral blood. Spontaneous detachment of niche-bound HSC is considered, as well as attachment of HSC from the peripheral blood to empty niches. Cellular death is modelled, with different rates dependent on whether the HSC is within the BM or in the peripheral blood. Through mathematical analysis of the model as well as stochastic simulations, Ashcroft et al. (2017) investigates different scenarios with clinical relevance. The results provides insight into clonal dominance when multiple distinct HSC clones are considered, revealing a relation between the growth advantage of a mutant clone and the time it takes for the clone to reach a certain level of clonality. In addition, stem cell transplantation is considered in the model, showing that multiple small doses of labelled HSC into the peripheral blood can lead to higher cell-count of the labelled HSC, compared to when a single large dose of HSC is transplanted. Hence, the model is in agreement with experimental evidence suggesting more efficient engraftment into the bone-marrow through splitting of the transplanted dose (Bhattacharya et al., 2009).

Lastly, the model of Becker et al. (2019) also considers HSC and their interaction with the bone-marrow niches. In contrast to other work, Becker et al. (2019) also models the numbers of niches dynamically. In the proposed model, HSC are able to divide and self-renew when they are not attached to the niches. When bound to the niches, signalling from the niche induces quiescence in the HSC, keeping them from self-renewal. A loss of HSC due to differentiation and cell death is considered, for both free and niche-bound HSC. Attachment of HSC to empty niches and detachment is modelled to occur spontaneously. To model niche dynamics, niche-cells with a HSC bound to them divide and give rise to new empty niches. When no HSC are bound to the niches, they can undergo apoptosis, and hence the niche-cell population diminishes in the absence of HSC. The interaction between HSC and niche-cells results in a system that returns to homeostasis after perturbations, and offers a hypothesis for why HSC reconstitution following stem cell ablation can be slow. The return to homeostasis occurs in a damped oscillatory way, which is determined to occur due to integral feedback of HSC numbers on HSC production, a robust and well-known type of feedback control.

The three models described all consider some type of interaction between HSC and HSC-specific bone-marrow niches. Together they provide different hypothesis for the details of the interaction while all agreeing with general HSC knowledge and experimental evidence. Since observing HSC within the bone-marrow is difficult *in vivo*, it is hard to determine on which points the three models are in agreement with real HSC behaviour and on which point they do not. In the next section, we present a novel model of HSC and their interaction with the niches, in an effort to further investigate HSC dynamics. The proposed model is inspired by the work described above, and under certain conditions, the models are in agreement. The proposed model is thus not a substitution of previous models but rather an extension and additional hypothetical description of the interaction between HSC and the niches.

Chapter 8

The HSC niche model

In this chapter, we present a mathematical model of HSC within the bone-marrow microenvironment and the interaction between HSC and the HSC-specific niches. The model was originally proposed and investigated in Pedersen et al. (nd1) (Supplementary material B). Further mathematical exploration of the model as well as a model reduction was the focus of the work of Pedersen et al. (nd2) (Supplementary material C). The mathematical analysis described here is similar to the one presented in Pedersen et al. (nd2), presented in detail here for completeness. We first give a description of the formulation of the model and the biological reasoning behind it.

8.1 Constructing the HSC niche model

To construct a simple mathematical model of HSC dynamics, we restrict our focus to just the HSC and the interaction with HSC-specific niches, and signalling or immune response from the rest of the body is not considered until later in this thesis.

We consider multiple distinguishable HSC sub-populations. Differences in sub-populations could be due to mutations with no significant effect, malignant mutations leading to hematologic malignancies such as MPN or chemically marked HSC in a transplantation scenario. We refer to such different sub-populations as different *clones* of HSC. All variables and rates pertaining to a given clone carry the same subscript in the model. In the most general terms, the model considers $n \in \mathbb{N}$ different distinguishable clones. In the following, we use j to denote any $j \in (1, \dots, n)$.

We describe the processes that we consider:

- **The niche and the release of HSC.** HSC that are in proximity with niche-cells are considered bound. We denote these N_j . These HSC are considered quiescent or dormant, and while the specifics of this quiescence is not considered in detail here, the interaction with the niche is assumed to be an important part of maintaining healthy stem cell function (Kumar and Geiger, 2017; Vaidya and Kale, 2015; Zhang and Gao, 2016). We refer to niches that have no HSC bound to them as empty niches, and denote them N_E . The release or unbinding of niche-bound HSC is assumed to occur spontaneously with rate u_j , and results in a free cell A_j and an empty niche:

$$N_j \xrightarrow{u_j} N_E + A_j \quad (8.1)$$

- **Self-renewing division of free HSC.** Free HSC can divide in a self-renewing way such that cells post-division are still considered HSC, i.e. their stemness is maintained. We

hypothesize that self-renewing division of free HSC cannot be done indefinitely and that after an unspecified number of divisions, the cell requires interaction with the niche to self-renew again. We denote free HSC that are active and can self-renew as A_j and the resulting inactive or inhibited free HSC as I_j :

$$A_j \xrightarrow{r_j} 2\gamma I_j \quad (8.2)$$

Where $\gamma \geq 1$ is a finite real number and r_j is the rate at which A_j divide. The exact number of self-renewing division possible is not well-known, and hence γ is not determined. In the supplementary material of (Pedersen et al., nd2) (available at the end of appendix C, page 197), we consider a model in which each self-renewing division is considered separately. Through a quasi-steady-state approximation of all intermediate steps between cells that have not yet divided, i.e. A_j , and the cells that can no longer divide, i.e. I_j , we determine that $\gamma \in \mathbb{R}$ is appropriate in general, and that assuming $\gamma = 1$ is suitable.

- **Differentiation of active HSC.** When active stem cells do not self-renew, they are modelled to divide and differentiate into two differentiated cells, considered progenitor cells:

$$A_j \xrightarrow{d_{A_j}} 2D_j \quad (8.3)$$

We only consider symmetric division, see the discussion in section 8.2 below.

- **Differentiation of inactive HSC.** The inactive HSC can no longer divide in a self-renewing way. Without interaction with the niche, the inactive HSC are modelled to differentiate spontaneously:

$$I_j \xrightarrow{d_{I_j}} D_j \quad (8.4)$$

Note that we do not assume that the inactive HSC divide before differentiation, but rather that the differentiation is spontaneous.

- **Attachment to empty niches.** Lastly, we model attachment or binding between empty niches and free HSC. In the most general form, free HSC, both active A_j and inactive I_j , are assumed to bind with empty niches spontaneously, resulting in bound HSC. The resulting niche-bound HSC are assumed to be indistinguishable, regardless of previous activity.

$$A_j + N_E \xrightarrow{b_{A_j}} N_j \quad (8.5)$$

$$I_j + N_E \xrightarrow{b_{I_j}} N_j \quad (8.6)$$

The different processes considered are illustrated in figure 8.1

Note that different clones interact only through the niches. No direct interaction between free cells are considered in the model. In the processes described above, we did not explicitly consider apoptosis. We assume that, compared to the other processes, HSC apoptosis is rare. Apoptosis of both free active and inactive HSC could be considered, by including the rates of apoptosis in the rates of differentiation d_{A_j} and d_{I_j} . Hence d_{A_j} and d_{I_j} would describe the effective rates of loss of HSC due to both differentiation and apoptosis. Hence, including HSC apoptosis in the model would not impact the dynamics of HSC. While changes in HSC apoptosis could affect the production of progenitor cells, we postpone this discussion for later parts of this thesis.

Assuming that all considered variables are numerous enough for the law of large numbers to be applicable, the processes described above can be modelled as a system of ODEs describing the dynamics of n clones:

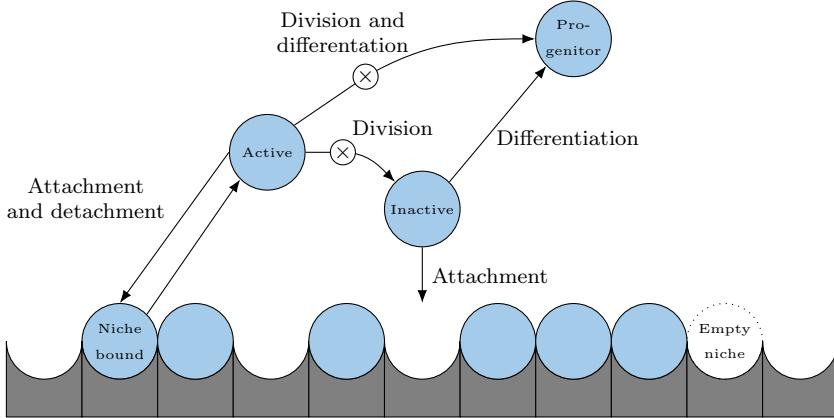


Figure 8.1: **Visual representation of the HSC mechanisms considered.** Blue circles represent cells, with text labelling their type (Niche-bound HSC, active HSC, inactive HSC or progenitor cells). HSC-specific niches are illustrated as gray “egg-trays”. Each arrow depicts one of the processes described above. Small circles with a \times symbol show processes where the cell divides.

$$\dot{N}_E = \sum_{i=1}^n u_i N_i - N_E \sum_{i=1}^n (b_{I_i} I_i + b_{A_i} A_i) \quad (8.7a)$$

$$\dot{N}_j = b_{I_j} N_E I_j + b_{A_j} N_E A_j - u_j N_j \quad (8.7b)$$

$$\dot{I}_j = 2\gamma r_j A_j - b_{I_j} N_E I_j - d_{I_j} I_j \quad (8.7c)$$

$$\dot{A}_j = u_j N_j - b_{A_j} N_E A_j - r_j A_j - d_{A_j} A_j \quad (8.7d)$$

where j refers to the j^{th} clone. For n clones, the full system consists of $3n + 1$ equations. Since all parameters describe the rates at which biological processes occur, they are all assumed to be non-negative. As the right-hand sides of the system fulfil a Lipschitz condition in the variables, a unique solution exists for which $N_E(t), N_j(t), I_j(t)$ and $A_j(t)$ are C^1 for all j .

Since the modelled variables are cell-counts, it is useful to make a distinction between solutions that have biological meaning and solutions that do not. In particular, solutions with non-negative cell-counts are defined as *feasible*.

Definition 8.1.1: Feasibility

A set of solutions N_E, N_j, I_j, A_j is *feasible* if all variables N_E, N_j, I_j, A_j are non-negative for all j .

All solutions of equations (8.7) are feasible for all $t > 0$, for non-negative initial conditions at $t = 0$. This can be seen by considering the derivatives of the system in the case where the

particular variable is zero:

$$\dot{N}_{E|N_E=0} = \sum_{i=1}^n u_i N_i \geq 0 \quad (8.8a)$$

$$\dot{N}_{j|N_j=0} = N_E (b_{A_j} A_j + b_{I_j} I_j) \geq 0 \quad (8.8b)$$

$$\dot{I}_{j|I_j=0} = 2\gamma r_j A_j \geq 0 \quad (8.8c)$$

$$\dot{A}_{j|A_j=0} = u_j N_j \geq 0 \quad (8.8d)$$

for all j .

The $3n + 1$ dimensional system of equations trivially reduces to a $3n$ dimensional system, since the sum of the occupied niches $\sum_{i=1}^n \dot{N}_i$ and empty niches, N_E is constant. Hence, we express the total number of niches as $N_E + \sum_{i=1}^n N_i = K$. For feasible solutions, it must hold that $K \geq 0$. As K corresponds to the total number of HSC niches in the bone marrow, we generally assume $K > 0$ and under realistic biological conditions, we assume $K \gg 1$. Having defined K , we have $N_E = K - \sum_{i=1}^n N_i$ and we define the model.

Definition 8.1.2: The HSC niche model

The $3n$ -dimensional *HSC niche model* is given by the system of equations:

$$\dot{N}_j = b_{I_j} \left(K - \sum_{i=1}^n N_i \right) I_j + b_{A_j} \left(K - \sum_{i=1}^n N_i \right) A_j - u_j N_j \quad (8.9a)$$

$$\dot{I}_j = 2\gamma r_j A_j - b_{I_j} \left(K - \sum_{i=1}^n N_i \right) I_j - d_{I_j} I_j \quad (8.9b)$$

$$\dot{A}_j = u_j N_j - b_{A_j} \left(K - \sum_{i=1}^n N_i \right) A_j - r_j A_j - d_{A_j} A_j \quad (8.9c)$$

$$N_E = K - \sum_{i=1}^n N_i \quad (8.9d)$$

for all $j \in [1, \dots, n]$.

Feasibility of solutions and initial conditions still depends on $N_E \geq 0$ and hence the range of $N_j(t)$ is restricted to $[0, K]$. In fact, the sum of niche-bound HSC is restricted, $\sum_{i=1}^n N_i(t) \leq K$ since $N_E(t) \geq 0$ holds for all t for feasible initial conditions. The range of $I_j(t)$ and $A_j(t)$ is $[0, \infty)$ for all j .

While the production of differentiated cells does not influence the dynamics of the HSC niche model, it has relevance for the production of blood-cells. Assuming that progenitors arising from active HSC and inactive HSC are identical, we can define the production of progenitor-cells based on the processes described in equations (8.3) and (8.4).

Definition 8.1.3: Production of progenitors

Production of progenitor cells of the HSC niche model in definition 8.1.2 is defined for the j^{th} clone as:

$$i_{D_j} = 2d_{A_j} A_j + d_{I_j} I_j \quad (8.10)$$

Note that definition 8.1.3 only describe the influx to a hypothetical progenitor compartment. Further details about progenitor dynamics is considered later in this thesis, but is not relevant for the present purposes.

For illustration, a compartment diagram of the HSC niche model with $n = 2$ is shown in figure 8.2.

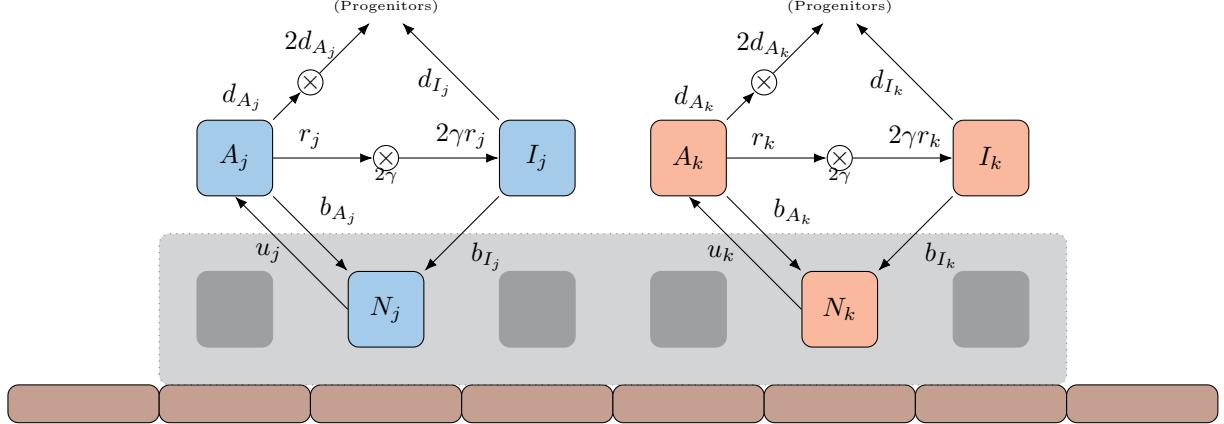


Figure 8.2: **Compartment diagram of the HSC niche model.** Two clones are considered, clone j shown in blue and clone k shown in red. The compartments of the six variables are shown as rounded boxes and the arrows illustrate the interaction between variables. The empty dark-gray boxes in the lower part of the figure illustrates empty niche-space. The dark-red cells depicted at the bottom are the unspecified niche-cells, shown for illustration.

8.2 Asymmetric division

Before analysing the model in detail, a brief comment on asymmetric cell division is warranted. As described in section 2.2, HSC can divide both symmetrically and asymmetrically, that is, it is possible that a cell-division results in two distinct cells; one similar to the original cell and one which is differentiated. In the processes described above however, only symmetric division of active cells was considered, giving rise to either two HSC (with rate r_j) or two progenitor-cells (with rate d_{A_j}). For the model presented, the consideration of asymmetric explicitly is not necessary, as the contribution of asymmetric division can be included in the rates r_j and d_{A_j} , which are hence considered *effective* rates for self-renewing division and differentiating division respectively.

To show this, consider asymmetric division by replacing the processes of equations (8.2) and (8.3) by:

$$A_j \xrightarrow{S_r} 2\gamma I_j \quad (8.11)$$

$$A_j \xrightarrow{a} \gamma I_j + D_j \quad (8.12)$$

$$A_j \xrightarrow{S_d} 2D_j \quad (8.13)$$

where S_r is the rate of symmetric self-renewing division, S_d the rate of symmetric differentiating division, and a is the rate of asymmetric division. Note that we assume that the cell resulting

from self-renewing division gives rise to γ inactive cells I_j , regardless of the symmetry of the division. The model formulation analogue to equations (8.9) is then:

$$\dot{N}_j = b_{I_j} N_E I_j + b_{A_j} N_E A_j - u_j N_j \quad (8.14a)$$

$$\dot{I}_j = (2S_r + a)\gamma A_j - b_{I_j} N_E I_j - d_{I_j} I_j \quad (8.14b)$$

$$\dot{A}_j = u_j N_j - b_{A_j} N_E A_j - (S_r + a + S_d) A_j \quad (8.14c)$$

$$N_E = K - \sum_{i=1}^n N_i \quad (8.14d)$$

$$i_{D_j} = (2S_d + a)A_j + d_{I_j} I_j \quad (8.14e)$$

By defining $r_j = S_r + \frac{a}{2}$, equation (8.14b) is identical to (8.7c). Substituting $a = 2r_j - 2S_r$ in the expression for A_j and for i_{D_j} , we obtain:

$$\dot{A}_j = u_j N_j - b_{A_j} N_E A_j - (S_d + 2r_j - S_r) A_j \quad (8.15)$$

$$i_{D_j} = (2S_d + 2r_j - 2S_r) A_j + d_{I_j} I_j \quad (8.16)$$

Defining $d_{A_j} = S_d + r_j - S_r = S_d + \frac{a}{2}$, we have:

$$\dot{A}_j = u_j N_j - b_{A_j} N_E A_j - (r_j + d_{A_j}) A_j \quad (8.17)$$

$$i_{D_j} = 2d_{A_j} A_j + d_{I_j} I_j \quad (8.18)$$

and exactly the system in equations (8.9) is obtained. Hence, it is not necessary to explicitly consider asymmetric division, as including $\frac{a}{2}$ in the effective rates for symmetric division is equivalent. Note however that while this holds for the system described, it does depend on the assumption that there is no distinction between the cells arising from asymmetric division and from symmetric division. Similarly, the assumption that active HSC always divide before differentiation is also a necessity for omission of explicit asymmetric division. As it has been suggested that differentiation can occur due to microenvironmental asymmetry rather than due to divisional asymmetry (Wilson and Trumpp, 2006), the omission of asymmetric division is a simplification that might not be appropriate.

8.3 Dynamics and behaviour of the HSC niche model

Analysis of the HSC niche model reveals some features of the model behaviour. In this section, we go through some of the immediate results of the model analysis.

8.3.1 Existence of an attracting trapping region

An attractive trapping region exists for the HSC niche model.

Theorem 8.3.1: Trapping region for the HSC niche model

For

$$\hat{A}_{j,\epsilon} = \frac{u_j K}{r_j + d_{A_j}} + \epsilon \quad (8.19)$$

and

$$\hat{I}_{j,\epsilon} = \frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}} \epsilon + \epsilon \quad (8.20)$$

where $\epsilon \geq 0$,

$$TR_\epsilon = \left\{ \begin{array}{l} \prod_{j=0}^n (N_j, A_j, I_j) \in \mathbb{R}^{3n} : N_j \geq 0, A_j \geq 0, I_j \geq 0, \\ \sum_{j=1}^n N_j \leq K, A_j \leq \hat{A}_{j,\epsilon}, I_j \leq \hat{I}_{j,\epsilon}, j = 1, \dots, n \end{array} \right\} \quad (8.21)$$

is a trapping region for the HSC niche model. In addition, for $\epsilon > 0$, it is an attracting trapping region.

The existence of the trapping region can be seen by evaluating \dot{A}_j in $\hat{A}_{j,\epsilon}$, yielding only negative contributions when $\epsilon > 0$. The situation is similar for \dot{I}_j evaluated in $\hat{I}_{j,\epsilon}$ and for any \dot{N}_j when $\sum_{i=1}^n N_i = K$. Further details of are described in (Pedersen et al., nd2), supplementary material C, along with proof that trajectories starting outside the trapping region enters the region in finite time.

8.3.2 Steady states of the HSC niche model

Steady states are denoted with an asterisk *. The steady states of \dot{A}_j and \dot{I}_j are trivially found from equations (8.9c) and (8.9b) respectively.

Lemma 8.3.1: Steady states of \dot{A}_j and \dot{I}_j depend on N_j^* and N_E^*

For any steady state with N_j^* and $N_E^* = K - \sum_{i=1}^n N_i^*$,

$$A_j^* = \frac{u_j N_j^*}{r_j + d_{A_j} + b_{A_j} N_E^*}. \quad (8.22)$$

and

$$I_j^* = \frac{2\gamma r_j u_j N_j^*}{(r_j + d_{A_j} + b_{A_j} N_E^*) (d_{I_j} + b_{I_j} N_E^*)} \quad (8.23)$$

If all N_j^* are within the trapping region TR_0 , then so is A_j^* and I_j^* .

Note that all points in the trapping region are non-negative, and hence determining a steady state with N_j^* within the trapping region for all j , implies that all variables are non-negative and hence that the given steady state solution is feasible.

By considering $\dot{N}_j^* = 0$, the following expression arises:

$$N_j^* = \frac{1}{u_j} N_E^* (b_{I_j} I_j^* + b_{A_j} A_j^*) \quad (8.24)$$

Due to lemma 8.3.1, $N_j^* = 0$ implies both $A_j^* = 0$ and $I_j^* = 0$. Hence, equation (8.24) has a trivial solution, for which $N_j^* = A_j^* = I_j^* = 0$. A trivial steady state exists, in which this holds for all j .

Theorem 8.3.2: Trivial steady state S_0^*

A *trivial steady state*, denoted S_0^* , always exists. In the trivial steady state $N_j^* = A_j^* = I_j^* = 0, \forall j \in \{1, \dots, n\}$.

The number of empty niches in the trivial steady state is $N_E^* = K - \sum_{i=1}^n N_i^* = K$.

Using the expression of lemma 8.3.1, equation (8.24) can be written as:

$$N_j^* = N_E^* \left(b_{I_j} \frac{2\gamma r_j}{(r_j + d_{A_j} + b_{A_j} N_E^*) (d_{I_j} + b_{I_j} N_E^*)} + b_{A_j} \frac{1}{r_j + d_{A_j} + b_{A_j} N_E^*} \right) N_j^* \quad (8.25)$$

Assuming $N_E^* \geq 0$ and $N_j^* > 0$, the expression simplifies to

$$N_E^* = K - \sum_{i=1}^n N_i^* = \frac{d_{I_j} (r_j + d_{A_j})}{b_{I_j} ((2\gamma - 1)r_j - d_{A_j})}, \quad (8.26)$$

under the assumption $(2\gamma - 1)r_j \neq d_{A_j}$. Hence, N_E^* has a unique value for any steady state with $N_j^* > 0$. This gives rise to a definition of HSC fitness.

Definition 8.3.1: Fitness

The *fitness* of the j^{th} clone is defined as:

$$F_j = \frac{b_{I_j} ((2\gamma - 1)r_j - d_{A_j})}{d_{I_j} (r_j + d_{A_j})}. \quad (8.27)$$

In a steady state with $N_j > 0$, $N_E^* = F_j^{-1}$, and thus, for feasibility of the steady state $0 \leq F_j^{-1} \leq K$. $F_j > 0$ implies that $(2\gamma - 1)r_j > d_{A_j}$. For $F_j^{-1} = K$, note that $K - F_j^{-1} = K - N_E^* = \sum_{i=1}^n N_i^* = 0$ and hence $N_j^* = 0$. Thus, two conditions for a non-trivial feasible steady state are $(2\gamma - 1)r_j > d_{A_j}$ and $F_j^{-1} < K$.

A unique value of $N_j^* > 0$ implies both a unique value of N_E^* as well as for A_j^* and I_j^* , due to equations (8.22) and (8.23). In addition, A_j^* and I_j^* are zero when $N_j^* = 0$, and hence only in the trivial steady state.

Lemma 8.3.2: Steady state values for free cells

For any non-trivial steady state, with a unique $N_j^* > 0$ and $K - \sum_{i=1}^n N_i^* = F_j^{-1}$, A_j^* and I_j^* are unique and given by equations (8.22) and (8.23) respectively.

Hence, determining a non-trivial feasible steady state depends only on the n values of N_j^* .

A *single-clone* steady state exists, in which all clones $k \neq j$ have $N_k^* = A_k^* = I_k^* = 0$, while only the j^{th} clone has $N_j^* > 0$.

Theorem 8.3.3: Single-clone steady state S_j^*

Given $(2\gamma - 1)r_j > d_{A_j}$ and $F_j^{-1} < K$, a *single-clone steady state* exists for which $N_j^* > 0, A_j^* > 0, I_j^* > 0$ and $N_E^* > 0$ while $\forall k \neq j : N_k^* = A_k^* = I_k^* = 0$. In particular, $N_j^* = K - F_j^{-1}$ and $N_E^* = F_j^{-1}$.

We denote the unique j^{th} single-clone steady state as S_j^* .

In addition to the trivial steady state and the n single-clone steady states, another type of steady state can exist, in which two or more clones have non-negative counts simultaneously. This occurs only when the fitnesses F_j are equal for the given clones. As an example, two-clone co-existence steady states with feasible N_1^* and N_2^* exist when $F_1 = F_2$. This implies

$$\frac{b_{I_1}((2\gamma - 1)r_1 - d_{A_1})}{d_{I_1}(r_1 + d_{A_1})} = \frac{b_{I_2}((2\gamma - 1)r_2 - d_{A_2})}{d_{I_2}(r_2 + d_{A_2})} \quad (8.28)$$

Note that for the existence of non-trivial steady states, the conditions $(2\gamma - 1)r_1 > d_{A_1}$, $(2\gamma - 1)r_2 > d_{A_2}$, $F_1^{-1} < K$ and $F_2^{-1} < K$ must be fulfilled. Defining $F = F_1 = F_2$, equation (8.26) implies $N_E^* = F^{-1}$. Assuming all other clones are zero, the condition $F^{-1} + N_1^* + N_2^* = K$ applies. Defining β such that $N_1^* = \beta$, we have $N_2^* = K - F^{-1} - \beta$. A line is parametrized by β , going through the single-clone steady states S_1^* (for $\beta = K - F^{-1}$) and S_2^* (for $\beta = 0$). The line is unique since F^{-1} is uniquely determined. All points along this line are feasible steady states with $N_1^* \geq 0$ and $N_2^* \geq 0$.

The notion of co-existence steady states extends to multiple clones having equal fitness. This is generalized in the following theorem.

Theorem 8.3.4: Co-existence steady states

Co-existence steady states in which multiple clones assume positive concentrations may exist.

A necessary and sufficient condition for co-existence is that F_j is equal for all co-existing clones.

The number of empty niches, N_E^* , in the co-existence steady states is given uniquely by $N_E^* = F_j^{-1}$. Additionally, the bound cells must fulfil the condition $\sum_{i \in \mathcal{C}} N_i = K - F_j^{-1}$ where \mathcal{C} is the set of all co-existing clones.

Co-existence steady states exist on a simplex where the dimension of the simplex is the number of co-existing clones, $\dim(\mathcal{C})$.

When there is no ambiguity, we denote the simplex of steady states S_C^* .

The existence of multiple simplexes of co-existence steady states simultaneously is possible. As a particular example, if $F_1 = F_2 \neq F_3 = F_4$, two lines of co-existence steady states exists. Along one line, there are feasible steady states with $N_1^* \geq 0$ and $N_2^* \geq 0$ but $N_3^* = N_4^* = 0$, while the other line has $N_1^* = N_2^* = 0$ and $N_3^* \geq 0$ and $N_4^* \geq 0$. Similarly, if k clones have equal fitness, a k -dimensional simplex of co-existence steady states exists. As a consequence, many different combinations of equal fitness could be considered when multiple clones are considered. However, we assume that multiple clones having exactly equal fitness is improbable in a noisy biological setting, and the different simplexes of co-existence steady states are not discussed in extensive detail in this thesis. For illustration, we do however discuss the simplest case of two

clones having equal fitness while all other clones have different fitness, as an exemplary situation of co-existence.

To summarize the steady states of the HSC niche model, a trivial steady state S_0^* always exists (Theorem 8.3.2), single-clone steady states S_j^* exists under certain conditions on parameters (Theorem 8.3.3). For equal fitness between clones, one or multiple simplexes of co-existence steady states exist (Theorem 8.3.4). Figure 8.3 illustrates all possible steady states for one, two and three distinct clones up to permutations.

When only two clones are considered, ordering writing a subset of variables as $(N_E^*, N_1^*, N_2^*)^T$ we write up the steady states:

$$S_0^* = \begin{pmatrix} K \\ 0 \\ 0 \end{pmatrix}, \quad S_1^* = \begin{pmatrix} F_1^{-1} \\ K - F_1^{-1} \\ 0 \end{pmatrix}, \quad S_2^* = \begin{pmatrix} F_2^{-1} \\ 0 \\ K - F_2^{-1} \end{pmatrix} \quad (8.29)$$

For equal fitness, $F = F_1 = F_2$, a line of co-existence steady states exists:

$$S_C^* = \begin{pmatrix} F^{-1} \\ \beta \\ K - F^{-1} - \beta \end{pmatrix} \quad (8.30)$$

where β is a number between 0 and $K - F^{-1}$.

Lemma 8.3.1 and 8.3.2 shows that for all j and all steady state, unique A_j^* and I_j^* can be determined from equations (8.22) and (8.23) respectively.

8.3.3 Local stability of steady states

The local stability of the steady states described above can be determined in the general n -clone case. Details of the calculations are omitted for brevity. Additional details can be found in the main text of (Pedersen et al., nd2) and the related supplementary material of the article.

We order the $3n$ variables as $(N_1, I_1, A_1, \dots, N_n, I_n, A_n)$. The Jacobian of the system can then be written as a block-matrix of the form:

$$Jac = \begin{pmatrix} D_1 & G_1 & G_1 & \dots & G_1 \\ G_2 & D_2 & G_2 & \dots & G_2 \\ G_3 & G_3 & D_3 & \dots & G_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_n & G_n & G_n & \dots & D_n \end{pmatrix} \quad (8.31)$$

where

$$D_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j - u_j & b_{I_j} (K - \sum_{i=1}^n N_i) & b_{A_j} (K - \sum_{i=1}^n N_i) \\ b_{I_j} I_j & -b_{I_j} (K - \sum_{i=1}^n N_i) - d_{I_j} & 2\gamma r_j \\ u_j + b_{A_j} A_j & 0 & -b_{A_j} (K - \sum_{i=1}^n N_i) - r_j - d_{A_j} \end{pmatrix} \quad (8.32)$$

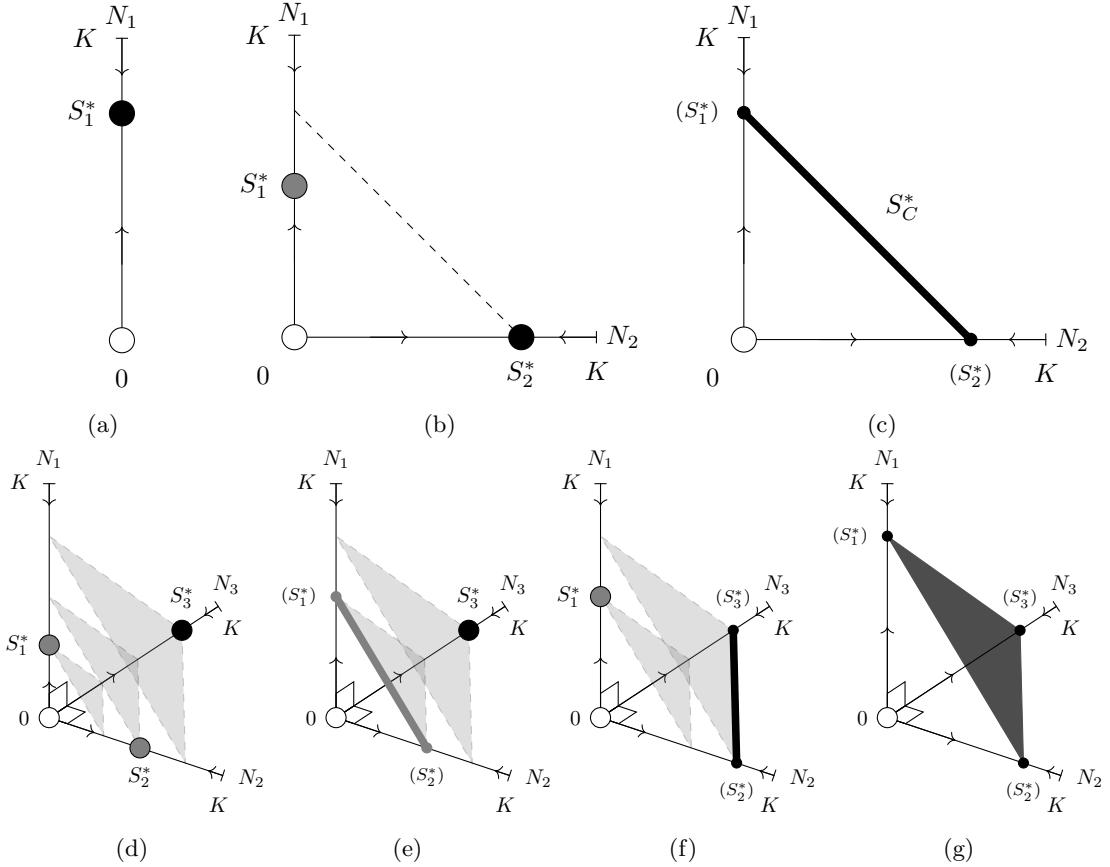


Figure 8.3: Illustrations of all steady states for one, two and three feasible clones. Assuming that all clones permit feasible single-clone steady states, the possible combinations are shown here. As such, this figure is a complete visual representation of one-, two-, and three-clone scenarios up to permutations. Only N_j -axes are shown, as A_j^* and I_j^* follow directly, due to lemma 8.3.1 and 8.3.2. White circles depict the trivial steady state S_0^* . Grey and black circles show the single-clone steady states S_j^* , with black circles being locally stable and grey circles being locally stable only within a single-clone subspace. Panel (a) depicts a single-clone scenario. When S_1^* is feasible it is locally stable. Scenarios with two clones are illustrated in panels (b) and (c). In panel (b) $F_1 < F_2$ and S_2^* is stable. The dashed line is illustrative to show that $N_2^* > N_1^*$. In panel (c), $F_1 = F_2$ and the line of co-existence steady states is shown in black. Numerical investigations show that this line is attracting, however, numerical investigations also reveal that the Jacobian evaluated in the steady states has a zero eigenvalue and hence the steady states are non-hyperbolic. Panels (d) through (g) depict the full range of possible scenarios with three clones. In (d) all three clones have different fitness. In (e) $F_1 = F_2 < F_3$, while panel (f) has $F_1 < F_2 = F_3$. Finally, panel (g) depicts the scenario where $F_1 = F_2 = F_3$. In panel (d-f) greyed-out triangles are shown as guide for the eye, while the black triangle in panel (f) is a triangle of co-existence steady states S_C^* .

and

$$G_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j & 0 & 0 \\ b_{I_j} I_j & 0 & 0 \\ b_{A_j} A_j & 0 & 0 \end{pmatrix} \quad (8.33)$$

We sketch how we determined local stability: Evaluating the Jacobian Jac at the trivial steady state S_0^* and at the single-clone steady states S_j^* reveals that the eigenvalues of the Jacobian are the eigenvalues of all n submatrices D_j evaluated at the corresponding steady states.

Lemma 8.3.3: Local stability of S_0^* and S_j^* is determined by D_j

In the trivial steady state S_0^* and in the single-clone steady states S_j^* , the eigenvalues of the Jacobian of the system of equations (8.9) evaluated in the given steady state values, are equal to the eigenvalues of the submatrices D_j , defined in equation (8.32).

Hence, the local stability of the steady states can be determined by considering the submatrices D_j .

Without determining explicit expressions for all eigenvalues of the system, it can be shown that the sign of the real parts of the eigenvalues is determined by the fitness of the clones. As a consequence, the Routh-Hurwitz stability criterion implies that local stability of the steady states can be determined and also depends on the fitness.

Lemma 8.3.4: Routh-Hurwitz stability criterion for D_j

Assume that S_j^* is feasible and hence $(2\gamma - 1)r_j > d_{A_j}$ as well as $F_j^{-1} < K$ holds.

The matrix D_j of equation (8.32) evaluated in the trivial steady state, S_0^* , i.e. $D_j|_{S_0^*}$, has at least one eigenvalue with positive real part.

Evaluated in S_j^* , the j^{th} single-clone steady state, all eigenvalues of the matrix $D_j|_{S_j^*}$ have negative real part.

The eigenvalues of $D_j|_{S_k^*}$ where $k \neq j$ (i.e. the matrix with parameters of the j^{th} clone, evaluated in the single-clone steady state of the k^{th} clone) depend on the fitnesses of both the j^{th} clone and the k^{th} clone. In particular, if $F_k > F_j$, all eigenvalues have negative real part, and when $F_k < F_j$ at least one eigenvalue has positive real part.

Hence the local stability of the trivial steady state and the single-clone steady states can be determined, assuming no clones have equal fitness.

Theorem 8.3.5: Local stability of S_0^* and S_j^* for all j

If the system has a feasible single-clone steady state S_j^* , the trivial steady state S_0^* is locally unstable. Otherwise, it is locally stable.

Assuming no clones have equal fitness, a particular single-clone steady state S_k^* is locally stable if and only if $F_k > F_j$ holds for all $j \neq k$. If it does not hold, the steady state S_k^* is locally unstable.

In the special case where only one clone is considered, S_1^* is locally stable.

Considering the possibility of equal fitness, it was shown in the supplementary material of (Pedersen et al., nd2) that in the two-clone scenario, an eigenvector along the line of co-existence has a corresponding eigenvalue zero. Numerical investigations revealed a similar structure when

equal fitness of more clones were considered, namely eigenvectors with eigenvalues zero along the simplex of co-existence steady states. Through numerical solutions of the differential equations under conditions for equal fitness of clones, the dependence on the fitness was found to depend on the relative fitness of clones in the expected way: When a single clone has the highest fitness, trajectories of solutions will move toward the corresponding single-clone steady state, and when multiple clones have equal and highest fitness, solution trajectories approach steady state on the simplex of co-existence steady states.

8.3.4 Estimates for parameters and numerical simulations

The main work of Pedersen et al. (nd1) consisted of numerical investigations of the HSC niche model with $\gamma = 1$ and $b_{A_j} = 0$. To obtain estimates of model parameters, the model was related to experimental data found in the literature. In particular, we considered data from mice-experiments presented by Bhattacharya et al. (2009). In these experiments, mice were transplanted with chemically marked HSC, followed by subsequent harvesting of the bone-marrow to determine the proportion of the bone-marrow HSC carrying the chemical marker at the time of harvest.

In the model, we considered all HSC to pertain to a single endogenous clone. We interpreted the transplanted HSC as a separate clone, with properties identical to the endogenous clone. From a single-clone equilibrium of the endogenous clone, we simulated a transplant of free HSC (A and I) from the transplanted HSC clone, similar to the experiment described by Bhattacharya et al. (2009). The relative frequency of the transplanted clone was then compared to the experimental data of Bhattacharya et al. (2009). Fitting to the experimental data with simulated annealing (using the MATLAB implementation due to Vandekerckhove (2008)) provided us with estimates for the parameters of the model. The parameter-estimates found for the long-term dynamics observed in data from (Bhattacharya et al., 2009) are shown in table 8.1. These values were used for the numerical investigations described in (Pedersen et al., nd1), see supplementary material B.

Table 8.1: **Default parameters for the niche model.** Parameters were found by relating the niche model to the data of Bhattacharya et al. (2009) using simulated annealing, see the supplementary material of (Pedersen et al., nd1), available at the end of appendix B for details. We assumed that all clones had identical parameters, and hence the shown parameters are default values for all clones j .

K	15000 cells	u_j	0.04 day^{-1}
r_j	2.32 day^{-1}	b_{I_j}	0.96 day^{-1}
d_{A_j}	2.06 day^{-1}	d_{I_j}	3.77 day^{-1}

Our numerical investigations considered the effect of each parameter on equilibrium HSC-counts and various transplant scenario. In particular, we investigated the effect of stem cell mobilization and preconditioning in relation to stem cell transplantation through simulation of temporary increases of u and by removing bound HSC before the transplant. The findings are presented in full in the article available as supplementary material B at the end of this thesis, but some important findings are described here.

Our results implied that increased mobilization of HSC must be done with great care if the mobilization affects both healthy and malignant cells. The release of bound malignant cells can

enhance their competitive advantage and speed up the eradication of healthy HSC. This was particularly clear following preconditioning in which great numbers of cells (both healthy and malignant) were removed initially.

However, we found a range of timing and mobilization efficiency for which the mobilization can be beneficial for patients and can delay disease progression. By increasing mobilization for a short time following a transplant of healthy cells, our simulations implied that the increased mobilization lead to an increased homing of the free cells and a smaller disease burden as a result. If intense preconditioning was simulated to precede the transplantation, the homing of free cells was not increased. Hence, for patient where only reduced intensity preconditioning is an option (such as for weak or elderly patients), we find that mobilization is more beneficial than for patient that can tolerate intense preconditioning.

We investigated the effect parameter-changes had on blood-production. We assumed that the production of progenitors as described in definition 8.1.3 correlated with the production of blood-cells. Simulating a transient increase in differentiation-rates d_{I_j} and d_{A_j} , the equilibrium value of i_{D_j} , I_j and A_j all decrease. However, progenitor production momentarily increase before converging to the decreased equilibrium value. Hence the model suggests that medical intervention that increase differentiation lead to a temporary increase of blood-cells. In addition, the model suggests that differences in HSC behaviours are not immediately clear from only blood-cell measurements. The rate of release from the niches u_j does not influence the steady state value of bound HSC, N_j^* , but does influence both free HSC and progenitor production. Hence, large differences in the release from the niche can result in over-representation of a clone in the progenitor production i_{D_j} and hence in the blood. This finding implies that blood-cell measurement cannot be sufficient for determining the state of HSC within the bone-marrow. However, it also implies that a malignant clone can be more numerous within the bone-marrow than otherwise implied by blood-cell measurements, which highlights the importance of HSC-targeted therapy, see figure 8.4.

8.4 Summarizing discussion

In this section, a mechanism-based mathematical model of the behaviour of HSC was suggested, analysed and parametrized. We here summarize the model formulation and the most significant findings.

8.4.1 Model formulation

Inspired by previous work on modelling HSC-behaviour (section 7), we made assumptions about which properties of HSC play the most significant role in determining the general behaviour of HSC in the human body. The role of HSC-specific niches within the bone-marrow is difficult to determine. We suggested a hypothetical interaction between HSC and the niche where HSC self-renewing division is limited and niche-binding is required to allow for continued division. This leads to a distinction between active HSC that are ready to divide in a self-renewing fashion and inactive HSC that can only differentiate or bind to the niche. To formulate the model, some additional assumptions were made. The total number of niches were assumed to remain constant. This is in contrast to the work of Becker et al. (2019), where the production and clearance of niche-cells were linked dynamically to the binding with HSC. Similar interaction could be included in our proposed HSC-model, by modifying the behaviour of empty niche N_E in equation (8.7a). However, the exact nature of how niches are produced are unknown, and we decided to omit dynamic niche-counts in our model. Some initial numerical investigations of

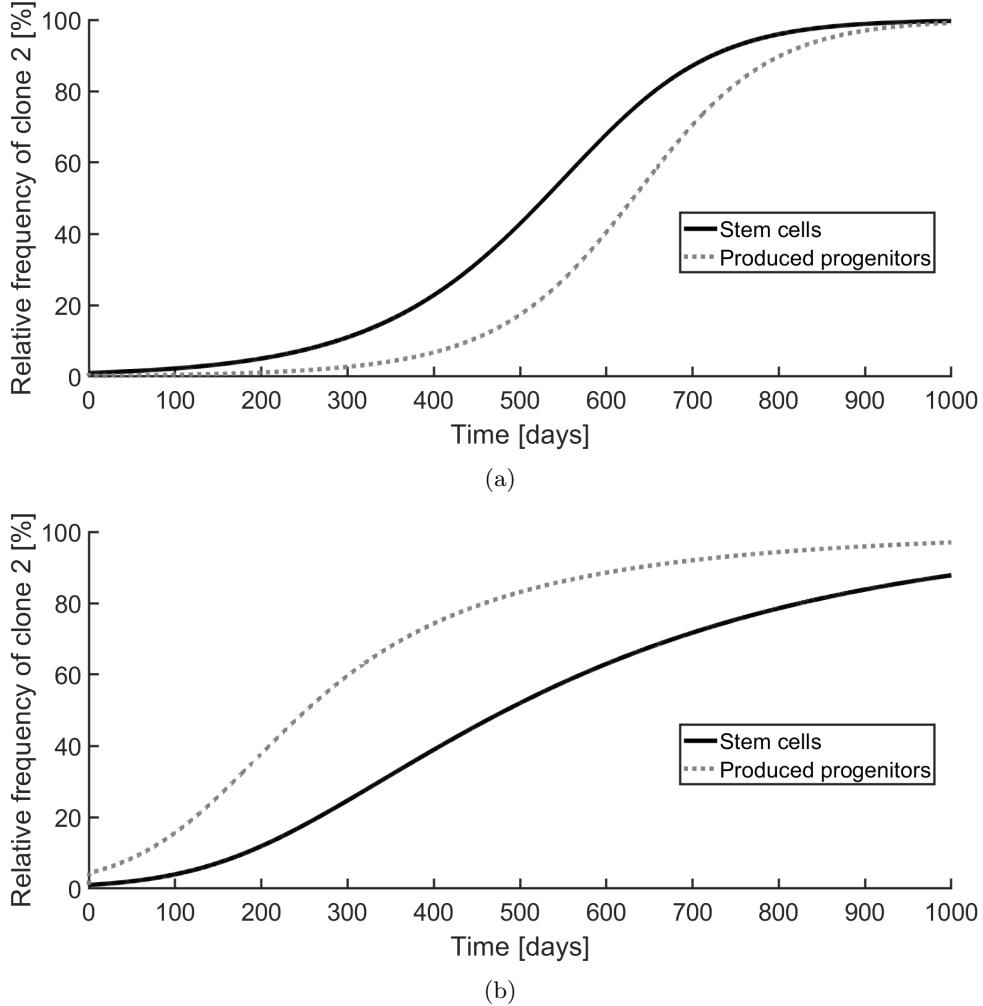


Figure 8.4: Production of progenitors composition does not necessarily reflect HSC composition. Two scenarios of clonal competition between two clones are simulated. In both cases, clone 2 has a competitive advantage and initially makes up 1% of all HSC. In panel (a) the fitness advantage of clone 2 is small, $r_2 = 1.2r_1$, but the release of the second clone is much faster than the first clone, $u_2 = 5u_1$. Due to the fitness advantage of the second clone, the relative frequency of clone 2 is increasing. The production of progenitor cells appears to be delayed compared to the time evolution of the stem cells. In the scenario depicted in panel (b), the competitive advantage is bigger, with $r_2 = 4r_1$. However, a slower release of HSC from the niche, $u_2 = 0.4u_1$ results in a development where the relative frequency of progenitor cells pertaining to the second clone increases earlier than the relative frequency of HSC from clone 2 out of all HSC.

implementing such differences did however show that as long as the number of niches are near a steady-state count, the effect of production and clearance of niches have only an insignificant effect on HSC-dynamics. In the absence of niches, HSC-production cannot be maintained. In our model, this occurs because HSC are lost to differentiation. Additional loss of HSC due to HSC apoptosis could be included, however, we assumed such stem cell death to be rare and omitted it from the model formulation. Note that including HSC apoptosis does not change the properties and dynamics of the model, since the effective loss of cells could be included in the rates for loss of cells due to differentiation, d_{A_j} and d_{I_j} . However, this would necessitate a different formulation of the production of progenitor cells. While such a reformulation has little effect on the numerical results discussed at the end of this section, it could have an effect in other scenarios. This is relevant in later parts of this thesis, and will be discussed in more detail then. Whether HSC differentiate during division, due to environmental factor, or both, is uncertain (Wilson and Trumpp, 2006). In our proposed model, active stem cells are modelled to always divide before differentiating (i.e. one HSC divides into two progenitors), while inactive cells differentiate without division (i.e. one HSC differentiate spontaneously and one progenitor arise). Hence both types of differentiation is considered in the model, and the distinction between active and inactive provides an explanation for both to occur naturally. However, we note that requiring division before differentiation of inactive cells only leads to a reformulation of the progenitor-production, and does not change the model. The model describes symmetric division of active stem cells, that is division leading to either two HSC or two differentiated cells. While the model thus does not explicitly consider asymmetric division, we showed in section 8.2 that the resulting model would be equivalent if asymmetric division was included.

Other mathematical models of stem cell behaviour explicitly models the cell cycle, and puts a time-dependence on HSC-quiescence. A notable example is the model of Michael Mackey, proposed in Mackey (1978) and analysed in detail throughout the literature, as described in section 4. In our model, niche-bound HSC are considered quiescent, and the detachment from the niche is modelled to occur spontaneously, and hence independent of how long the HSC had been niche-bound. If quiescence of HSC must be maintained for an extended time before the HSC regains its self-renewing potential, the model could include an explicit time-delay from attachment of inactive HSC to detachment as active HSC. While this could have a significant effect on the dynamics of the system, in particular when simulating HSC transplantation, we chose to omit such effects of time-delays for simplicity.

8.4.2 Fitness and structure of steady states

Analysis of the model revealed that an attractive non-negative trapping region exists, and hence that solutions to the system with non-negative initial conditions are non-negative for all time t . This is important for the biological applicability of the model as solutions that lead to non-feasible variables or variables that approach infinity would not correctly model reality.

Within the trapping region, a number of steady state of the HSC niche model exists. A trivial steady state S_0^* always exists, in which all niches are empty. Under conditions on parameters, n single-clone steady states S_j^* exist in which only cells pertaining to a given clone, j , are non-zero. The specific conditions depend only on the properties of the clone as well as the total number of niches. This gives rise to the notion of a stem cell fitness, F_j . When multiple cells have the same fitness, co-existences steady states exist. These steady states exist along a k -simplex, where k is the number of clones with equal fitness.

Local stability of steady states where determined to depend on the relative fitness of the different HSC-clones considered. When one clone had higher fitness than all other clones, the related single-clone steady state was locally stable, while all other steady states were locally

unstable. Hence, the use of the term fitness was justified, as high fitness of one clone leads to out-competition of other clones in the long-term. The expression of the fitness of a given clone does not depend on properties of the cells it is competing with. Hence, the model suggests that it is possible to determine the fitness of a clone by investigating it in isolation. This could theoretically imply that the fitness of two clones could be determined independently and the outcome of clonal competition could be predicted. This could be beneficial in HSC transplantation, as it could be used to predict if a transplanted HSC clone will out-compete a malignant clone in the patient.

Equal fitness of clones lead to a complicated scenario, where the co-existence steady states are highly degenerate. It is possible for multiple groups of equal fitness to arise, such as e.g. two clones with fitness F_a , and three other clones with fitness F_b where $F_a \neq F_b$. In such a scenario, a line of steady states between single-clone steady states of the clones with fitness F_a exists, while a plane of steady states also exists, going through the single-clone steady states of the clones with fitness F_b . The complexity of such scenarios makes them difficult to analyse, however, we assume that equal fitness of clones is unlikely to arise in a biological system, and is an artefact of the simplicity of the modelling approach. However, the co-existence steady states are important from a mathematical view, as they represent the boundary between biologically attainable situations, that is, between the scenario where one clone has the higher fitness, and one where another clone has.

8.4.3 Model parametrization and numerical findings

The HSC niche model was parametrized using data from Bhattacharya et al. (2009). While the data arose from a mouse HSC-transplantation study, we assumed that the dynamics of HSC in mice are similar enough to the dynamics in the human body. As the experimental data of Bhattacharya et al. (2009) came from studies of HSC clone expected to have similar properties, it was only necessary to determine one set of cell-properties, and use those parameter-values for both clone in the experiment. However, we emphasize that the parameters are indeed still uncertain, and comparing the model behaviour to additional HSC data could lead to other estimates for parameters. The default parameters of the HSC niche model given in table 8.1 should thus be considered first estimates of HSC parameters. Having an estimate of the parameters, albeit an uncertain one, allowed us to investigate the model numerically. A range of biologically relevant scenarios were simulated, and the relation between HSC properties and the outcome of e.g. HSC transplantation could be found in the model. The model can be used as a tool to suggest possible transplantation schemes that would not be possible or ethical to investigate *in vivo*. In particular, we simulated the effect of drug-induced HSC-mobilization combined with HSC-transplantation and found that short-term mobilization can be beneficial when treating malignancies with HSC-transplantation.

Investigating how differences between clones affect the production of progenitors, we found that there can be significant discrepancy between the relative frequency of different clones when comparing the HSC and the production of progenitors, both transiently after sudden changes and sustained for clones with equal or similar fitness. As the progenitors in turn give rise to the mature blood-cells, this result implies that the relative frequency of clones can be significantly different when comparing HSC and blood cells. Note that this difference can go either way; it is both possible for a HSC clone (healthy or malignant) to be over-represented in the blood and to be under-represented in the blood. Hence, our investigations suggest that great care must be taken to ensure that the composition of clones observed in blood is indicative of the composition of HSC in the bone-marrow, particularly when malignancies are considered.

Chapter 9

Model reduction

We here present the mathematical analysis of the HSC niche model described above, as well as a number of model assumptions that lead to different reduced and simplified forms of the HSC niche model. The majority of the considerations described in this chapter were also described in (Pedersen et al., nd2), available in section C.

9.1 Reducing the HSC niche model

For simplicity, we first assume that $b_{A_j} = 0$ is valid for all j . The considerations about steady states and local stability described in the previous section is unchanged by this assumption. b_{A_j} is the rate at which active stem cells re-attach to an empty niche immediately after detaching. Hence, the simplification is equivalent to considering the population of cells that detach and immediately re-attach as a subpopulation of the bound HSC. While other free cells could intercept such immediate re-attachment we assume this occurs rarely, and hence that A_j represent the free HSC that are committed to division or differentiation.

Two additional assumptions form the basis for the model reductions we consider. These are:

Assumption R1 Most HSC niches are occupied.

Assumption R2 Niche-bound HSC make up the majority of the total HSC population.

Assumption R1 is based on experimental evidence from Bhattacharya et al. (2009), while we base Assumption R2 on the low frequency of HSC division (Lee-Six et al., 2018) and studies of HSC mobilization in the literature (Yang et al., 2009).

From mice experiments, Bhattacharya et al. (2009) conclude that only few niches were available for engraftment of transplanted HSC. Unless the total number of HSC is small, this implies that Assumption R1 is a reasonable biological assumption. In the niche model, Assumption R1 implies that the number of empty niches, N_E , at steady state should be low compared to the number of occupied niches, $\sum_{i=1}^n N_i$. From the steady state considerations discussed in the previous section, we found that the steady state values of A_j and I_j scale with u_j , while N_E and N_j were independent of u_j . The steady state value of N_E was however determined to scale with $b_{I_j}^{-1}$. Hence, if the dimensionless quantity $\frac{u_j}{b_{I_j}} \ll 1$, Assumption R1 is fulfilled.

For purposes of reducing the niche model, Assumption R2 is interpreted as niche-bound HSC counts, N_j , being numerous compared to A_j and I_j for all j in steady state. When $b_{A_j} = 0$, both the steady state value of A_j and I_j scales with $\frac{u_j}{r_j + d_{A_j}} N_j$, as seen in the previous section. Hence, if the dimensionless quantity $\frac{u_j}{r_j + d_{A_j}} \ll 1$, Assumption R2 is fulfilled.

To show how these assumptions and our model-interpretation of them leads to a reduced form of the model, we first present a scaling of variables. Letting superscript o denote the unscaled variables, the scaled variables are denoted with the original notations. The following scalings are chosen:

$$N_j^o = KN_j \quad (9.1)$$

$$I_j^o = \frac{u_j}{b_{I_j}} I_j \quad (9.2)$$

$$A_j^o = \frac{u_j}{r_j} K A_j \quad (9.3)$$

In addition, we define U as the smallest u_j , that is, the specific u_j such that $U = u_j \leq u_k$ holds for all $k \in [1, n]$. Introducing a scaled time $\tau = Ut$, the scaled model is:

$$\frac{dN_j}{d\tau} = \frac{u_j}{U} \left(\left(1 - \sum_{i=1}^n N_i \right) I_j - N_j \right) \quad (9.4a)$$

$$U \frac{dI_j}{d\tau} = 2\gamma b_{I_j} K A_j - b_{I_j} K \left(1 - \sum_{i=1}^n N_i \right) I_j - d_{I_j} I_j \quad (9.4b)$$

$$U \frac{dA_j}{d\tau} = r_j N_j - (r_j + d_{A_j}) A_j \quad (9.4c)$$

Note that N_j , I_j and A_j now denote the scaled variables. While the domain of the scaled variables I_j and A_j is still the non-negative real numbers $\mathbb{R}_{\geq 0}$, the scaling of niche-bound HSC has solutions $N_j(t) \in [0, 1]$. This also implies a scaling of the empty niches: $N_E = 1 - \sum_{i=1}^n N_i$, also on $[0, 1]$.

Equation (9.4b) can be written as:

$$\frac{U}{b_{I_j} K} \frac{dI_j}{d\tau} = 2\gamma A_j - \left(1 - \sum_{i=1}^n N_i \right) I_j - \frac{d_{I_j}}{b_{I_j} K} I_j \quad (9.5)$$

In general $K > 1$, and hence $\frac{U}{b_{I_j} K} \leq \frac{u_j}{b_{I_j} K} < \frac{u_j}{b_{I_j}}$. Under Assumption R1 we thus have $\frac{U}{b_{I_j} K} \ll 1$ which suggests the approximation $\frac{U}{b_{I_j} K} \rightarrow 0$ is appropriate. Hence a quasi-steady-state approximation of $\frac{dI_j}{d\tau}$ arise, in which the left-hand-side of equation (9.5) is set to zero.

Lemma 9.1.1: Quasi-steady-state approximation, I_j

The expression

$$I_{j,red} = \frac{2\gamma}{\frac{d_{I_j}}{b_{I_j} K} + 1 - \sum_{i=1}^n N_i} A_j \quad (9.6)$$

is a quasi-steady-state approximation of equation (9.4b), which is valid under Assumption R1.

Hence $I_{j,red}$ scales with A_j . Dividing equation (9.4c) by $r_j + d_{A_j}$, we obtain

$$\frac{U}{r_j + d_{A_j}} \frac{dA_j}{d\tau} = \frac{r_j}{r_j + d_{A_j}} N_j - A_j \quad (9.7)$$

Since Assumption R2 implies that $\frac{u_j}{r_j+d_{A_j}} \ll 1$ holds for all j , and $\frac{U}{r_j+d_{A_j}} \leq \frac{u_j}{r_j+d_{A_j}}$, we approximate $\frac{U}{r_j+d_{A_j}} \rightarrow 0$. A quasi-steady-state approximation of $\frac{dA_j}{d\tau}$ is then appropriate, as the left-hand-side of equation (9.7) is zero.

Lemma 9.1.2: Quasi-steady-state approximation, A_j

The expression

$$A_{j,red} = \frac{r_j}{r_j + d_{A_j}} N_j \quad (9.8)$$

is a quasi-steady-state approximation of equation (9.4c), which is valid under Assumption R2.

The two quasi-steady-state approximations allows us to define the reduced model. We first define the domain \mathbb{T} on which solutions of the reduced model exist.

Definition 9.1.1: Reduced domain, \mathbb{T}

We define the *reduced domain* as:

$$\mathbb{T} = \{(N_1, \dots, N_n) : N_j \geq 0, j = 1, \dots, n \wedge \sum_{j=1}^n N_j \leq 1\} \quad (9.9)$$

Replacing A_j in equation (9.6) with the expression for $A_{j,red}$ in lemma 9.1.2 yields an expression for $I_{j,red}$ which depends only on parameters and N_j for all j . Using $I_{j,red}$ in equation (9.4a) hence leads to a model which has solutions that can be described on the domain \mathbb{T} .

Definition 9.1.2: Reduced Model

On the domain \mathbb{T} , the *Reduced Model*, is given by:

$$\dot{N}_j = u_j \left(\frac{2\gamma\rho_j(1 - \sum_{i=1}^n N_i)}{\alpha_j + 1 - \sum_{i=1}^n N_i} - 1 \right) N_j \quad (9.10)$$

where $\rho_j = \frac{r_j}{r_j+d_{A_j}}$ and $\alpha_j = \frac{d_{I_j}}{b_{I_j}K}$, for all j . The variables I_j and A_j are given by equations (9.6) and (9.8) respectively.

Note that since all parameters of the niche model are positive, it must hold that $0 \leq \rho_j \leq 1$, since $\rho_j > 1$ would imply $d_{A_j} < 0$.

Basic analysis of steady state and stability of the reduced model is completely analogue to the analysis of the full model, albeit shorter and simpler. For this reason, we summarize the results rather than going through the calculations here. Details are described in (Pedersen et al., nd2). Before summarizing the results however, two definitions are useful. These definitions are used throughout the analysis of the reduced model, but also prove useful in later parts of this thesis.

Definition 9.1.3: Scaled fitness

The scaled fitness of the j^{th} clone is defined as $f_j = KF_j$, which can be written as:

$$f_j = \frac{2\gamma\rho_j - 1}{\alpha_j} \quad (9.11)$$

where $\rho_j = \frac{r_j}{r_j+d_{A_j}}$ and $\alpha_j = \frac{d_{I_j}}{b_{I_j}K}$.

When there is no ambiguity, the scaled fitness will be referred to as just *fitness*.

Definition 9.1.4: Reduced pre-factor

We define the function $g_j(N_1, \dots, N_n)$ as

$$g_j(N_1, \dots, N_n) = u_j \alpha_j f_j \left(\frac{1 - \sum_{i=1}^n N_i - f_j^{-1}}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) \quad (9.12)$$

where $\rho_j = \frac{r_j}{r_j+d_{A_j}}$ and $\alpha_j = \frac{d_{I_j}}{b_{I_j}K}$.

As the functional dependence of $g_j(N_1, \dots, N_n)$ is clear, we write just g_j in most cases. Note that the expression for j^{th} clone of the reduced model can be written $\dot{N}_j = g_j N_j$.

As for the full model, a trivial steady state exists regardless of parameter values.

Theorem 9.1.1: Trivial steady state S_0^* , reduced form

A *trivial steady state*, denoted S_0^* , always exists. In the trivial steady state $N_j^* = 0$, $\forall j \in \{1, \dots, n\}$.

When $0 < f_j^{-1} < 1$, a feasible single-clone steady state exists.

Theorem 9.1.2: Single-clone steady state S_j^* , reduced form

Given $2\gamma\rho_j > 1$ and $f_j^{-1} < 1$, a *single-clone steady state* exists for which $N_j^* > 0$ while $\forall k \neq j : N_k^* = 0$. In particular, $N_j^* = 1 - f_j^{-1}$ and $N_E^* = f_j^{-1}$.

We denote the j^{th} single-clone steady state as S_j^* .

Co-existence steady states exists for the reduced model, analogous to the co-existence steady states of the full model. These are characterized by equal fitness of two or more clones, and constitute a k -dimensional simplex of steady states where k is the number of clones with equal fitness. For all steady states on the simplex, $N_E^* = f_j^{-1}$ and hence $\sum_{i \in \mathcal{C}} N_i = 1 - f_j^{-1}$ where \mathcal{C} is the set of clones with equal fitness.

Stability considerations of the reduced model mirror the full system. The single-clone steady state of any clone with fitness higher than all other clones is locally stable while all other steady state are locally unstable. For cases where a number of clones have equal fitness, a simplex of steady states exists and numerical investigations have shown that an eigenvector along the simplex has corresponding eigenvalue zero.

Thus, we have found the reduced model maintains the same stability structure as the full model, with the same dependence on fitness.

When evaluating the k^{th} reduced pre-factor, g_k , in the steady state pertaining to a clone $j \neq k$, g_k can be written:

$$g_k|_{S_j^*} = \frac{u_k}{\alpha_k + \frac{1}{f_j}} \left(\frac{f_k}{f_j} - 1 \right) \quad (9.13)$$

Hence, for $f_k > f_j$, $g_k|_{S_j^*} > 0$, while $f_k < f_j$ implies that $g_k|_{S_j^*} < 0$. As the sign of \dot{N}_k is the same as g_k , we see that the sign of \dot{N}_k at the j^{th} steady state depends only on the relation between fitnesses f_k and f_j .

9.2 Production of progenitors in the reduced niche model

Before introducing further simplifications of the niche model, we here show that the production of progenitors HSC give rise to can be written in terms of the reduced parameters ρ_j and α_j . This is relevant when considering the effect that changes to the reduced model has on blood-cell production since some parameter-changes can affect progenitor production and must be carefully considered in further work using the niche model, described in part V of this thesis.

In terms of the unscaled variables, production of progenitors was given in definition 8.1.3 as $i_{D_j} = 2d_{A_j} A_j^o + d_{I_j} I_j^o$. By replacing A_j^o and I_j^o by the scaled variables, we see:

$$i_{D_j} = 2d_{A_j} A_j^o + d_{I_j} I_j^o \quad (9.14a)$$

$$= 2d_{A_j} \frac{u_j}{r_j} K A_j + d_{I_j} \frac{u_j}{b_{I_j}} I_j \quad (9.14b)$$

$$= \left(2 \frac{d_{A_j}}{r_j} K + \frac{d_{I_j}}{b_{I_j}} \frac{2\gamma}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) u_j A_j \quad (9.14c)$$

$$= \left(2 \frac{d_{A_j}}{r_j} + \frac{2\gamma\alpha_j}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) u_j K \rho_j N_j \quad (9.14d)$$

$$= \left(2(1 - \rho_j) + \frac{2\gamma\alpha_j \rho_j}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) u_j K N_j \quad (9.14e)$$

For the last step, observe that $\frac{2d_{A_j}}{r_j} \rho_j = \frac{2d_{A_j}}{r_j + d_{A_j}} = \frac{2r_j - 2r_j + 2d_{A_j}}{r_j + d_{A_j}} = 2(1 - \rho_j)$. Hence, the production of progenitors of the j^{th} clone can be written in terms of the niche-bound HSC of the reduced model, N_j , and the set of parameters related to the j^{th} clone; u_j , ρ_j , α_j and K .

Definition 9.2.1: Production of progenitors, reduced model

Production of progenitor cells of the reduced model in definition 9.1.2 is defined for the j^{th} clone as:

$$i_{D_j} = \left(2(1 - \rho_j) + \frac{2\gamma\alpha_j \rho_j}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) u_j K N_j \quad (9.15)$$

Since $0 \leq \rho_j \leq 1$ by definition of the reduced parameters, $\frac{i_{D_j}}{N_j}$ is always non-negative, and hence any non-negative number of HSC N_j results in a non-negative production of progenitors, i_{D_j} .

9.3 The transformed HSC niche model

As discussed previously, the relative frequency of a given clone is an important clinical measure, used in case of malignancies to e.g. identify the progression of disease for a patient. Additionally, the total HSC cell-count can be easier to estimate than the numbers of cells from a given clone.

Excluding the trivial steady state, we define the set $\mathbb{T} \setminus S_0^*$ in \mathbb{R}^n . On this set, transformed variables are defined. The sum of total niche-bound cells is given as:

$$T = \sum_{i=1}^n N_i \quad (9.16)$$

and for $j \in (2, 3, \dots, n)$ we define the relative frequency of the j^{th} clone as:

$$C_j = \frac{N_j}{\sum_{i=1}^n N_i} \quad (9.17)$$

The relative frequency of the 1^{st} clone, C_1 , is calculated as $C_1 = 1 - \sum_{i=2}^n C_i$. Note that $T > 0$ holds since the trivial steady state $S_0^* = \{(0, \dots, 0)\}$ was omitted, and hence C_j is well-defined for all j .

The definition of T and C_j allows for a transformation of the reduced model.

Definition 9.3.1: Transformed Model

On the domain $(0, 1] \times [0, 1]^{n-1}$ the *transformed model* is given as:

$$\dot{T} = T \sum_{i=1}^n g_i(T) C_i \quad (9.18a)$$

$$\dot{C}_j = \left(g_j(T) \sum_{i=1}^n C_i - \sum_{i=1}^n g_i(T) C_i \right) C_j \text{ for } j = 2, \dots, n \quad (9.18b)$$

$$C_1 = 1 - \sum_{i=2}^n C_i \quad (9.18c)$$

where $g_j(T)$ is given as

$$g_j(T) = u_j \alpha_j f_j \left(\frac{1 - T - f_j^{-1}}{\alpha_j + 1 - T} \right) \quad (9.19)$$

A transformation \mathcal{F} exists:

$$\begin{aligned} \mathcal{F} : (N_1, \dots, N_n) \in \mathbb{T} \setminus S_0^* \subset \mathbb{R}^n &\mapsto (T, C_2, \dots, C_n) = \\ &\left(\sum_{i=1}^n N_i, \frac{N_2}{\sum_{i=1}^n N_i}, \dots, \frac{N_n}{\sum_{i=1}^n N_i} \right) \in (0, 1] \times [0, 1]^{n-1} \subset \mathbb{R}^n \end{aligned} \quad (9.20)$$

with an inverse function:

$$\begin{aligned} \mathcal{F}^{-1} : (T, C_2, \dots, C_n) \in (0, 1] \times [0, 1]^{n-1} \subset \mathbb{R}^n &\mapsto (N_1, \dots, N_n) = \\ &\left(T(1 - \sum_{i=2}^n C_i), T C_2, \dots, T C_n \right) \in \mathbb{T} \setminus S_0^* \subset \mathbb{R}^n \end{aligned} \quad (9.21)$$

The two functions \mathcal{F} and \mathcal{F}^{-1} are bijective and transforms points between the domain of the reduced model and the domain of the transformed. Hence, a notion of equivalence between the transformed model and the reduced model exists.

Lemma 9.3.1: Equivalence of the reduced model and the transformed model

When excluding the trivial steady state S_0^* from the domain of the reduced model, the transformed model is equivalent to the reduced model from definition 9.1.2 in the sense that solutions of one model can be transformed to equivalent solutions of the other model. In particular \mathcal{F} of equation (9.20) and its inverse transforms solutions between the two models.

In the transformed model, the single-clone steady states of the reduced model, S_j^* , have corresponding transformed steady states:

$$\bar{S}_1 = \begin{pmatrix} 1 - f_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \bar{S}_2 = \begin{pmatrix} 1 - f_2 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \bar{S}_3 = \begin{pmatrix} 1 - f_3 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \quad \bar{S}_n = \begin{pmatrix} 1 - f_n \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad (9.22)$$

where \bar{S}_j describes $(T, C_2, C_3, \dots, C_n)$.

For equal fitness between clones, we consider as an example where clone 2 and 3 have equal fitness. This would allow for a line of steady states given as:

$$\bar{S}_{2,3} = \begin{pmatrix} 1 - f \\ \beta \\ 1 - \beta \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (9.23)$$

where β is a number between 0 and 1, and the fitness of clone 2 and 3 is $f = f_2 = f_3$. As for the full HSC model, k -dimensional simplexes exists when k clones have the same fitness. The structure when multiple clones have the same fitness is thus maintained.

In general, different HSC clones can arise that are functionally similar to each-other albeit genetically different. For simplification, we consider a scenario in which all healthy clones are considered a single sub-population, and only a single clone that differ from the main population is considered. This allows us to consider disease scenarios or HSC-transplantation scenarios by modelling just two distinct clones. Writing the relative frequency of the second clone, C_2 as simply C , the transformed model can be written as:

$$\dot{T} = [g_1(T)(1 - C) + g_2(T)C] T \quad (9.24a)$$

$$\dot{C} = (g_2(T) - g_1(T))(1 - C)C \quad (9.24b)$$

This model is referred to as the *two-clone transformed model*. In the supplementary material of (Pedersen et al., nd2) this version of the model is analysed in detail, and the global dynamics are determined. We write the two-clone versions of the steady states:

$$\bar{S}_1 = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix}, \quad \bar{S}_2 = \begin{pmatrix} 1 - f_2^{-1} \\ 1 \end{pmatrix} \quad (9.25)$$

Theorem 9.3.1: Global dynamics of the transformed two-clone system

For $f_1 \neq f_2$, the system of equations (9.24) has just two steady states, \bar{S}_1 and \bar{S}_2 , as defined in equations (9.25). There are no periodic solutions.

For $f_1 > f_2$, all solutions with $C(0) < 1$ are attracted toward \bar{S}_1 .

For $f_1 < f_2$, all solutions with $C(0) > 0$ are attracted toward \bar{S}_2 .

9.4 The logistic approximation in the two-clone case

As shown in equation (9.13), evaluating the reduced pre-factor of the k^{th} clone in the steady state of the j^{th} clone shows a simple relation between the sign of \dot{N}_k and the relative fitness between clone k and j . This importance is carried over to the transformed model, and paves the way for an additional simplification of the two-clone transformed model. Evaluating $g_2(T)$ in the steady state \bar{S}_1 we obtain

$$g_2(T_1^*) = u_2 \frac{\alpha_2}{\alpha_2 + f_1^{-1}} \left(\frac{f_2}{f_1} - 1 \right) \quad (9.26)$$

where $T_1^* = 1 - f_1^{-1}$. Note that $g_1(T_1^*) = 0$.

The final reduction of niche model assumes that $\dot{T} = 0$ is appropriate and that the behaviour of the model in general resembles the behaviour close to \bar{S}_1 . From these assumptions, a logistic equation arises, in which equation (9.26) determines the asymptotic behaviour.

Definition 9.4.1: Logistic Approximation of the relative frequency of the 2nd clone

The *Logistic Approximation Model* is given as

$$\dot{C} = \phi(1 - C)C \quad (9.27)$$

where $\phi = g_2(T_1^*) = u_2 \frac{\alpha_2}{\alpha_2 + f_1^{-1}} \left(\frac{f_2}{f_1} - 1 \right)$. The sum of HSC is constant with $T = 1 - f_1^{-1}$.

Note that ϕ has units of $[time]^{-1}$, due to the factor u_2 .

Determining the steady states of equation (9.27) is trivial: One has $C = 0$, corresponding to \bar{S}_1 , and the other has $C = 1$, corresponding to \bar{S}_2 . The sign of ϕ determines whether \dot{C} is positive or negative for $0 < C < 1$, and determining the sign of ϕ is enough to determine the asymptotic behaviour. For $\phi > 0$, the steady state $C = 1$ is attracting, while $C = 0$ is attracting for $\phi < 0$. As $f_2 > f_1$ implies $\phi > 0$, and $f_2 < f_1$ implies $\phi < 0$, the stability of the steady states for the logistic approximation is in agreement with the transformed model and in turn the full HSC niche model.

Changes to parameters affect the logistic approximation similar to the other forms of the model. However, as an artefact of the approximation, the parameter u_1 has no effect on ϕ and hence on the logistic approximation. However, numerical investigations based on the default values of parameters determined that u_1 and u_2 had to be much different for the omission of u_1 to cause significant differences between the logistic approximation and the two-clone transformed model.

In section 3, we discussed data from patients with the JAK2 mutation, from the DALIAH clinical trial. As described previously in (Pedersen et al., 2020), the kinetics of the JAK2 allele burden was approximately exponential for both patients not receiving treatment and patients receiving IFN treatment. For low numbers, exponential growth is similar to the growth of a logistic equation, such as the logistic approximation of the HSC niche model. In figure 9.1a data for untreated growth of the JAK2 allele burden is shown, together with patient-specific fits of ϕ in the logistic approximation. This shows that the single-parameter logistic approximation allows for growth-curve in good agreement with patient-data. Similarly, figure 9.1b depicts patient from the DALIAH trial under IFN treatment. Here the logistic approximation also allows for a fit to data which visually appears to agree well with data. Hence, the logistic approximation provides a one-parameter fit to JAK2 allele burden data, both before and during treatment, suggesting that the underlying mechanism giving rise to both disease growth and decay could be explained by the HSC dynamics captured in the HSC niche model.

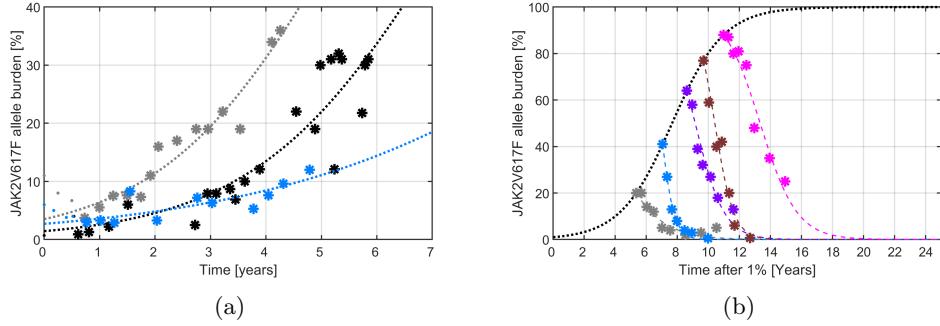


Figure 9.1: The logistic approximation agrees with data from the DALIAH trial. Ignoring changes in IFN-dose, the logistic approximation of definition 9.4.1 is fit to patient-data. In panel (a) data is shown from the patients used to determine the exponential growth of the JAK2 allele burden in (Pedersen et al., 2020). As in (Pedersen et al., 2020), the first 200 days were not used in the fit, since all three patients previously received treatment. In panel (b) data for the JAK2 allele burden of five patients of the DALIAH trial is shown as colored *, time-shifted to coincide at baseline with growth similar to the fits shown in panel (a), shown in dotted black. Dashed colored lines depict the logistic approximation with a fitted value of ϕ which minimizes the sum of squared differences between the model and the data.

9.5 Summarizing discussion

We have presented a series of model reductions, leading to reduced and simpler forms of the HSC niche model. Through assumptions based on biology, we showed that two quasi-steady-

state approximations of the model are appropriate. For n distinct clones, this reduced the $3n$ -dimensional HSC niche model to a reduced model of n dynamic variables. With the reduction, the HSC-count was scaled with the total number of available niches, leading to an expression of scaled fitness, f_j . As the reduction was based on quasi-steady-state approximations, the reduced model has the same steady states, and analysis reveals that the local stability of the steady states depends on the scaled fitness in the same way as the full non-scaled model depended on the non-scaled fitness. Hence, the general structure of the model is maintained when reducing the model. In addition, we found that production of progenitors could be described in terms of the reduced parameters of the reduced model, allowing for an interpretation of the behaviour of the reduced model to the production of progenitors and in turn the production of mature blood cells.

As discussed in section 3, two important measures of disease are the count of mature cells and the relative frequency of a malignant clone, expressed through the JAK2 allele burden for the MPNs. Introducing a transformation of variables, it is possible to formulate a transformed form of the reduced model in which the total sum of HSC, T , and the relative frequency of a given clone j , C_j , are the dynamic variables. The resulting model is hence more directly comparable to clinical measures. Considering a two-clone scenario, it is possible to show that the transformed model for two clones cannot have periodic solutions, and determine the basin of attraction for the possible steady states under all circumstances.

Finally, we showed that when a malignant clone makes up a small part of the cells, the two-clone transformed system can be approximated by a one-parameter logistic expression. The sign of the single parameter of the logistic equation determines which clone out-competes the other.

Interestingly, the HSC niche model and the reduced and simplified forms were determined to depend on the fitness in a similar way. Whenever one clone had a higher fitness than all other clones, it was found to out-compete the others and drive them to extinction. This suggests that the concept of fitness could be an inherent feature of HSC competition, and draw parallels to the types of competition seen in ecology such as between populations of animals. While it is possible that the HSC niche model omits important features of HSC behaviour, asymptotic behaviours with similar dependence on HSC fitness could arise, and hence the notion of HSC fitness must be investigated further, both mathematically and experimentally.

The HSC niche model, reduced niche model, the transformed model and the logistic approximation provide four different levels of complexities of modelling HSC. While most detail about biological features of HSC could be obtained through comparing the HSC niche model with experimental or clinical data, challenges of identifiability arise due to the number of model parameters.

From the analysis described in section 3.2 and originally presented in (Pedersen et al., 2020), data for the IFN-treated MPN patients in the DALIAH trial showed a decline in JAK2 allele burden which in many case resembled a simple exponential decay. In addition, we found that the increasing JAK2 allele burden of untreated patients resembled exponential growth in the early stages for the patients considered in (Pedersen et al., 2020). For the JAK2 mutation, the allele burden of HSC has been found to agree well with the allele burden of mature cells (Takahashi et al., 2013), and hence comparing the data of blood measurements from the DALIAH trial with the HSC niche model and the related reduced forms is reasonable for the JAK2 mutation. The simplest form, the logistic approximation, is capable of capturing both the initial increase in JAK2 allele burden and the IFN-induced decay, as shown in figure 9.1. Hence, all other forms of the model will fit at least as well to the patient-data. This demonstrates the challenge with identifiability of parameters, since the logistic approximation is a single-parameter model that agrees with data. The full HSC niche model for two clones has 14 distinct parameters, and identifying each parameter from this data is not reasonable. As such, model parameters identified

by the relative frequency of a malignant clone are shrouded in uncertainty, and consequently interpretation of model results will also be uncertain.

Coincidentally, the success of the logistic approximation to fit to data also explains the success of using an exponential expression to describe data in our work in (Pedersen et al., 2020). Hence, the HSC niche model and the reduced and simplified forms lay a theoretical foundation for interpretation of the otherwise empirical model, which can be used to understand the dynamics of the JAK2 allele burden or similar measure for other hematologic malignancies.

Part V

Combining the Cancitis model and the bone-marrow model

Chapter 10

Formulation of a combined Cancitis-Niche model

While fits of the Cancitis model to the data from the DALIAH trial, as described in chapter 5 and in (Ottesen et al., 2020) results in good fits to patient data and a reasonable hypothesis about the biological response to IFN, the mathematical interpretation of the effect of IFN is still uncertain. In particular, IFN was previously modelled to interact only with the rate of apoptosis of malignant stem cells, i.e. through the parameter \hat{d}_{y_0} . In the following, the Cancitis model and the niche model will be combined to allow for a IFN-interaction with parts of the HSC system not considered in the Cancitis model.

Before combining the two models, we first present some general considerations about how mature blood-cells relate to the production of progenitors due to HSC.

10.1 Connecting HSC and mature cells

From differentiation of HSC to the mature cells that constitute the blood, multiple stages of specialized progenitor cells give rise to the increase in numbers that occurs from few HSC to the abundant blood-cells. Progenitor cells are generally stem-like cells, characterized by a limited capacity for self-renewal, which cannot regain stem-ness. In the niche model, differentiating HSC were simply considered a loss of cells. However, to also model the blood cells, we must carefully consider how differentiated HSC give rise to progenitors which in turn give rise to the mature blood-cells. For simplicity, we refer to all intermediate steps between HSC and blood as progenitors, and consider no significant differences between different cell lines such as leukocytes or thrombocytes. We here present an argument for making a quasi-steady-state approximation for the progenitors. Such simplification was done implicitly in the formulation of the Cancitis model in (Andersen et al., 2017).

We first describe the biological assumptions that we consider:

Assumption P1 When hematopoietic progenitors cells divide, the resulting cells are more likely to be differentiated cells than cells identical to the original cell.

Assumption P2 Differentiation of progenitor cells occurs more frequently than HSC are released from the HSC-niches.

Assumption P1 is in agreement with the idea that progenitor self-renewal is limited and progenitor cells cannot sustain their numbers indefinitely. The existence of multiple intermediate

steps of progenitors between HSC and mature cells and the slow time-scale of HSC division suggest that individual cells must differentiate relatively quickly for stem cell changes to have an effect on the mature blood cells. This is the base of assumption P2.

We consider a general formulation of a mathematical model of stem cells, progenitor cells and mature blood cells. The model is illustrated in figure 10.1. Consider an ODE for HSC defined as:

$$\dot{N} = u(f(N) - d_0 N) \quad (10.1)$$

where, for generality of the argument, $f(N)$ is an unspecified positive function and ud_0 denotes a rate at which differentiated cells are produced. First stage of differentiated cells are given as:

$$\dot{P}_1 = ud_0 N + r_1 P_1 - d_1 P_1 \quad (10.2)$$

where r_1 is self-renewal of progenitors and d_1 is the rate of differentiation. Apoptosis of progenitors are assumed to be rare enough to be negligible. The differentiating cells, $d_1 P_1$ give rise to two cells of the next stage of progenitors, P_2 . For k stages of progenitors, we write up the j^{th} stage as:

$$\dot{P}_j = 2d_{j-1} P_{j-1} + r_j P_j - d_j P_j \quad (10.3)$$

Following the k 'th stage, the differentiated cells no longer divide and are considered mature. The mature blood-cells remain in circulation for an extended period of time before they undergo apoptosis. Denoting this rate by a , the differential equation for the mature cells is:

$$\dot{M} = 2d_k P_k - a M \quad (10.4)$$

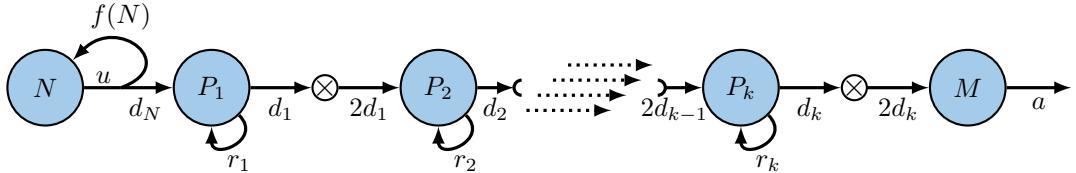


Figure 10.1: **Diagram of the relation between the HSC, mature cell and the intermediate progenitor cells.** Each cell-type is shown as a blue circle. Arrows depict rates of flow from one variable to another. The circle with the \times symbolizes a division and hence doubling of cells. The dotted arrows in the middle illustrate an undetermined number of intermediate step of progenitors.

Assumption P1 implies that $r_j < d_j$ holds for all progenitors j and we define $\delta_j = d_j - r_j > 0$. This allows us to write:

$$\dot{P}_1 = ud_0 N - \delta_1 P_1 \quad (10.5)$$

$$\dot{P}_j = 2d_{j-1} P_{j-1} - \delta_j P_j \quad (10.6)$$

Scaling time and variables

We scale time by u , and let ' denote $\frac{d}{d\tau}$ where $\tau = ut$. Hence the system described above is:

$$N' = f(N) - d_0 N \quad (10.7a)$$

$$uP'_1 = d_0 N - \delta_1 P_1 \quad (10.7b)$$

$$uP'_2 = 2d_1 P_1 - \delta_2 P_2 \quad (10.7c)$$

$$\vdots \quad (10.7d)$$

$$uP'_k = 2d_{k-1} P_{k-1} - \delta_k P_k \quad (10.7e)$$

$$uM' = 2d_k P_k - aM \quad (10.7f)$$

Next, we scale the variables. N remains unscaled, the progenitors are scaled such that $p_j = \delta_j P_j$ and the mature cells are scaled such that $m = aM$. The scaled system is then:

$$N' = f(N) - d_0 N \quad (10.8a)$$

$$\frac{u}{\delta_1} p'_1 = d_0 N - p_1 \quad (10.8b)$$

$$\frac{u}{\delta_2} p'_2 = 2\frac{d_1}{\delta_1} p_1 - p_2 \quad (10.8c)$$

$$\vdots \quad (10.8d)$$

$$\frac{u}{\delta_k} p'_k = 2\frac{d_{k-1}}{\delta_{k-1}} p_{k-1} - p_k \quad (10.8e)$$

$$\frac{u}{a} m' = 2\frac{d_k}{\delta_k} p_k - m \quad (10.8f)$$

Slow manifold / Quasi-steady-state approximation

Note that $\frac{u}{\delta_j} = \frac{u}{d_j - r_j} < \frac{u}{d_j}$ for any stage j . Assumption P2 implies that $\frac{u}{d_j} \ll 1$. Thus, for each j , we consider the limit where $\frac{u}{\delta_j} \rightarrow 0$ and quasi-steady-state approximations of equations (10.8b) through (10.8e) (i.e. the expressions for all stages of progenitors) are appropriate.

The system in equations (10.8) can now be approximated as:

$$N' = f(N) - d_0 N \quad (10.9a)$$

$$0 = d_0 N - p_1 \quad (10.9b)$$

$$0 = 2\frac{d_1}{\delta_1} p_1 - p_2 \quad (10.9c)$$

$$\vdots \quad (10.9d)$$

$$0 = 2\frac{d_{k-1}}{\delta_{k-1}} p_{k-1} - p_k \quad (10.9e)$$

$$\frac{u}{a} m' = 2\frac{d_k}{\delta_k} p_k - m \quad (10.9f)$$

or equivalently:

$$N' = f(N) - d_0 N \quad (10.10)$$

$$m' = a \left(\prod_{j=1}^k 2\frac{d_j}{\delta_1} \right) d_0 N - \frac{a}{u} m \quad (10.11)$$

By scaling back time and m back to M we finally have:

$$\dot{N} = uf(N) - ud_0N \quad (10.12a)$$

$$\dot{M} = u\omega N - aM \quad (10.12b)$$

where

$$\omega = 2 \left(\prod_{j=1}^k 2 \frac{d_j}{\delta_1} \right) d_0 \quad (10.13)$$

Hence, when Assumption P1 and Assumption P2 holds, the system of HSC, progenitors and mature cells can be approximated by equations (10.12). Further quasi-steady-state approximation leading to $\dot{M} = 0$ would be appropriate only if $\frac{u}{a} \ll 1$. However, based on our previously estimated parameter values for the release of HSC from the niche (u in table 8.1) and the death-rate of mature blood-cells (d_{x_1} and d_{y_1} in table 5.1), this does not appear to be appropriate.

10.2 The combined model

10.2.1 Formulation of the combined model

The production of progenitor cells in the reduced niche model was described in section 9.2. In particular, the influx to the first stage of progenitors was found to be described by

$$i_{D_j} = \left(2 - 2\rho_j + \frac{2\gamma\alpha_j\rho_j}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) u_j K N_j \quad (10.14)$$

for all j clones. Through the argument described above, it is appropriate to make a quasi-steady-state approximation of dynamics of the intermediate step between HSC and mature cells, and consider just the mature blood-cells that the final stage of progenitors give rise to.

Mature blood cells can influence HSC behaviour. The same feedback as in the Cancitis model is considered: Apoptosis of blood cells lead to cellular debris, D , and cleanup of debris can up-regulate the immune system, modelled by an abstract measure of the inflammatory load, S . An unspecified exogenous stimulus, I , guarantees immune system activity at all time.

As discussed in section 2.1, chronic inflammation is considered a hallmark of cancer (Colotta et al., 2009) and for MPN such inflammation of the bone-marrow microenvironment is believed to result in an egress of HSC from the niches (Hasselbalch and Bjørn, 2015). For this reason, we connect the inflammatory stimulus, S , with the release of HSC from the niches and we replace u_j by $u_j S$ in the reduced niche model and in the expression i_{D_j} for the production of progenitors, equation (10.14).

With this connection of the HSC and inflammation, it is now possible to propose a model which is a combination of the Cancitis model, and the niche model. The combined model has both the detailed description of HSC-mechanisms of the niche model, along with the intricate relationship between the production of mature blood cells and feedback on HSC production as considered in the Cancitis model. The added HSC-mechanisms allows for a more complete description of the effect of the immune system on the HSC, and enables a mechanistic interpretation of the interaction between HSC properties and chronic inflammation.

Definition 10.2.1: The Combined Cancitis-Niche model

The *Combined Cancitis-Niche model* is given by the system of differential equations:

$$\dot{N}_H = u_H S \left(\frac{2\gamma\rho_H (1 - N_H - N_L)}{\alpha_H + 1 - N_H - N_L} - 1 \right) N_H \quad (10.15a)$$

$$\dot{N}_L = u_L S \left(\frac{2\gamma\rho_L (1 - N_H - N_L)}{\alpha_L + 1 - N_H - N_L} - 1 \right) N_L \quad (10.15b)$$

$$\dot{M}_H = \omega_H i_{D_H} S - d_{M_H} M_H \quad (10.15c)$$

$$\dot{M}_L = \omega_L i_{D_L} S - d_{M_L} M_L \quad (10.15d)$$

$$\dot{D} = d_{M_H} M_H + d_{M_L} M_L - e_D D S \quad (10.15e)$$

$$\dot{S} = r_S D - e_S S + I \quad (10.15f)$$

where $i_{D_j} = \left(2 - 2\rho_j + \frac{2\gamma\alpha_j\rho_j}{\alpha_j + 1 - N_H - N_L} \right) u_j K N_j$. All parameters are non-negative. In addition, $\rho_j \leq 1$ and $\gamma \geq 1$.

Due to the symmetry of different clones, it is trivial to extend the combined Cancitis-niche model to consider more than two clones. For generality, we define the n -clone combined Cancitis-niche model.

Definition 10.2.2: The n -clone Combined Cancitis-Niche model

The n -clone *Combined Cancitis-Niche model* is given by the $2n + 2$ -dimensional system of differential equations:

$$\dot{N}_j = u_j S \left(\frac{2\gamma\rho_j (1 - \sum_{i=1}^n N_i)}{\alpha_j + 1 - \sum_{i=1}^n N_i} - 1 \right) N_j \quad (10.16a)$$

$$\dot{M}_j = \omega_j i_{D_j} S - d_{M_j} M_j \quad (10.16b)$$

$$\dot{D} = \sum_{i=1}^n d_{M_i} M_i - e_D D S \quad (10.16c)$$

$$\dot{S} = r_S D - e_S S + I \quad (10.16d)$$

where $i_{D_j} = \left(2 - 2\rho_j + \frac{2\gamma\alpha_j\rho_j}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) u_j K N_j$. All parameters are non-negative. In addition, $\rho_j \leq 1$ and $\gamma \geq 1$.

Note that extending the original Cancitis model to consider multiple clones is possible, and different ways to do this was considered in the work of Bangsgaard et al. (nd). The n -clone Combined Cancitis-Niche model is thus comparable to a hypothetical n -clone Cancitis model.

Although some results generalize to the n -dimensional combined Cancitis-Niche model, we restrict our focus to the two-clone version in definition 10.2.1. Hence “the combined model” will refer to equations (10.15) for the remainder of this thesis. A schematic diagram of the model for two clones is shown in figure 10.2.

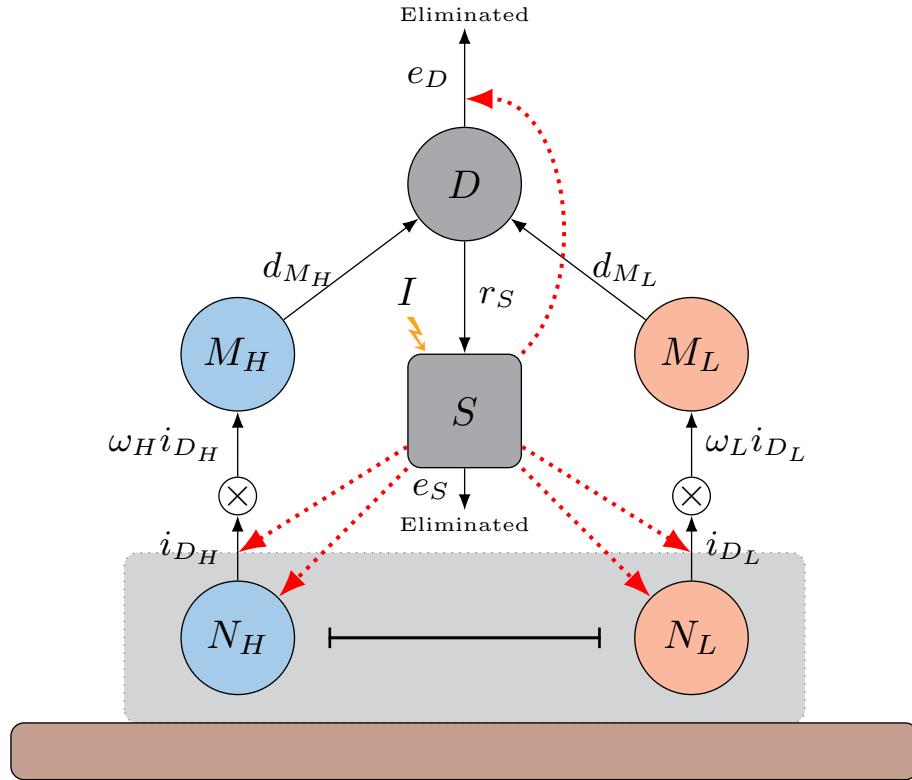


Figure 10.2: **Compartment diagram for the combined Cancitis-Niche model.** Healthy HSC and mature cells, N_H and M_H , are shown in the left-hand side as blue circles, while leukemic cells and leukemic mature cells, N_L and M_L , are shown in the right-hand side as red circles. The dead cells, D and the immune response, S is shown in the middle as a grey circle and a grey box respectively. The black arrows represents flows between compartments, while the red arrows signify rates that are upregulated by the immune response. The box with the \times symbolizes a multiplication due to progenitor proliferation. The reproduction of healthy and leukemic stem cells is not shown, however the interaction and self-renewal through the niche is illustrated by the grey background and the dark-red box at the bottom of the figure.

10.2.2 Brief comments on the formulation of the combined model

In the formulation of the combined model, we assumed that immune response, S , affects the detachment of HSC from the bone-marrow niches, u_j , in a linearly increasing way. Naturally, it is possible that multiple HSC properties are directly affected by the immune system, and that the contribution of immune cells in various biological pathways should be included as well. S as a linear factor on u_j is thus a simplification. An alternative contribution of the immune response could be directly on the HSC niches. Increased inflammation of the immune system is known to affect HSC and the bone-marrow (Hasselbalch and Bjørn, 2015), which could affect the number and properties of available HSC niches. Hence the number of free niches K could be perturbed by S leading to changes for the HSC and the production of progenitors. However, the details of how changes of K affects the model dynamics should first be considered separately in the niche model, to ensure that the model reduction is analogous. Lacking a particular mode of effect from the immune response on the HSC niches, we omit any possible effect of S on K .

Two features of the Cancitis model are not included in the formulation of the combined model: The mutation-rate r_m and the T-cell induced death-rate of leukemic stem cells \tilde{d}_{y_0} . While including a flow of healthy stem cells N_H to leukemic stem cells N_L through a mutation rate would be trivial in the niche model, we omit this contribution in the combined model, since the specifics of the rate of mutation is largely unknown. In previous investigations of the Cancitis model, as well as in section 6, we assumed $r_m = 0$ for simplicity, with the interpretation that the rate of mutation is so rare that within a simulated time-period, no mutations occur. The T-cell induced death-rate \tilde{d}_{y_0} were introduced in (Ottesen et al., 2019), to include the contribution from the immune system assumed to be most significant in regards to clearing of malignant cells. Analysis showed that changes to \tilde{d}_{y_0} changed the stability structure of the system, and hence suggested that T-cell activation could be beneficial to the patient. In the niche model, HSC death-rates were assumed to be insignificant and loss of cells were modelled to occur only through differentiation. However, an additional rate of loss through HSC apoptosis (either naturally occurring or T-cell induced) could be included in the rates of differentiation, hence modelling the combined loss of cells. Changing the definition of differentiation in the niche model such that $d_{A_j} = \Delta_{A_j} + \delta_{A_j}$ where Δ_{A_j} is the rate of differentiation and δ_{A_j} is the rate of apoptosis would lead to an equivalent model. However, the definition of the production of progenitors would change to only consider the rates of differentiation Δ_{A_j} and Δ_{I_j} . Thus, including HSC death in the niche model would not change the dynamics, but would reduce the production of progenitors. For simplicity, we assume that it is possible to capture such reduction of progenitor-production through changes to the multiplication factors ω_H and ω_L . Put differently, our assumption is that increased HSC apoptosis primarily affects the production of mature cells, and that it does so linearly. Increased apoptosis of active HSC A_j is modelled as an increase of d_{A_j} which in turn decrease ρ_j and reduction of ω_j . Simiarly for I_j , apoptosis is assumed to be captured through a decrease of ω_j , as well as an increase of d_{I_j} , increasing the reduced parameter α_j .

In the Cancitis model, and in turn also in the combined model, there is no distinction between different cell-lines such as thrombocytes or leukocytes. This has the consequence that different cell-lines behave identically, and hence mutations, immune-system feedback or treatment that affect cell-lines differently cannot be considered, and different cell-lines must be considered as different scaled versions of the same count of mature cells for the given clone.

We present a simplified way to extend the model to consider separate cell-lines. By splitting the mature cell counts, M_H of equations (10.15c), separate expressions for particular cell-lines could be described, by considering different multiplication factors such as e.g. $\omega_{H,1} = \eta_H \omega_H$ and $\omega_{H,2} = (1 - \eta_H) \omega_H$ with $\eta_H \in [0, 1]$. This could model e.g. thrombocytes and leukocytes

separately:

$$\dot{M}_{H,trom} = \omega_1 i_{D_H} S - d_{M_{H,trom}} M_{H,trom} \quad (10.17a)$$

$$\dot{M}_{H,leuk} = \omega_2 i_{D_H} S - d_{M_{H,leuk}} M_{H,leuk} \quad (10.17b)$$

with both contributing to the debris D and similar expression for the malignant clone. Such splitting of cell-lines allows for different modelling opportunities, like modelling diseases that increase just thrombocyte production (by increasing $\eta_L > \eta_H$) as well as treatment that directly affects one cell-line but not another (through e.g. an increased cell-line-specific death-rate). Additionally, one could consider immune-system feedbacks that affect each cell-line in a unique way. Such separate modelling of different cell-lines is not uncommon in the literature, and is considered in, for example the Mackey model (Colijn and Mackey, 2005b). We direct the reader to the ongoing research of the Cancitis group, in particular the Ph.D.-thesis of Zamra Sajid, in which such extension is considered for the Cancitis model (in preparation at the time of writing). Since different signalling is known to regulate the progenitors of particular cell-lines, this concept is biologically reasonable. However, considering multiple cell-lines would increase the number of parameters and make identification of parameter from data more difficult.

The mathematical structure of the combined model is similar to the reduced niche model. The mathematical properties of the model are discussed below.

10.2.3 Feasibility

As for the niche model, the notion of feasibility of solutions and steady states (definition 8.1.1), has similar importance for the combined model. Only solutions that remain feasible (i.e. non-negative) for all time can be considered biologically reasonable. The definition of feasibility for the combined model is trivially that all variables are non-negative.

We show feasibility of the n -clone version. Considering the derivatives of the system in the cases where the particular variable is zero, we find:

$$\dot{N}_j|_{N_j=0} = 0 \quad (10.18)$$

$$\dot{M}_j|_{M_j=0} = \omega_j i_{D_j} S \geq 0 \quad (10.19)$$

$$\dot{D}|_{D=0} = \sum_{i=1}^n d_{M_i} M_i \geq 0 \quad (10.20)$$

$$\dot{S}|_{S=0} = r_S D + I \geq I > 0 \quad (10.21)$$

In addition, considering the sum of HSC approaching zero, $\sum_{i=1}^n N_i \rightarrow 0$, we find:

$$\frac{\dot{N}_j}{N_j} \xrightarrow{\sum_{i=1}^n N_i \rightarrow 0} u_j S \frac{1}{\alpha_j + 1} (2\gamma\rho_j - 1 - \alpha_j) \quad (10.22)$$

and hence when $2\gamma\rho_j - 1 > \alpha_j$, $\frac{\dot{N}_j}{N_j}$ is positive for a vanishing sum of HSC. As will be shown below, this is exactly the requirement for a feasible steady state for clone j , as was the case for the niche model.

10.2.4 Steady states

The steady state of the combined model resemble the steady states of the niche model. Only feasible steady states are considered. Steady state values are denoted with an asterisk *.

We first consider a scenario with no malignant clone, i.e. $N_L = M_L = 0$. As all parameters are positive, $\dot{N}_H = 0$ implies either $S^* = 0$, $N_H^* = 0$ or $\frac{2\gamma\rho_H(1-N_H^*)}{\alpha_H+1-N_H^*} = 1$. For $S^* = 0$, $\dot{S} = 0$ implies that $D^* = -\frac{I}{r_S} < 0$. Such steady state would not be feasible, and we omit further investigations. For $N_H^* = 0$, the steady state production of progenitors, $i_{D_H}^* = 0$ which in turn implies $M_H^* = 0$. As a consequence $\dot{D} = 0 \Rightarrow D^* = 0$ and $\dot{S} = 0 \Rightarrow S^* = \frac{I}{e_S}$. We refer to the steady state identified as the *trivial steady state*, E_0^* , which can be written in the order $(N_H^*, M_H^*, N_L^*, M_L^*, D^*, S^*)^T$ as:

$$E_0^* = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{I}{e_S} \end{pmatrix} \quad (10.23)$$

The case $\frac{2\gamma\rho_H(1-N_H^*)}{\alpha_H+1-N_H^*} = 1$, implies $N_H^* = 1 - \frac{\alpha_H}{2\gamma\rho_H-1}$. Note that this is exactly the same as for the reduced niche model. In particular $N_H^* = 1 - f_H^{-1}$ where f_H is the scaled fitness of the reduced model from in definition 9.1.3. The production of progenitors can be simplified for $1 - N_H^* = f_H^{-1}$ as

$$i_{D_H}^* = (1 + 2\rho_H(\gamma - 1)) u_H K (1 - f_H^{-1}) \quad (10.24)$$

where $f_H = \frac{2\gamma\rho_H - 1}{\alpha_H}$.

Trivially $M_H^* = \frac{\omega_H}{d_{M_H}} i_{D_H}^* S^*$ arises from $\dot{M}_H = 0$. Writing up $\dot{D} = 0$ with $M_L = 0$, and assuming $S^* > 0$, we find:

$$D^* = \frac{\omega_H}{e_D} i_{D_H}^* \quad (10.25)$$

which allows us to determine S^* by setting $\dot{S} = 0$ and obtaining:

$$S^* = \frac{r_S}{e_S} D^* + \frac{I}{e_S} \quad (10.26)$$

Hence, we have determined a *healthy steady state*, E_H^* :

$$E_H^* = \begin{pmatrix} 1 - f_H^{-1} \\ \frac{\omega_H}{d_{M_H}} i_{D_H}^* \left(\frac{r_S \omega_H}{e_S e_D} i_{D_H}^* + \frac{I}{e_S} \right) \\ 0 \\ 0 \\ \frac{\omega_H}{e_D} i_{D_H}^* \\ \frac{r_S \omega_H}{e_S e_D} i_{D_H}^* + \frac{I}{e_S} \end{pmatrix} \quad (10.27)$$

where $f_H = \frac{2\gamma\rho_H - 1}{\alpha_H}$ and $i_{D_H}^*$ is given in equation (10.24). The steady state solution is feasible

only when $0 < f_H^{-1} < 1$. Analogously, we find a *leukemic steady state*, E_L^* :

$$E_L^* = \begin{pmatrix} 0 \\ 0 \\ 1 - f_L^{-1} \\ \frac{\omega_L}{d_{M_L}} i_{D_L}^* \left(\frac{r_S \omega_L}{e_S e_D} i_{D_L}^* + \frac{I}{e_S} \right) \\ \frac{\omega_L}{e_D} i_{D_L}^* \\ \frac{r_S \omega_L}{e_S e_D} i_{D_L}^* + \frac{I}{e_S} \end{pmatrix} \quad (10.28)$$

where $f_L = \frac{2\gamma\rho_L - 1}{\alpha_L}$ and $i_{D_L}^*$ is given by

$$i_{D_L}^* = (1 + 2\rho_L(\gamma - 1)) u_L K (1 - f_L^{-1}). \quad (10.29)$$

The leukemic steady state is feasible when $0 < f_L^{-1} < 1$.

Co-existence

As in the niche model, *co-existence* steady states are possible, when $f_H = f_L$ i.e. $\frac{2\gamma\rho_H - 1}{\alpha_H} = \frac{2\gamma\rho_L - 1}{\alpha_L}$.

By defining $f = f_H = f_L$ and observing that $N_H^* + N_L^* = 1 - f^{-1}$ implies both $\dot{N}_H = 0$ and $\dot{N}_L = 0$. Hence, as for the niche model, a line of co-existence steady states exists, determined by $N_H^* + N_L^* = 1 - f^{-1}$. Defining $\beta \in [0, 1 - f^{-1}]$ allows us to write $N_L^* = \beta$ and $N_H^* = 1 - f^{-1} - \beta$.

As in equation (10.24) and (10.29), the production of progenitors can be simplified in co-existence steady states:

$$i_{D_H, \beta}^* = (1 + 2\rho_H(\gamma - 1)) u_H K N_H^* = (1 + 2\rho_H(\gamma - 1)) u_H K (1 - f^{-1} - \beta) \quad (10.30a)$$

$$i_{D_L, \beta}^* = (1 + 2\rho_L(\gamma - 1)) u_L K N_L^* = (1 + 2\rho_L(\gamma - 1)) u_L K \beta \quad (10.30b)$$

The steady state values of M_H^* , M_L^* , D^* and S^* can be determined in the same way as in the single-clone steady state:

$$M_H^* = \frac{\omega_H}{d_{M_H}} i_{D_H, \beta}^* S^* \quad (10.31)$$

$$M_L^* = \frac{\omega_L}{d_{M_L}} i_{D_L, \beta}^* S^* \quad (10.32)$$

For D^* , we assume $S^* > 0$ as above and the expression $\dot{D} = 0$ simplifies to:

$$D^* = \frac{\omega_H}{e_D} i_{D_H, \beta}^* + \frac{\omega_L}{e_D} i_{D_L, \beta}^* S^* \quad (10.33)$$

$\dot{S} = 0$ yields $S^* = \frac{r_S}{e_S} D^* + \frac{I}{e_S}$ as above.

The line of co-existence steady states, parametrized by β can then be written as:

$$E_{\beta}^* = \begin{pmatrix} 1 - f^{-1} - \beta \\ \frac{\omega_H}{d_{M_H}} i_{D_H, \beta}^* \left(\frac{r_S}{e_S} D^* + \frac{I}{e_S} \right) \\ \frac{\beta}{\frac{\omega_L}{d_{M_L}} i_{D_L, \beta}^* \left(\frac{r_S}{e_S} D^* + \frac{I}{e_S} \right)} \\ D^* \\ \frac{r_S}{e_S} D^* + \frac{I}{e_S} \end{pmatrix} \quad (10.34)$$

where D^* is given in equation (10.33), $i_{D_H, \beta}^*$ and $i_{D_L, \beta}^*$ are given in equations (10.30), and $\beta \in [0, 1 - f^{-1}]$. The line of co-existence steady states is feasible only when $0 < f^{-1} < 1$.

Note that for $\beta = 0$, $E_{\beta}^* = E_H^*$ and for $\beta = 1 - f^{-1}$, $E_{\beta}^* = E_L^*$. Hence, the line of co-existence steady states connects the two single-clone steady states.

Local stability

The local stability of the steady states can be determined from the eigenvalues of the Jacobian evaluated at the steady states. Analysing the 6×6 Jacobian is primarily done numerically in the following sections. However, for the trivial steady state E_0^* , the eigenvalues of the Jacobian is a simple expression, which alludes to the stability of the niche model. The eigenvalues of the Jacobian evaluated at E_0^* are:

$$\lambda_{Jac, E_0^*} = \left\{ -e_S, -\frac{e_D}{e_S} I, -d_{M_H}, -d_{M_L}, \frac{u_H}{1 + \alpha_H^{-1}} \frac{I}{e_S} (f_H - 1), \frac{u_L}{1 + \alpha_L^{-1}} \frac{I}{e_S} (f_L - 1) \right\} \quad (10.35)$$

From equations (10.27) and (10.28) $f_H > 1$ and $f_L > 1$ are necessary for feasibility of E_H^* and E_L^* respectively. Hence two of the eigenvalues λ_{Jac, E_0^*} are positive, and E_0^* is locally unstable. Only in the case where neither of the steady states E_H^* and E_L^* are feasible are all the eigenvalues λ_{Jac, E_0^*} negative. As such, when there are no other feasible steady states, E_0^* is locally stable.

Numerical investigations suggest that local stability depends on f_H and f_L in the same way as in the niche model: When $f_H > f_L$, the eigenvalues of the Jacobian evaluated in the healthy steady state all have negative real parts, while both the trivial and the leukemic steady state has at least one eigenvalue with positive real part. For $f_H < f_L$, only the leukemic steady state has negative real parts of all eigenvalues and the trivial and healthy steady states both have at least one eigenvalue with positive real part. In the co-existence scenario, evaluating the Jacobian along the line of co-existence, one eigenvalue is zero.

Summary of steady states

The steady states of the combined model are analogous to the steady states of the niche model, that is, a trivial steady state, E_0^* , always exists. Under the assumption $0 < f_H^{-1} < 1$ (or equivalently $\alpha < 2\gamma\rho_H - 1$, since $0 < \alpha$) a healthy steady state, E_H^* , exists. For $0 < f_L^{-1} < 1$ a leukemic steady state, E_L^* , exists. In the special case where $f_L = f_H$, a line of co-existence steady states exists, connecting E_H^* and E_L^* .

Investigating the eigenvalues of the Jacobian numerically revealed that the fitnesses of the clones determine the local stability. When $f_H < f_L$, only the leukemic steady state is locally

Parameter	Value	Origin	Parameter	Value	Origin
u_H	0.0376	Niche model	u_L	0.0432	See text
ρ_H	0.5289	Niche model	ρ_L	0.5310	See text
α_H	0.0053	Niche model	α_L	0.0051	See text
K	15000	Niche model	γ	1	Niche model
ω_H	$4.7 \cdot 10^6$	Cancitis A_x	ω_L	$11.75 \cdot 10^6$	See text
d_{M_H}	0.0129	Cancitis d_{x_1}	d_{M_L}	0.0129	Cancitis d_{y_1}
e_D	$2 \cdot 10^5$	Cancitis e_a	r_S	0.0003	Cancitis r_s
e_S	2	Cancitis e_s	I	7	Cancitis I

Table 10.1: **Default parameters used in simulations of the combined model.** Parameters u_H , U_L , d_{M_H} , d_{M_L} , e_D , r_S and e_S are in units of [$days^{-1}$], while the other parameters are without unit. For parameter where the values come from the Cancitis model, the name of the original parameter is included. Parameters u_L , ρ_L , α_L and ω_L were determined from the healthy counterparts, modified to agree with the disease progression shown in figure 10.3, see main text for details.

stable. When $f_H > f_L$, only the healthy steady state is locally stable. From numerical investigations of the fitnesses, it is possible to determine which steady state is stable for a given set of parameters. In the following sections, this is done in treatment simulations to illustrate whether the flow for a given set of parameters approach E_0^* , E_H^* or E_L^* (or whether co-existence is possible).

10.2.5 Parametrization and numerical investigations

The default parameters of the combined model come from the parameters of the Cancitis model and the HSC niche model, since all terms in the combined model are analogous to terms in the original models. In table 10.1 the default parameters are shown, along with an explanation of where the values came from. The advantage of the leukemic clone was determined by modifying the parameter values by hand to obtain a visual agreement with both the exponential growth described in section 3 and the default parameters of the Cancitis model when simulating disease progression starting with a single malignant stem cell. The values of ρ_L , α_L and u_L were all modified. In figure 10.3 a comparison is shown. To have a larger count of mature cells in the leukemic steady state, we set $\omega_L = 2.5\omega_H$. The analogous parameter in the Cancitis model, A_x and A_y were equal, however, in the Cancitis model the leukemic steady state was characterized by a large count of leukemic stem cells, resulting in large numbers of mature cells. Since stem cells counts in the leukemic steady state in the niche model are only slightly more numerous than in the healthy steady state, the increased counts of malignant mature cells must arise due to an excessive proliferation of malignant progenitor cells. While the logistic approximation of the niche model was related to MPN in section 9.4, we determined that unique values of ρ_L , α_L and u_L cannot be estimated from the data considered. Hence the default values for these are particularly uncertain, but provide possible reasonable first estimates of values.

Figure 10.4 shows all variables of an exemplary simulation in which a single mutation at time

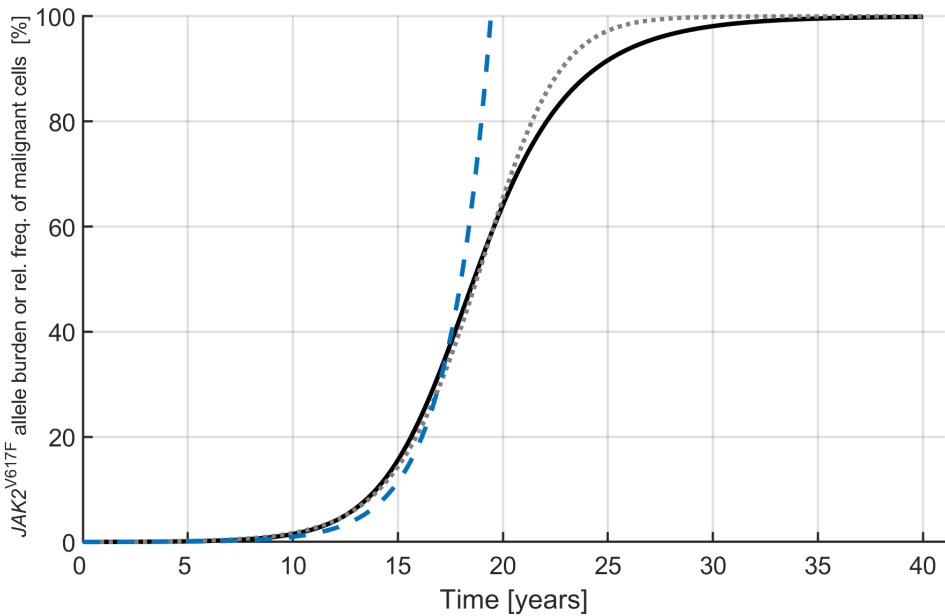


Figure 10.3: **Disease progression in the emperical model, the Cancitis model and the combined model.** The exponential growth of the JAK2 allele burden determined in section 3 is shown in dashed blue. The relative frequency of malignant mature cells in the Cancitis model is shown in black, while the grey dotted line depict the relative frequency of malignant mature cells in the combined model, using the default parameters given in table 10.1.

$t = 0$ is simulated as K^{-1} cells subtracted from N_H in the healthy steady state and added to N_L . The default parameters imply that the clones have fitness $f_H = 11.0$ and $f_L = 12.1$, and hence the leukemic steady state is locally stable since $f_H < f_L$.

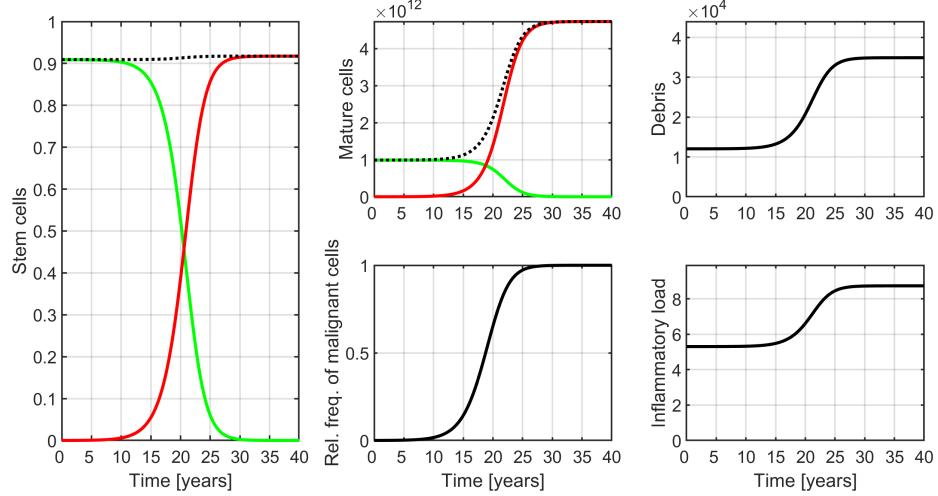


Figure 10.4: **Illustration of a model simulation of the combined model.**

At time $t = 0$, the system is perturbed out of the healthy steady state, by removing $K^{-1} \approx 7 \cdot 10^{-5}$ from N_H and adding the same amount to N_L . In the left panel and the top mid panel green curves depict the healthy clone N_H and M_H respectively, while the red curves are the leukemic clone N_L and M_L . The dotted black curves shows the sum. The middle panel at the bottom shows the relative frequency $\frac{M_L}{M_H + M_L}$. In the top-right, the debris of dead cells D is depicted, while the bottom-right depicts the inflammatory load, S . The default parameters of table 10.1 were used.

As the parameters of the combined model have biological interpretations, it is possible to simulate treatment that affects particular biological processes. From the steady state considerations in the previous section, disease progression toward the leukemic steady state occurs due to the leukemic clone having a higher fitness than the healthy clone, $f_L > f_H$. Hence it is clear that successful treatment must perturb parameters such that the healthy clone has a higher fitness than the leukemic clone. This can be achieved by either increasing f_H , decreasing f_L or a combination thereof. Figure 10.5 depicts a simulation such a treatment scenario, in which ρ_L is significantly decreased in a two-year period. The parameter-perturbation changes f_L to 5.86 and hence the healthy steady state becomes locally stable because $f_H > f_L$. This results in a decline of leukemic cells (red curve) and an increase of healthy cells (green curve). The relative frequency of malignant mature cells out of all mature cells also decreases significantly as a consequence.

The relative frequency of the malignant clone in the simulation shown in figure 10.5 agrees well with the decay of the JAK2 allele burden observed in IFN-treated patients from the DALIAH trial. An initial delay before decreasing is in agreement with the slow respond observed in some patients, and could explain the pseudo-progression response-type discussed in the end of section 3. However, the DALIAH trial data showed that during IFN treatment, blood-cells counts decreased on a time-scale much faster than the decrease observed in the JAK2 allele burden. Hence, figure

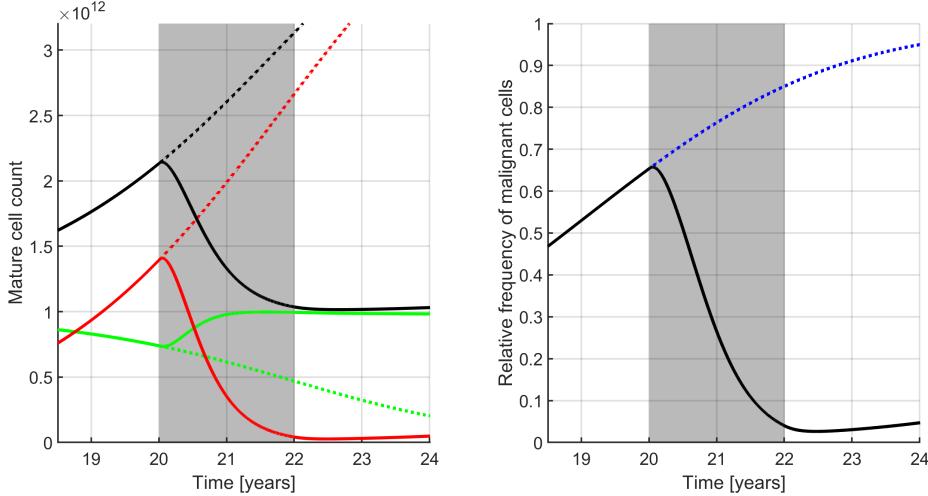


Figure 10.5: Illustrative example of ρ_L treatment in the combined model. Starting in the healthy steady state S_0^* at time $t = 0$ (with $M_H \approx 1 \cdot 10^{12}$), the equivalent of one HSC, $K^{-1} \approx 7 \cdot 10^{-5}$, is removed from N_H and added to N_L . Due to the growth advantage of the malignant clone, the disease develops, the relative frequency of malignant mature cells, $\frac{M_L}{M_H+M_L}$ reaches approximately 65% within 20 years. Between year 20 and 22 treatment is simulated, setting $\rho_L = 0.515$. This reduces the fitness of the malignant clone, f_L , from 12.1 to 5.9, resulting in the healthy clone having the higher fitness during treatment since $f_H = 11$. At year 22, the ρ_L parameter is reset to the default value. The grey background shows the period of treatment. The left panel depicts the mature cell-counts, with full lines depicting the simulated scenario and the dotted lines showing the scenario without treatment. M_H is shown in green, M_L is magenta and the sum $M_H + M_L$ is shown in black. The right-hand panel shows the relative frequency of malignant mature cells, with the dotted blue curve showing the scenario without treatment.

10.5 suggests that while a decreased value of ρ_L can explain the JAK2 decrease, it does not lead to a fast decrease of blood cell counts and additional parameters must be perturbed to model the full effect of IFN. In figure 10.6 such a scenario is depicted. In addition to a significant reduction of ρ_L during treatment, we also model a decrease of ω_H and ω_L . The two parameters are reduced by the same magnitude. This results in treatment which features an immediate and significant decrease in mature blood cells while maintaining a slower decay in the relative frequency of malignant mature cells. Hence both the relative difference in blood cell counts (given by the JAK2 allele burden) and the absolute blood cell counts can be captured by the combined model when ω_H and ω_L is perturbed in addition to ρ_L . Note that ω_H and ω_L does not influence the fitness of either clone.

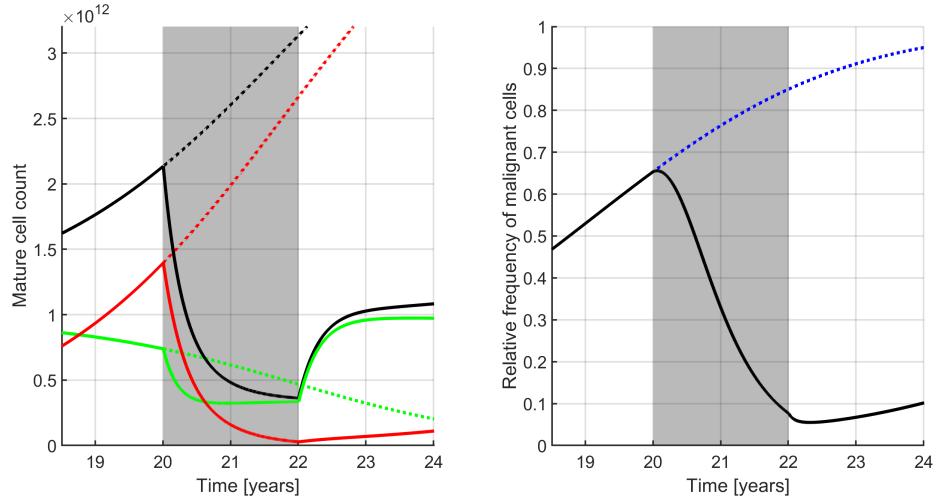


Figure 10.6: Illustrative example of treatment affecting ρ_L , ω_H and ω_L in the combined model. Starting in the healthy steady state S_0^* at time $t = 0$ (with $M_H \approx 1 \cdot 10^{12}$), the equivalent of one HSC, $K^{-1} \approx 7 \cdot 10^{-5}$, is removed from N_H and added to N_L . Due to the growth advantage of the leukemic clone, the disease develops, the relative frequency of malignant mature cells, $\frac{M_L}{M_H + M_L}$ reaches approximately 65% within 20 years. Between year 20 and 22 treatment is simulated, setting $\rho_L = 0.515$, $\omega_H = 2 \cdot 10^6$ and $\omega_L = 5 \cdot 10^6$ (and hence maintaining the relationship $\omega_L = 2.5\omega_H$). At year 22 parameter are reset to the default parameters. The left panel depicts the mature cell-counts, with full lines depicting the simulated scenario and the dotted lines showing the scenario without treatment. M_H is shown in green, M_L is magenta and the sum $M_H + M_L$ is shown in black. The right-hand panel shows the relative frequency of malignant mature cells, with the dotted blue curve showing the scenario without treatment.

During simulated treatment, the healthy steady state is only approached asymptotically and hence never reached. For this reason, treatment cessation will always be followed by a relapse, since some non-zero (possibly very small) number of N_L is present and is competitively advantageous without treatment. Increased thrombocyte counts are one of the primary risk-factors of MPNs, as it is associated with an increase risk of thrombosis. Investigating the time it takes from treatment cessation until dangerous high counts of thrombocyte arise again is therefore

an important measure for the effect of treatment. By defining enhanced thrombocyte counts as mature cells being 50% increased compared to the healthy steady state, we simulated the model following treatment cessation. Figure 10.7 illustrates this definition of relapse, using the treatment shown in figure 10.6 and continuing simulation after treatment cessation.

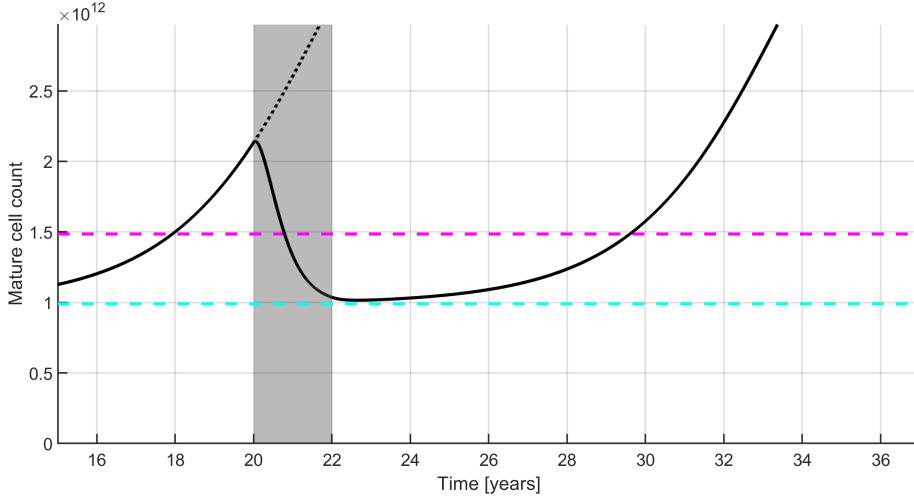


Figure 10.7: Illustration of relapse after treatment cessation. Simulating the same scenario as in figure 10.6, the figure shows the increase in mature cell counts following treatment cessation. The dashed cyan line shows the value of N_H^* in the healthy steady state, while the dashed magenta line shows our definition of a relapse threshold, 50% higher than healthy cell-counts. Hence, even with almost normalized cell-counts due to treatment, the model predicts that the patient will relapse approximately 8 years after treatment cessation.

Considering a range of perturbations of ρ_L , ω_H and ω_L , the relationship between specific treatment-related changes to biology and the time to relapse can be investigated. A contour-plot is shown in figure 10.8. Additional investigation not shown revealed that changing the time of treatment initiation and/or cessation made no significant difference. The specific time to relapse depends on all model-parameters and is thus uncertain. However the general relation between how the treatment affects the patient and the expected time to relapse is hypothesized to be robust in the sense that it is indicative for real predictions. Hence, determining the patient respond to treatment could be used to make long-term predictions and ultimately to determine the best time for diagnosing for relapse of the leukemic clone.

Although a reduction of ω_H and ω_L leads to shorter time to relapse as seen in figure 10.8, we note that the immediate decrease in cell-counts associated with the reduction of ω_H and ω_L is not only observed in data, but also an important clinical goal of treatment, as decreasing thrombocyte-counts immediately is important to reduce the risk of thrombosis for the patient. Hence, even if the reduction of ω_H and ω_L can worsen long-term patient outcome, it is beneficial for the patient in the short-term. In figure 10.9, the short-term benefits of reducing both ρ_L , ω_H and ω_L is illustrated. Considering two independent clinical goals, a reduction of blood-cell counts and a reduction of the relative frequency of the malignant clone, the figure shows how fast these goal are reached. Notably, the reduction of the relative frequency of the malignant clone is shown in panel 10.9a is similar to the time to relapse shown in figure 10.8, while the

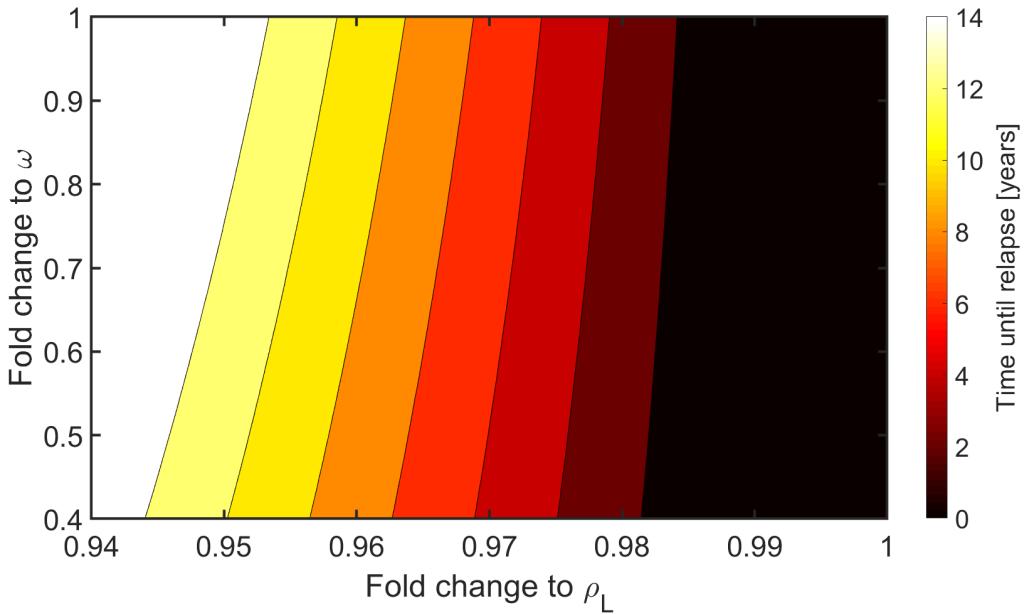


Figure 10.8: **The relationship between modelled treatment respond and time to relapse.** The model was simulated as in figure 10.6 and 10.7, and the time to relapse was found, defined as mature cell count 50% above the healthy steady state. Fold changes to ρ_L was considered as well as fold changes to both ω_H and ω_L . Both ω_H and ω_L were perturbed equally, such that the relation $\omega_L = 2.5\omega_H$ was maintained. From the contours, we see that decreasing ρ_L is associated with a longer time to relapse, while reducing ω_H and ω_L can shorten the time to relapse, although not as significantly. The perturbation illustrated in figures 10.5 and 10.7 corresponds to a 0.97 fold decrease of ρ_L , while the additional perturbation of ω_H and ω_L shown in figure 10.6 corresponds to 0.43 fold decrease of both.

reduction of blood-cells in panel 10.9b is different. Since both are clinical goals, it is relevant to consider the maximal time it takes before both are attained. This is shown in panel 10.9c. The specifics of the simulations in figure 10.9 depends on many elements including the model parameters, the definition of cell-count normalization and on the stage of disease progression at which treatment is initiated. Regardless, the figure illustrates that successful treatment must target both the behaviour of stem cells (through e.g. a reduction of ρ_L) and the production of mature cells (through e.g. a reduction of ω_H and ω_L). Treatment that only targets either would not allow for short-term clinical benefit, even if the time to relapse after treatment cessation is long.

In section 3.1, we noted that the thrombocyte-counts and the JAK2 allele burden did not correlate at baseline. Investigating the parameter-space of the combined model suggests possible explanations for why the thrombocyte counts can be significantly different for equal levels of disease, as assessed by the JAK2 allele burden. By assuming that default parameters, table 10.1, describe an average person well, the mature cell counts in the healthy steady state should be within the healthy interval. As such, we assume that a scaling of thrombocytes is possible such that N_H^* is $300 \cdot 10^3 (\mu L)^{-1}$ at the healthy steady state for default parameters. In figure 10.10 disease progression is simulated as in figure 10.4, for different sets of parameters. By maintaining the same scaling of mature cells, we observe that changes in parameters gives rise to differences in mature cell counts, both before and during disease progression. In particular, figure 10.10a shows that increasing external inflammatory stimulus, I , can increase cell-counts, while figure 10.10b shows that the same is the case for the total number of HSC niches, K . We propose two hypotheses for the lack of correlation between thrombocytes and JAK2 allele burden observed at baseline. We consider a hypothetical patient which gets diagnosed when the thrombocyte-count crosses an arbitrary threshold of $800 \cdot 10^3 (\mu L)^{-1}$. For default parameters, the relative frequency of the malignant clone is 80% when the threshold is crossed. However, if the external inflammatory stimulus, I , was tripled compared to the default value, the relative frequency of the malignant clone was just 20% when the thrombocyte-count exceeds $800 \cdot 10^3 (\mu L)^{-1}$. Similarly, small changes in the count of HSC niches is found to influence this difference significantly, with just a 33% increased niche-count leading to a similar difference in the relative frequency of the malignant cells at diagnosis. These starting conditions are shown as red asterisks * in figure 10.10, with simulated treatment scenarios shown as red lines.

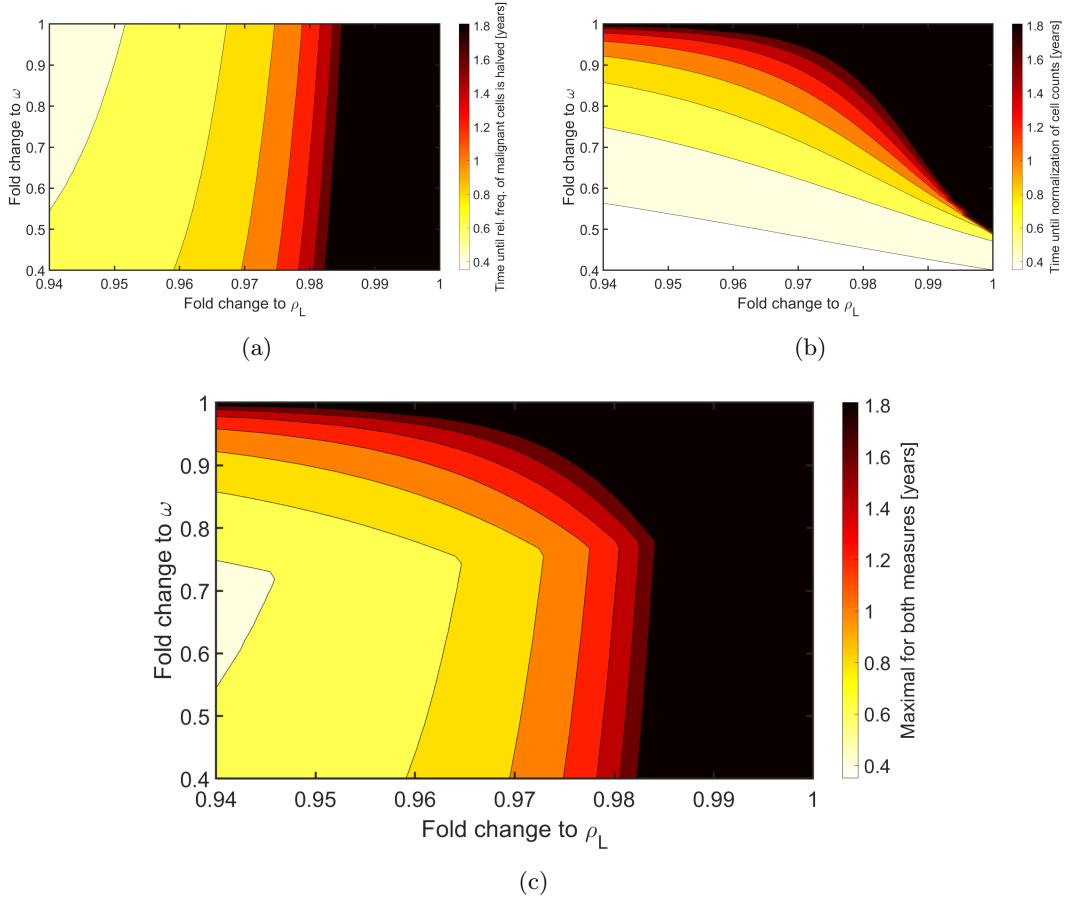


Figure 10.9: The relationship between modelled treatment perturbation and the short-term response. Simulating the same scenario as other figures in this section (figures 10.6 and 10.7), a range of changes to ρ_L , ω_H and ω_L is simulated. As above, $\omega_L = 2.5\omega_H$ is maintained. Panel (a) shows the time from treatment initiation until the relative frequency of the malignant clone is halved compared to the initial level. A significant reduction of ρ_L is found to be necessary for the reduction, whereas the reduction of ω_H and ω_L has little effect. In panel (b) the time to normalization of mature cell counts is illustrated. The time to normalization is here defined as the time from treatment initiation until mature cell counts return to the cell-counts in the healthy steady state. Note that in some cases, the cell-counts are reduced such that they are lower than the healthy level, which can also have negative consequences for the patient. We omit this consideration for simplicity. The worst-case scenario is shown in panel (c), by considering the maximal time it takes before both goals of treatment are attained.

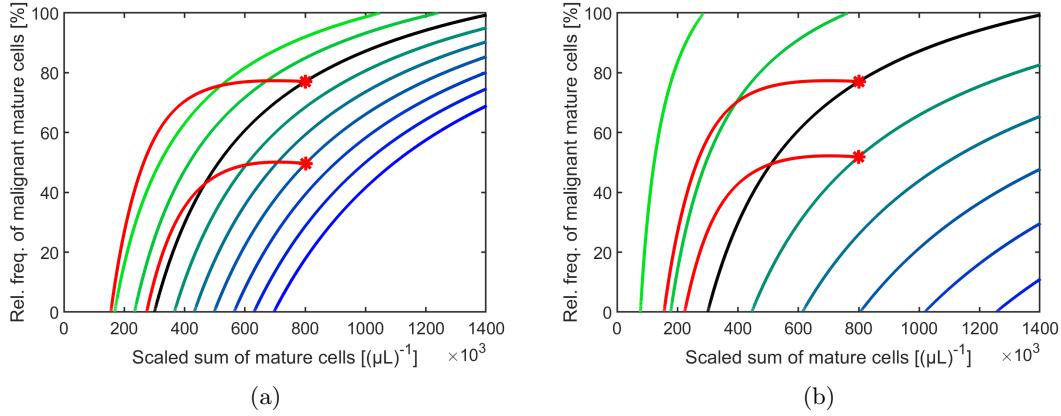


Figure 10.10: Changes in parameters influences the disease progression in the combined model. Disease progression is simulated for different values of I and K , while all other parameters attained the default values. The model was simulated with initial conditions corresponding to one malignant stem cell added in the healthy steady state. The sum of mature cells is scaled such that the healthy steady state for default parameters is $300 \cdot 10^3 (\mu L)^{-1}$, corresponding to a healthy thrombocyte count. In both panels, the black line depicts the scenario where all parameters attain their default value. Disease progression follows the curve from a relative frequency of the malignant clone close to 0 toward a relative frequency of 1. In panel (a) the model was simulated over a range of values for I , from $\frac{1}{3}$ of the default value shown in green on the far left, increasing in intervals of $\frac{1}{3}$ of the default up to a three-fold increase from default, shown in blue on the far right. Similarly, panel (b) depicts simulations where K was perturbed instead of I , starting from $\frac{1}{3}$ the default on the left in green, to a three-fold increase of the default value, shown on the right in blue. The intermediate steps similarly had steps of $\frac{1}{3}$ of the default value between them. In both panels, simulated treatment scenarios are shown for illustration, initiated as the red asterisks, see the main text. In the treatment-scenarios, treatment is simulated as a 0.98-fold change of ρ_L , while both ω_H and ω_L are reduced by a factor 0.6.

10.3 Summarizing discussion

In this section, the Cancitis model of section 5 and the HSC niche model of section 8 were combined to formulate the combined Cancitis-Niche model, definition 10.2.1 and 10.2.2. The combined model includes the features of HSC behaviour from the HSC niche model, as well as blood-cell feedback through an abstract notion of the immune system and inflammation as in the Cancitis model. Hence the combined model describes the full hematopoietic system with possible biological interpretation of the parameters included. We showed that the combined model has steady states that are very similar to those of the HSC niche-model. Single-clone steady states exist, and an expression for HSC fitness determines which steady state is locally stable. The fitness expression is found to be the same as for the reduced HSC niche model. As the combined model is derived from two models for which parameter values were already determined, the values of the parameters of the combined model were based on the values of the corresponding original parameters. Numerical investigations showed that the dynamics of the model is similar to the Cancitis model, see figures 5.2 and 10.4. To investigate the model behaviour, we simulated different treatment scenarios and determined how the clinically relevant measures of total mature cell count and relative frequency of the malignant mature cells responds to hypothetical treatment.

To combine the models, choices about the modelling of progenitor cells were necessary. We formulated a hypothetical scheme of HSC behaviour and blood-cell production with multiple steps of cells of increasing maturity. Intermediate steps were considered progenitor cells, while the final step described the mature blood-cells. Two biological assumptions were made: Limited self-renewal of progenitors and frequent differentiation of progenitors. Under these assumptions, quasi-steady-state approximations of the intermediate steps are appropriate, revealing a model of just HSC and mature cells with progenitors primarily playing a role of multiplication of cell-numbers. We found that a quasi-steady-state approximation of mature cells was not appropriate, based on the comparable time-scales of HSC release from the niche and the death-rate of mature cells. This suggests a connection between the slow dynamics of HSC and the life-span of mature cells. For the HSC niche model, we previously determined that the a slow detachment rate of HSC were related to the assumption that most HSC are niche-bound. Together, these findings suggests that the proportion of HSC that are quiescent and the average life-span of mature cells could be related. This will be discussed further in chapter VI.

The combined model does not include HSC apoptosis, in contrast to the Cancitis model where the parameters d_{x_0} , \hat{d}_{y_0} and \tilde{d}_{y_0} all relate to death-rates of HSC. As perturbation of \hat{d}_{y_0} were in focus in section 6, some comments on HSC apoptosis in the combined model are necessary. The introduction of HSC apoptosis would not affect the HSC niche model or the structure of steady states of the combined model, since the differentiation rates of HSCs could be redefined to include the effective loss of HSC due to both differentiation and apoptosis. However, the definition of the production of progenitors would change, and hence also the expressions i_{D_H} and i_{D_L} of the combined model. For simplicity, we assume that apoptosis occurs rarely enough to not have an effect under normal circumstances. During apoptosis-increasing treatment, we assume that the resulting decrease in progenitor production is either insignificant or can be approximated by a linear decrease in the production of mature cells that the progenitors in turn produce. Since a linear decrease of i_{D_H} or i_{D_L} is equivalent to a decrease of ω_H and ω_L respectively, we assumed that increasing the death-rates of HSC and malignant stem cells can be approximated by changes in ω_H and ω_L . Additionally, if treatment that induces apoptosis of stem cells also induces apoptosis of progenitors, we assume that the effect on the production of mature cells can also be captured by changes in ω_H and ω_L .

As a particular example of the effect of increasing the death-rate of stem cells, we simulated

a scenario where the malignant clone is expanding from a single malignant clone at time $t = 0$ to full-blown malignancy. In accordance with the discussion above, increasing the death-rate of stem cells corresponds to increasing the differentiation rates of the niche model, corresponding to a decrease of ρ_H and ρ_L or an increase of α_H and α_L , as well as an increase of ω_H and ω_L . While not shown here, preliminary numerical investigations suggested that the model dynamics that arise from increasing α_L were similar to the dynamics that arise from decreasing ρ_L . For simplicity, therapeutic intervention was simulated as only a decrease of ρ_L and a decrease of ω_H and ω_L and the remaining parameters were kept fixed. We defined three measures of successful treatment. Firstly, a fast decrease in mature cell counts, secondly, a fast halving of the relative frequency of the malignant clone compared to the baseline value, and thirdly, an increased time from treatment cessation until cell-counts are heightened again. Simulating a wide range of values of the death-rate related parameters mentioned above, we found that while a decrease of ρ_L is important for reducing the relative frequency of the malignant cells and for postponing relapse, a reduction of ω_H and ω_L must occur if cell-counts are to be lowered. Hence, reduction of ρ_L as well as of ω_H and ω_L is necessary to attain the response to treatment seen for IFN-treated patients in the DALIAH trial, see section 3. Patient-specific fits of exactly such responses is the focus of the next section.

The effect of different values of the external inflammatory stimulus I and the number of HSC-niches K was shown in figure 10.10. While this is not a substitute for a thorough sensitivity analysis of the relation between model parameters and different outputs of importance, the numerical investigations were indicative of how changes in parameters can lead to large changes in cell-counts during disease progression. Importantly our findings suggest that different combinations of counts and relative frequencies of cells can be observed at different stages of disease progression. As a consequence, patients with e.g. different external inflammatory stimulus could have experienced very different disease progression up to the point of diagnosis. In particular, a patient with high inflammatory stimulus could experience heightened levels of blood-cells much earlier than a patient with low inflammatory stimulus. Similarly, heightened number of HSC-niches, K , could lead to heightened levels of blood-cells. Conversely, the model suggests that a diminished pool of HSC-niches would lead to reduced blood-cell counts.

Chapter 11

Modelling MPN-patient responses with the combined model

In this section, the combined model is compared to the data from DALIAH trial. The methodology is based on the methodology used for comparing the Cancitis model to the same data-set in section 6, with some modifications. In particular, certain non-fixed model parameters are fitted to data of the blood-cell counts, in addition to a linear scaling of mature cells. The resulting fits are interpreted across the population, and a proof-of-concept of a population-modelling approach is presented.

11.1 Mapping blood-concentration to parameter perturbation

In section 6.1, we discussed how the blood-concentration of IFN was modelled for the study of the DALIAH trial patients for the work described in section 6. Parameters were perturbed relative to the simple pharmacokinetic model of IFN concentration given in equation (6.5). For the combined model, however, the dependence of parameters on the blood concentration of IFN was modified slightly. This was done to easier allow for negative contribution due to treatment, while avoiding the theoretical boundaries of parameter values. In particular, all parameters of the combined model are positive. In addition, the parameters ρ_H and ρ_L must both be less than 1, since e.g. $\rho_H > 1$ would imply $d_{A_H} < 0$.

When considering the DALIAH trial data, the blood concentration of IFN, B , changes with time in accordance with equation (6.4), page 44, and is hence generally a function of time. For simplicity, we write just B instead of $B(t)$ here. The parameters ρ_H and ρ_L are modelled to have the following dependence on B :

$$\hat{\rho}_j(B) = \frac{\rho_j}{\rho_j + (1 - \rho_j)e^{-\nu_{\rho_j}B}} \quad (11.1)$$

where ρ_j is the default parameter value and ν_{ρ_j} is a parameter determining the degree of response to treatment. The expression in equation (11.1) is a logistic expression from 0 to 1, going through ρ_j at $B = 0$, with the slope being determined by ν_{ρ_j} .

For all other parameters, two cases were considered. For $\nu_\theta \geq 0$ a linear increase as in

equation (6.5) and for $\nu_\theta < 0$, an exponential decrease toward zero:

$$\hat{\theta}(B) = \begin{cases} (1 + \nu_\theta B)\theta & \text{for } \nu_\theta \geq 0 \\ e^{\nu_\theta B}\theta & \text{for } \nu_\theta < 0 \end{cases} \quad (11.2)$$

Hence, $\hat{\theta} : (0, \infty) \rightarrow (0, \infty)$, with $\hat{\theta}(B) = \theta$ when $\nu_\theta = 0$. Note that $\hat{\theta}(B)$ is continuous in ν_θ . In practice, this meant that no restrictions on ν_θ were necessary when fitting to patient data. From the data of the DALIAH trial, B takes values between 0 and 20. The linear relation for $\nu_\theta \geq 0$ was chosen as the simplest possible relation between increases in dose and increases in parameter perturbation. We assess that the effect of increases in dose would have been to significant if an exponential function $e^{\nu_\theta B}\theta$ had been used for $\nu_\theta \geq 0$ as well.

Additional details could be considered for the parameter perturbations. This could include a lower threshold for an effect (i.e. an interval of B in which $\hat{\theta} = \theta$), or an upper limit of how significant an effect treatment could have (i.e. a threshold \bar{B} such that $\hat{\theta}(B) = (1 + \nu_\theta \bar{B})\theta$ for all $B \geq \bar{B}$). While such details are biologically realistic, including more details would introduce additional fitting parameters, greatly increasing the complexity of the fitting procedure, and hence obscuring or overcomplicating the interpretation of the response. For this reason, the single-parameter perturbations of equations (11.1) and (11.2) were chosen.

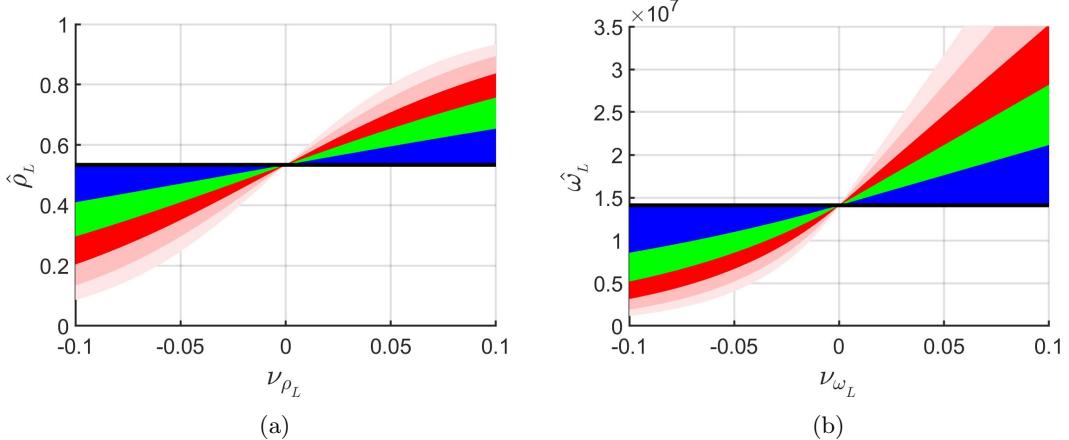


Figure 11.1: **Parameter perturbation in the combined model.** The relation between the fit-parameter ν_θ , the blood-concentration $B(t)$ and the resulting value of the perturbed parameter is shown. For a given value of ν_θ , the top of the blue area shows the value of the perturbed parameter for a blood-concentration of $B = 5$, corresponding to sustained $5 \mu\text{g}$ daily IFN, the top of the green corresponding to $B = 10 \mu\text{g}$ daily IFN, and top of the full red corresponds to $B = 15 \mu\text{g}$ daily IFN. The two light-pink areas show $B = 20$ and $B = 25 \mu\text{g}$ for illustration. Panel (a) shows the parameter perturbation of equation (11.1) used for ρ_C and ρ_L with the perturbation of the latter shown here. All other parameters were perturbed according to equation (11.2), and panel (b) shows the perturbation of ω_L , as an example.

11.2 Methodology for fitting the Combined model to data

The behaviour of the combined model under certain perturbations were discussed in section 10.2.5. In particular, it was shown in figure 10.5 that decreasing ρ_L can lead to a decline in the relative frequency of malignant cells that is visually similar to the decays observed in the JAK2 allele burden of the IFN-treated patients in the DALIAH trial. We found that decreasing both ω_H and ω_L only made little difference in the dynamics of the relative frequency of mature cells, but resulted in a fast decline of the cell-counts. Combining these observations with dose-dependent parameter perturbation and the procedure used for fitting the Cancitis model to data (as described in chapter 6), different possible ways to relate the combined model to data arise. We here describe the methodology that we have used. Some of the considerations from previous sections bears repeating. The relative frequency of malignant mature cells in the model, $\frac{M_L}{M_H+M_L}$, is assumed to agree with the JAK2 allele burden, and a linear scaling is assumed to relate the sum of mature cells in the model ($M_H + M_L$ for the combined model) to patient blood-cell counts (both thrombocyte- and leukocyte-counts). The blood-concentration of IFN is modelled in the same way as described in previous sections. All patients are considered on an individual level, and the found model-response thus reflects a patient-specific response to treatment. Initial investigations suggested that optimizing for both the blood-cell counts and relative frequency of the malignant cells was difficult due to scaling of errors-measures, leading to model-fits that only agreed with either the cell-counts or the relative frequency of malignant mature cells, but rarely both simultaneously. To alleviate this problem, we chose a three-step methodology:

1. ρ_L and the JAK2 allele burden.

Initially, dose-dependent perturbation of just ρ_L was considered. We minimized the squared difference of the JAK2 allele burden measurement of the given patient and the relative frequency of the malignant cells at the corresponding times, by fitting the parameter ν_{ρ_L} , see equation (11.1).

As in section 6.6, we considered a linear scaling of the sum of mature cells, $M_{tot} = M_H + M_L$, in the model. We defined the error of the model prediction and the observed cell count as in equation (6.8):

$$E_{c,i} = (M_H(t_i, \nu_\theta) + M_L(t_i, \nu_\theta)) \mathcal{R}_c - D_{c,i} \quad (11.3)$$

where $D_{c,i}$ were the c cell-line measurements at time t_i and \mathcal{R}_c was the model-predicted fraction of mature cells from the cell line c .

Using the dose-dependent perturbation of ρ_L , patient-specific scaling factors for thrombocytes and leukocytes were determined.

2. Multiplication factors fitted to linearly scaled blood-cell counts.

The multiplication factors ω_H and ω_L were determined to affect the mature cell counts M_H and M_L greatly, without much change to the relative frequency of the cells. For this reason, we fitted dose-dependent perturbations of ω_H and ω_L simultaneously with cell-line specific linear scaling as above. The perturbation of ω_H and ω_L was equal, to reduce the number of fitted parameters. Hence, a single parameter $\nu_\omega = \nu_{\omega_H} = \nu_{\omega_L}$ was used in the fitting procedure. Initial investigations revealed that considering independent fits of ν_{ω_H} and ν_{ω_L} leads to an improved fit for some patients, however, the resulting parameter values correlated in most cases, suggesting that the two parameters cannot be independently identifiable.

3. Final optimization of ρ_L to JAK2 allele burden.

Finally, we fitted the relative frequency of malignant cells to the JAK2 allele burden by maintaining the found value of $\nu_\omega = \nu_{\omega_H} = \nu_{\omega_L}$ and subsequently optimizing ν_{ρ_L} to reduce the JAK2 allele burden error a second time. The changes to ω_H and ω_L had little effect on this measure, however in some cases, the effect could be alleviated by this step. For most patients, the change to ν_{ρ_L} was minor. After optimizing ν_{ρ_L} , the scaling-factors were also fitted again.

To summarize, step 1 consisted of fitting the model to JAK2 data with parameter ν_{ρ_L} , followed by a fitting of the scaling factors \mathcal{R} to thrombocyte- and leukocyte-counts. In step 2 the parameter $\nu_\omega = \nu_{\omega_H} = \nu_{\omega_L}$ were fitted to the thrombocyte- and leukocyte-counts and the scaling factors were updated. The final step consisted of fitting the parameter ν_{ρ_L} again. Hence, two model-parameters were considered, as well as two scaling factors; one for thrombocytes and one for leukocytes. Fitting of parameters were done for 63 patients, using the MATLAB function `fminsearch`. Figures of the resulting model-dynamics after each step are shown in Electronic Supplementary Material 4. In the section below, some examples of particular patients are highlighted and discussed in detail.

11.3 Examples of patient fits

A dose-dependent $\hat{\rho}_L(B)$ was obtained for each patient by fitting ν_{ρ_L} to the JAK2 allele burden. This resulted in a good visual agreement between data and model-fit for most patients. For some patients however, such dose-dependent perturbation of ρ_L does not result in the behaviour seen in data. In figure 11.2 examples are shown of both good and bad fits.

As shown in section 10.2.5, equal perturbation of ω_H and ω_L had little influence on the behaviour of the relative frequency of malignant mature cells out of all mature cells. Hence, for patients where perturbing ρ_L is not sufficient to obtain a satisfactory fit to data, perturbing ω_H and ω_L equally does not change the agreement between the model and the JAK2 allele burden much. However, by maintaining the dose-dependence of $\hat{\rho}_L$ found above, equal dose-dependent perturbation of ω_H and ω_L can greatly improve the agreement between the scaled sum of mature cells in the model and the leukocyte- and thrombocyte counts. This is achieved in the second step of the methodology by finding a value $\nu_\omega = \nu_{\omega_H} = \nu_{\omega_L}$ that minimizes the blood-cell error as defined in equation 11.3. In figure 11.3 examples of the resulting fits are illustrated. Only thrombocytes are shown. For most patients, the dynamics of thrombocytes and leukocytes during IFN treatment were visually similar, however, due to differences in the order of magnitude of data, fitting often lead to better fit for thrombocytes than for leukocytes whenever there were differences in dynamics, see Electronic Supplementary Material 4 for all data and fits. The fits with $\nu_\omega = 0$, shown in blue in figure 11.3, show a slow decrease of thrombocytes, and the initial count at baseline is underestimated. This is the case for almost all patients. Fitting with ν_ω as well as the thrombocyte-specific scaling factor leads to a higher estimate of baseline thrombocyte counts, in good agreement with data, and fast decrease in counts within the first half year. Fitting a dose-dependent perturbation of ω_H and ω_L suggests that changes in dose directly influence the thrombocyte count. In both examples shown in figure 11.3, the IFN-dose is increased approximately eight months after baseline. The effect of the increase as well as the effect of later decreases in dose are seen clearly in the model-fit. Hence the model suggests a that IFN has significant effect of reducing the production of blood cells within the time-scale of a few months of treatment. Importantly, for both patients shown in figure 11.3, our modelling of the effect of IFN leads to good agreement between the data and the model for the initial half-year of fast decrease.

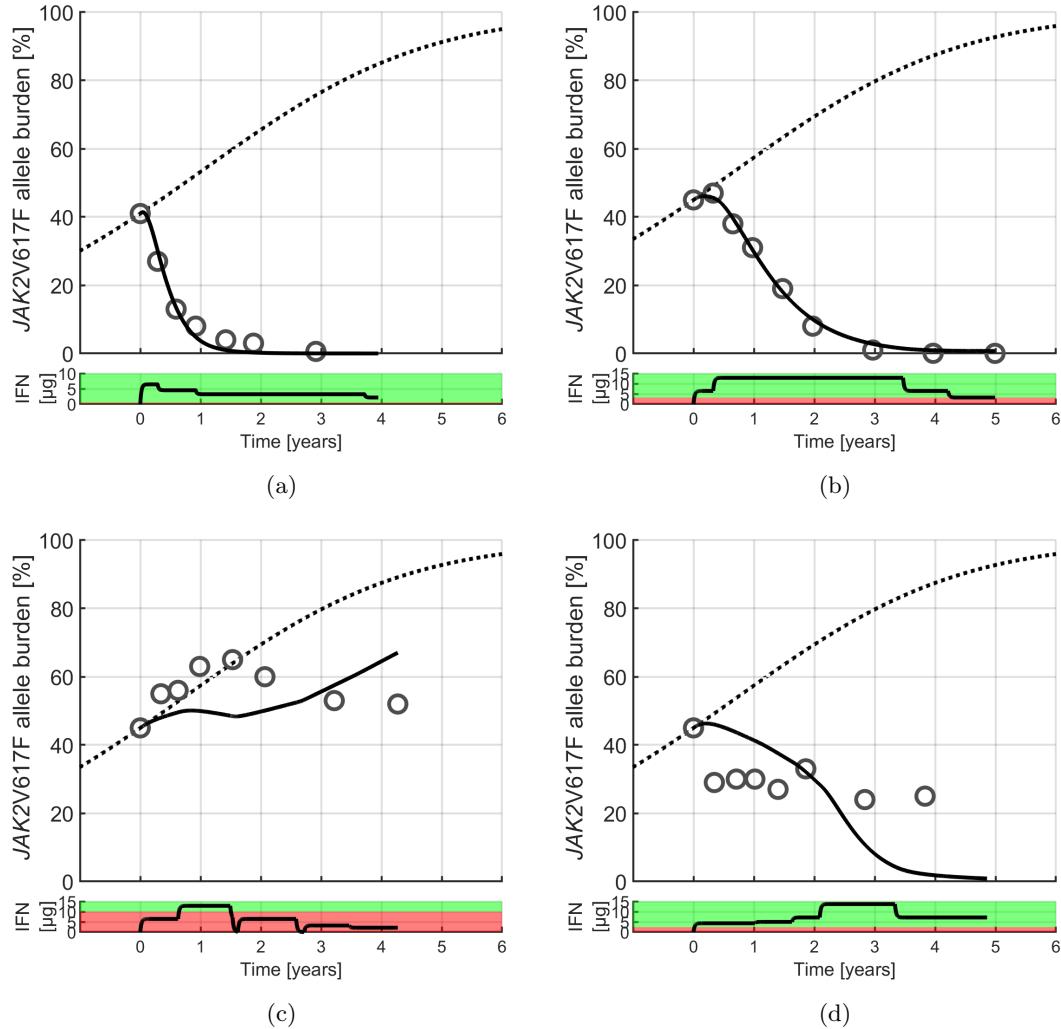


Figure 11.2: Perturbing ρ_L provides a good visual fit for most patients. The results of the first step are shown here. Panel (a) shows patient “P184”, while panel (b) shows patient “P198”, previously shown in examples in section 6. The full black curve in the figures show the relative frequency of the malignant clone in the model. In panel (c), patient “P111” is depicted, as an exemplary patient with a bad response in the *JAK2* allele burden. Based on the particular treatment dosing for the patient, the model is unable to capture the response, with black model-curve showing a visually different response. Another example of a bad fit of the model to data is shown in panel (d). Here the patient, “P164” experiences a sustained level of *JAK2* allele burden, while the model shows a decreasing level, due to the increasing dose. The best-fit parameters ν_{ρ_L} for each patient were -0.024 for “P184”, -0.004 for “P198”, -0.001 for “P111” and finally -0.005 for “P164”. In regards to the adjusted R^2 values described later in the text, the fit to “P184” gave a value of 0.94, “P198” was 0.99, “P111” was -2.88 , and “P164” was -5.30 .

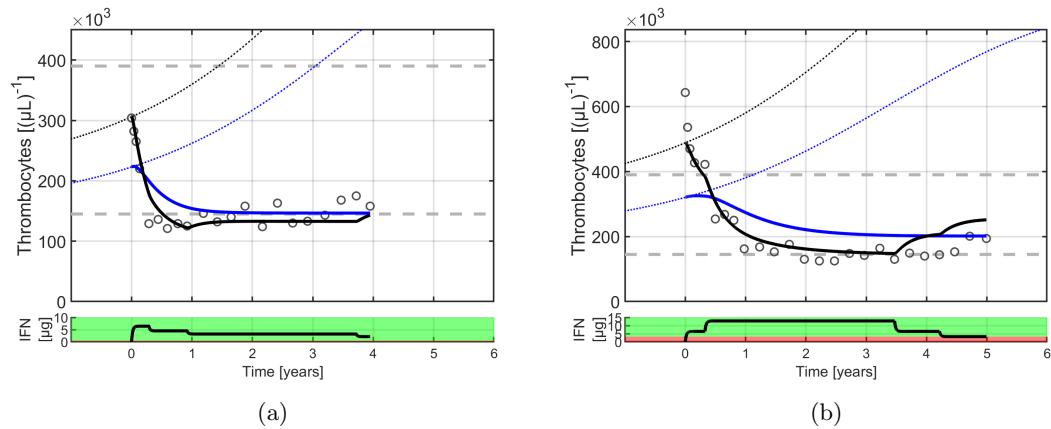


Figure 11.3: Subsequent optimization of ω_H and ω_L can greatly improve the fit to blood-cell counts. Panel (a) displays the differences between the first and second step for patient “P184”, while panel (b) displays the same for patient “P198”. In both panels, the blue curve displays the mature cell counts resulting from perturbing just ρ_L to fit the relative frequency of malignant cells to the JAK2 allele burden measurement of the given patient, scaled to minimize the difference between the sum of mature cells in the model and the thrombocyte counts. The simulated scenarios without treatment is shown as dotted lines. With the same perturbation of ρ_L , we subsequently make a fit of the scaling-parameter and a perturbation of ω_H and ω_L and obtain the curve shown in black. In both of the shown cases, this reduced the difference between data and model. The bottom parts of the figures display the dosing of the given patient, as in previous figures. Patient “P184” has $\nu_\omega = -0.1$ and $\nu_{\rho_L} = -0.024$, while patient “P198” has $\nu_\omega = -0.046$ and $\nu_{\rho_L} = -0.004$.

While the effect of non-zero values of ν_ω leads to significant differences in dynamics of blood-cells, the relative frequency of blood-cells is only changed to a small degree. However, for some patients, the optimization of ν_ω to fit with thrombocyte- and leukocyte-date yields a model-fit where the relative frequency of malignant cells in the model no longer agrees with the JAK2 allele burden measurements. To improve this, the third step consisted of an additional optimization of ν_{ρ_L} . Examples are shown in figure 11.4. While some discrepancies between data and model are still present, the model simulation with the new value of ν_{ρ_L} shown in black in the figure reduces the sum of squared errors. This final optimization makes only insignificant differences in blood-cell counts, and hence the combined set of fit values consisting of ν_{ρ_L} , ν_ω , a leukocyte scaling-factor and a thrombocyte scaling-factor, provides a dose-dependent model fit which allows for good visual agreement between both blood-cell counts and relative frequency of clones. Collecting these fit-parameters for all patients allows for population-level interpretation of how the model fits to data, and in turn, the effect of IFN.

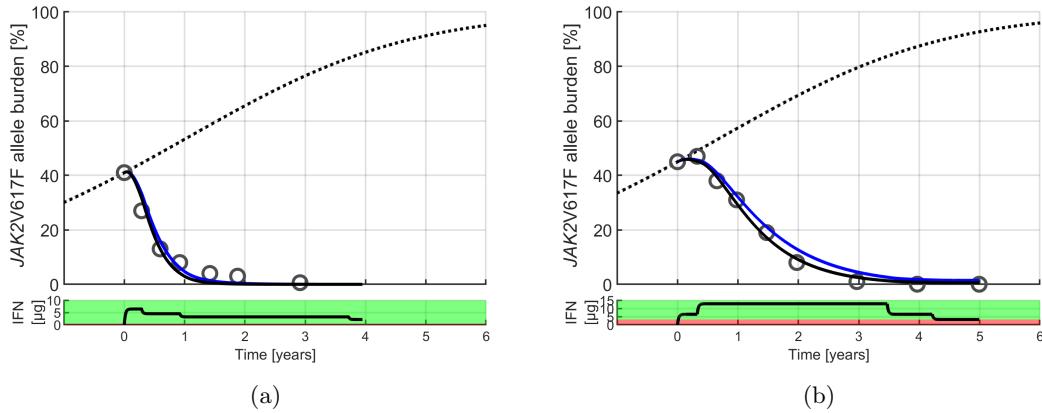


Figure 11.4: Final optimization of ν_{ρ_L} improves agreement with JAK2 allele burden. Changing ν_ω to fit with thrombocyte and leukocyte counts can cause minor changes in the relative frequency of malignant cells, leading to a sub-optimal fit to the JAK2 allele burden. The resulting fit for patient “P184” is shown in panel (a), while the fit for patient “P198” is shown in panel (b). Making a new fit of ν_{ρ_L} to the JAK2 allele burden while maintaining the error-minimizing value found for ν_ω improves the agreement with data. This is shown in black in both figures. In the shown scenarios, ν_{ρ_L} is changed from -0.024 to -0.030 for patient “P184”, and from -0.0035 to -0.0040 for patient “P198”. Note that for most patients, this additional fitting lead to only minor changes in the dynamics of the relative frequency of clones.

11.4 Population-level interpretation of fits to data

For all patients, a linear scaling of the sum of mature cells in the model was made, interpreted as a scaling factor describing how big a ratio of the modelled mature cells that pertain to the given cell-type. Since these scaling factors are independent of how far the disease of the given patients has progressed, the factors can be used to determine a model-estimate of the cell-counts in the healthy steady state. Assuming the model accurately describes the progression that the patient experiences before diagnosis, the found cell-counts are estimates for the actual cell-counts that the given patient had before the disease-initiating HSC-mutation, ignoring other changes to cell-counts unrelated to MPN. Additionally, we can predict how large cell-counts for the given patient would be in the full-blown disease steady state. While treatment is always to be preferred before such state is approached, determining the predicted counts of specific cell-lines in this full-blown disease state can illustrate how bad a worst-case scenario could be for MPN-patients. The distribution of patient-specific cell-counts in the healthy steady state is shown in figure 11.5, while the full-blown disease steady state is shown in 11.6.

We see that for most patients, scaling the healthy steady state leads to cell-counts within the expected interval. Hence, the scaling of the mature cell counts is not only valid during disease and treatment (as suggested by the agreement with data), but also suggests realistic cell-counts in the absence of disease.

For the full-blown disease steady state, both cell-counts are predicted to be significantly heightened. In particular, we observe that the distribution of thrombocyte counts are more varied than leukocytes, and that relative to upper limit of the healthy interval, thrombocytes are more heightened than leukocytes. We confirmed by visual inspection of data from individual patients that heightened thrombocyte-counts were more common than heightened leukocyte-counts. As increased thrombocyte-levels is one of the diagnostic criteria for ET, this was expected.

It is possible to use the fits of the combined model to determine a lower threshold for the IFN-dose required for successful treatment on the level of individual patients. This threshold is defined by the dose at which the leukemic steady state becomes locally unstable and the healthy steady state becomes locally stable, indicative of long-term eradication of the malignant clone. Figure 11.7 shows the distribution of these thresholds. As for the Cancitis model-fits, we predict that treatment can be attained for a dose of $5\mu\text{g}$ for most patients, while a subset of 10 patients out of 63 (16%) require doses above $10\mu\text{g}$ IFN daily.

Since the stability does not depend on ω_H and ω_L , figure 11.7 depends only on the fitted value of ν_{ρ_L} , and the model suggests that higher doses will always lead to faster treatment and hence better patient prognosis. However, by fitting with both ρ_L as well as ω_H and ω_L , the benefits of treatment is separated into a long-term benefit due to ρ_L perturbation and a shorter-term benefit due to ω_H and ω_L , as discussed in section 10 and illustrated in figure 10.9. In figure 10.9c we showed how sufficient perturbations of both ρ_L , ω_H and ω_L can lead to a 50% decrease in the relative frequency of malignant cells and cause a decline of all mature cells such that cell-counts equal to the healthy steady state are achieved. For each patient, we determined the dose required to achieve both of these goals within a five year period of sustained treatment, based on simulation of an identical treatment scenario. Twenty years after initial mutation, as in the examples considered throughout section 10, treatment was initiated with the response to treatment mimicking those of the DALIAH trial patients. Determining the threshold between doses that lead to both goals being achieved and doses that did not, allowed us to interpret the patient-specific fit-parameters as indicative of the dose required for treatment within a five year period. The distribution of dose-thresholds is shown in figure 11.8. These results suggest that for most patient, achieving both the goal of reducing the relative frequency of malignant cells by 50% and attaining healthy cell-counts within a five year period, an average daily dose of at least

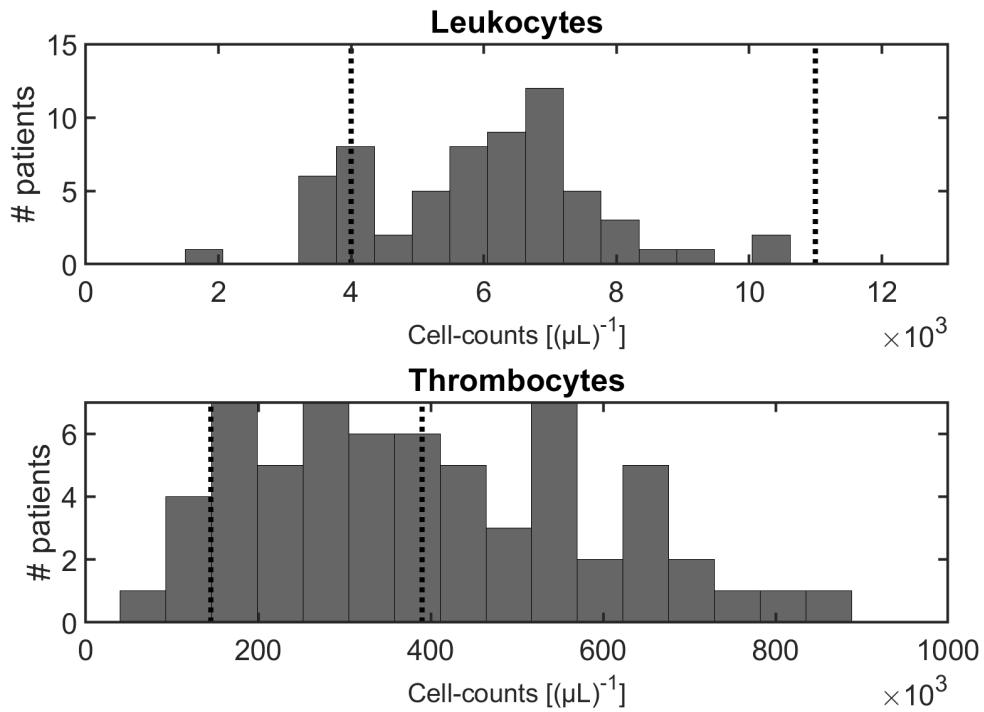


Figure 11.5: Distribution of counts of mature cells in the healthy steady state. Patient-specific scaled values of leukocytes and thrombocytes in the healthy steady state shown as histograms. The approximate “normal range” of cell-counts is shown with dotted black lines, defined as between $4 \cdot 10^3 (\mu L)^{-1}$ and $11 \cdot 10^3 (\mu L)^{-1}$ for leukocytes and $145 \cdot 10^3 (\mu L)^{-1}$ and $390 \cdot 10^3 (\mu L)^{-1}$ for thrombocytes. The model-fits suggest that most patients had cell-counts within these intervals before disease onset. The leukocyte-counts are below $4 \cdot 10^3 (\mu L)^{-1}$, and hence possibly underestimated for 7 patients, while thrombocyte-counts are above the interval for approximately a third of patients, suggesting a possible overestimation of the thrombocyte count.

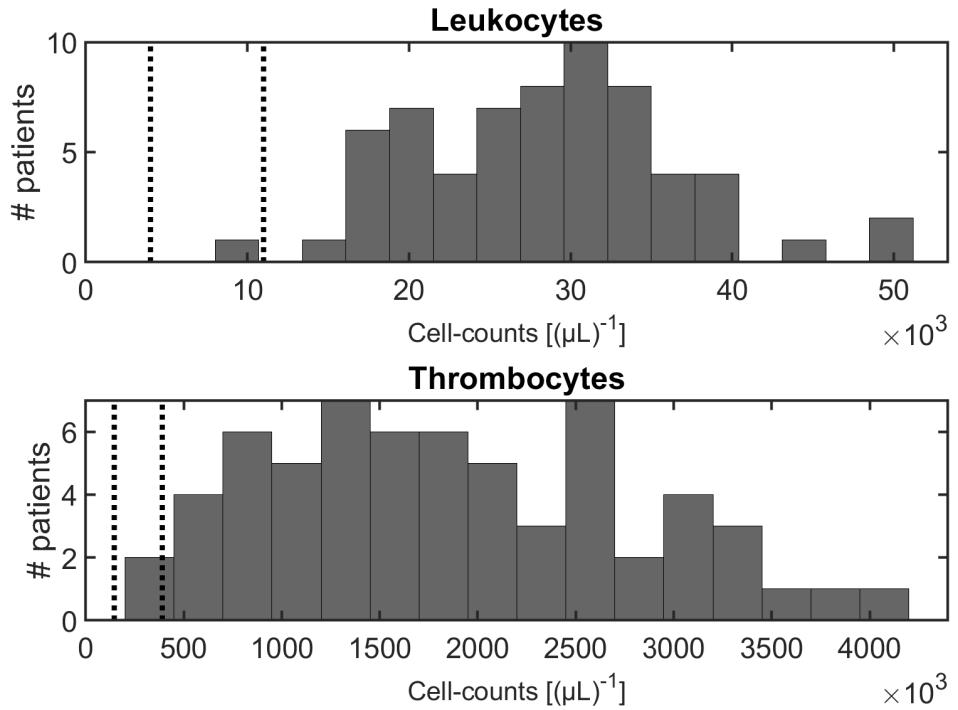


Figure 11.6: **Distribution of counts of mature cells in the full-blown disease steady state.** Histogram of patient-specific scaled values of leukocytes and thrombocytes for the full-blown disease steady state. Healthy intervals are shown for comparison, as in figure 11.5. For almost all patients, the model predicts that full-blown disease would result in significantly increased cell-counts, with leukocyte counts as high as $50 \cdot 10^3 (\mu L)^{-1}$, and thrombocyte counts up to $4.1 \cdot 10^6 (\mu L)^{-1}$. The median value of leukocyte counts is $29 \cdot 10^3 (\mu L)^{-1}$, corresponding to 264% above the upper part of the healthy interval, while the median thrombocyte count is $1831 \cdot 10^3 (\mu L)^{-1}$, corresponding to 469% above the threshold.

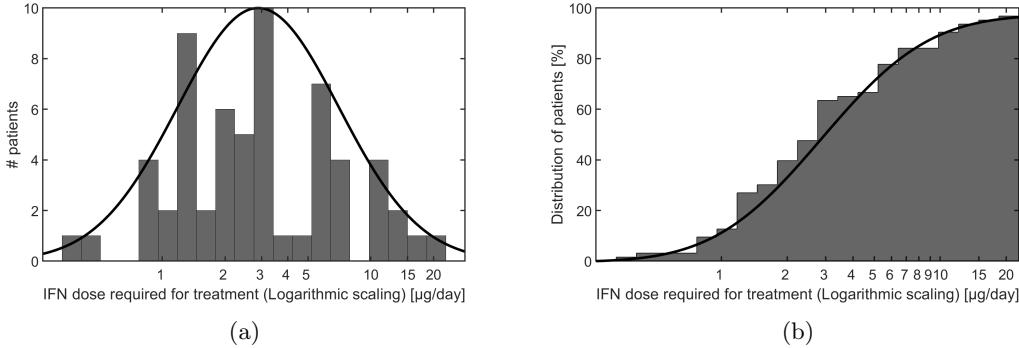


Figure 11.7: Distribution of minimum IFN-dose resulting in successful treatment based on the fits of the combined model. Patient-specific estimates of the required IFN-dose were determined. Panel (a) shows a histogram of the found threshold-doses, with a log-normal distribution shown in black to guide the eye. In panel (b) the cumulative percentage of patients that are predicted to have successful treatment at a given dose is shown, with a cumulative curve of the log-normal distribution from panel (a) shown in black. Two patients were found to require doses greater than $30\mu\text{g}$. As this is outside the normal range of IFN-doses, the two patients were not included in panel (a). Hence, a maximum of 97% of patients (61/63) is shown in panel (b).

$5\mu\text{g}$ IFN must be sustained throughout the period.

11.5 Proof-of-concept of population-level responses

The fitted parameters ν_{ρ_L} and $\nu_{\omega_H} = \nu_{\omega_L}$ resulted in model behaviour that agreed well with patients on an individual level. On a population level, patient responses differed significantly, partly due to the heterogeneity in data, previously discussed in section 3. This limits the reliability of any interpretation and prediction about population-level responses. However, as a proof-of-concept, we here show how the modelling framework described can lead to pre-treatment predictions that in some cases visually agrees well with patient responses.

Adjusted R^2 values were computed for the JAK2 allele burden measurements and for the thrombocyte measurements, defined as:

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n E_i^2}{\sum_{i=1}^n (D_i - \bar{D})^2} \frac{(n-1)}{(n-p-1)} \quad (11.4)$$

where E_i is the error of the model at the i 'th data-point, D_i the i 'th data-point, \bar{D} the mean of data, n the number of data-points and p is the degrees of freedom. As two parameters were considered ν_{ρ_L} and $\nu_{\omega} = \nu_{\omega_H} = \nu_{\omega_L}$, $p = 2$. An adjusted R^2 close to 1 indicates a good fit, while low or negative values indicate a bad fit. The adjusted R^2 was used as it will be lower for patients with fewer data-points, compared to patients where the model fits well to more data-point. By visual inspection of model-fits, we defined a cut-off between good and bad fits, defined as $\bar{R}_J^2 = 0.5$ for the JAK2 allele burden and $\bar{R}_T^2 = 0.6$ for thrombocytes. This gave us a population of the 20 best fits of the combined model to data. For all 20 patients, the values of ν_{ρ_L} and $\nu_{\omega} = \nu_{\omega_H} = \nu_{\omega_L}$ were negative. Considering the negative of the fitted values, and

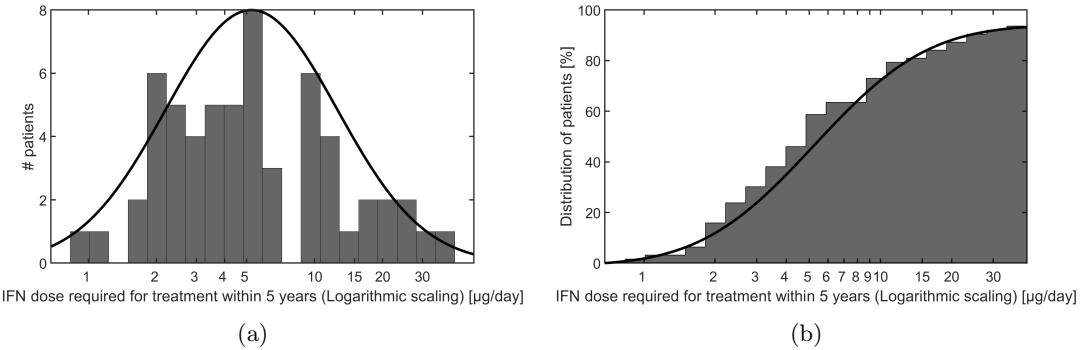


Figure 11.8: **Distribution of minimum IFN-dose resulting in blood-cell normalization and halving of the relative frequency of malignant cells within one year of treatment.** A common scenario similar to the examples shown in section 10 was simulated in which treatment was initiated 20 years after the first mutation. The response to treatment was simulated in accordance with the fitted parameters of each individual patient. Five years of treatment was simulated at a constant dose-level. Panel (a) shows a histogram of the distribution of doses required to attain both blood-cell normalization and a reduction of the relative frequency of the malignant cells by 50%, with a logarithmic x-axis. A log-normal distribution is shown as a black curve to guide the eye. In panel (b) a cumulative histogram is shown, together with a cumulative curve of the log-normal distribution from panel (a). For six patients, the necessary average daily dose was found to be greater than $30\mu\text{g}$. Four of these were excluded from panel (a), and hence panel (b) has a maximal number of patients at 93.7% (59/63) patients.

assuming a log-normal distribution of parameters, distributions of ν_{ρ_L} and $\nu_\omega = \nu_{\omega_H} = \nu_{\omega_L}$ were determined, see figure 11.9.

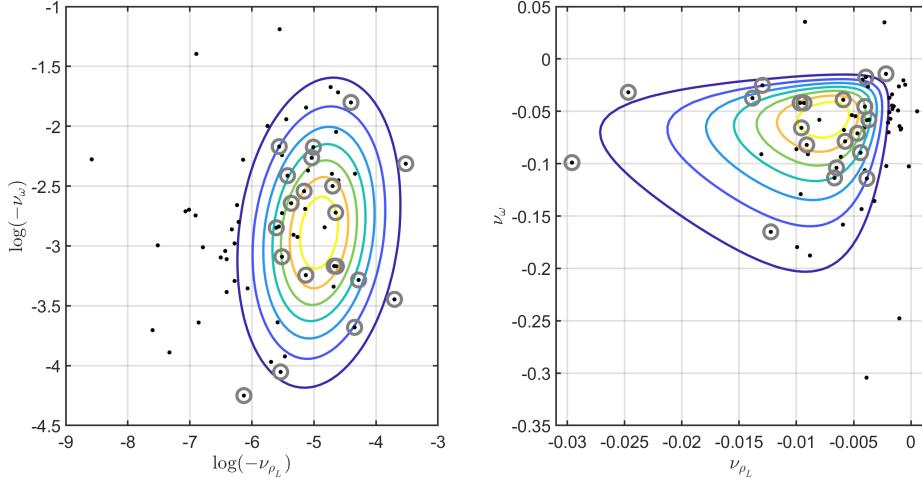


Figure 11.9: **Distribution of fitted parameters ν_{ρ_L} and $\nu_\omega = \nu_{\omega_H} = \nu_{\omega_L}$.**

All patients are shown as black dots, while grey circles are added to patients for which the adjusted R^2 values above the chosen thresholds described in the text. The left-hand panel depicts the natural logarithm of the $-\nu_{\rho_L}$ and $-\nu_\omega$ for each patient, while the right-hand panel depict the untransformed fitted values. Note that two patients had $\nu_\omega > 0$ (Patients “P014” and “P070”), and one patient had $\nu_{\rho_L} > 0$ (Patient “P089”) and are only shown in the right panel. The adjusted R^2 for these three patients were all below the threshold, and hence they were excluded due to measure of the goodness-of-fit and not because of the positive fit-parameter values. The lines depict the two-dimensional log-normal distribution that fits best to the 20 patients with acceptable adjusted R^2 values in the transformed space in the left-hand panel. The distribution was transformed back and shown in the right-hand panel as well. In the transformed

log-space depicted in the left-panel the distribution has $\mu = \begin{pmatrix} -4.92 \\ -2.89 \end{pmatrix}$ and

covariance matrix $\Sigma = \begin{pmatrix} 0.42 & 0.07 \\ 0.07 & 0.41 \end{pmatrix}$. Hence, the mean of the distribution corresponds to $\bar{\nu}_{\rho_L} = -0.0073$ and $\bar{\nu}_\omega = -0.0556$.

For each of the 20 patients whose data the combined model fit well with, 1000 sets of ν_{ρ_L} and $\nu_{\omega_H} = \nu_{\omega_L}$ were randomly chosen from the distributions, leading to 1000 virtual patient-responses. The initial conditions of the virtual patients were determined from the baseline JAK2 allele burden measurement of the real patient. The scaling-factors for thrombocytes and leukocytes were chosen such that the simulated model agreed with baseline measurements of the given cell-counts. Hence, the virtual patients were all made to agree perfectly with patient-data at baseline.

Since changes in doses had a significant impact on patient-response, all virtual patients were

assumed to undergo the same changes in IFN dose, regardless of whether criteria for dose-increases or dose-decreases were met or not. This was done to avoid discrepancies between the real patients and the virtual patients that would arise solely due to changes in dose.

Figure 11.10 displays an example of the distribution of virtual patient-responses, compared to patient data for the particular real patient. The rest of the 20 patients which the combined model fitted well with along with similar distribution of virtual patient-responses is shown in Electronic Supplementary Material 5. The example shown in figure 11.10 is an example of a patient where the virtual patients agree particularly well with data. For some of the patients, the distribution of virtual patient-responses appears to agree well visually, while other patients had responses different from the virtual responses. The linear scaling of mature blood cells counts is observed to lead to bad fits for particular patients. As the scaling-factors determined during patient-fitting was based on all data, such factors would not be available based solely on the baseline measurements, and hence scaling with just the baseline measurement is the only possible option for this type of pre-treatment prediction.

Finally, we consider an entire simulated patient-response. For patients diagnosed with PV, the mean leukocyte-count at baseline was $11.4 \cdot 10^3 (\mu L)^{-1}$, the mean thrombocyte-count at baseline was $571 \cdot 10^3 (\mu L)^{-1}$, while the mean JAK2 allele burden was 44%. Considering a PV patient with these baseline values and a constant dose of $5\mu g$ IFN (the typical initial dose for interferon-alfa2b “PegIntron”), we simulated 1000 virtual responses. The resulting distribution is shown in figure 11.11. The figure illustrates how the model predicts an idealized PV patient to respond to treatment.

We emphasize that figures 11.10 and 11.11 depict a proof-of-concept of how population modelling of dose-dependent IFN-responses could be used to assess possible patient responses before treatment. Due to the various restrictions on the data used and other limitations, the reliability of these virtual patient responses are uncertain, and further work must be done before the prediction of the virtual patients can be used in the clinic to predict patient response.

11.6 Summarizing discussion

The combined model was shown to allow for dynamics that are in agreement with the patient data of the DALIAH trial. IFN-dose dependence of parameters were modelled in a way similar to how the Cancitis model was fitted to data in section 6 and the work described in (Ottesen et al., 2020). As the combined model must have $0 \leq \rho_H \leq 1$ and $0 \leq \rho_L \leq 1$, these parameters were scaled in accordance with equation (11.1), while all dose-dependent perturbation of any other parameter is determined by equation (11.2).

In section 10 we found that the effect of IFN observed in patient-data from the DALIAH trial could be captured by perturbing ρ_L and ω_H and ω_L , with the former primarily affecting the long-term response of the relative frequency of the clones and the latter affecting short-term decrease of blood-cells. For this reason, a multi-step procedure was adopted, in which ρ_L was perturbed initially to allow for agreement with the JAK2 allele burden measurements of individual patients. This was followed by a scaling of the mature blood-cell-counts and an equal perturbation of ω_H and ω_L , to allow for agreement with both thrombocyte and leukocyte-counts. After a final minor improvement of ρ_L , a fit was obtained such that the model were in good visual agreement with most patients. Through scaling of the mature cells to fit with blood-cell counts, we found that, for most patients, the model predicted realistic counts of cells in the absence of disease. Additionally, the scaling suggests that were treatment not initiated, most patients would ultimately have reached a stage of disease with thrombocytes counts above one million per microliter, significantly above the level at which the patient would be at risk of thrombosis.

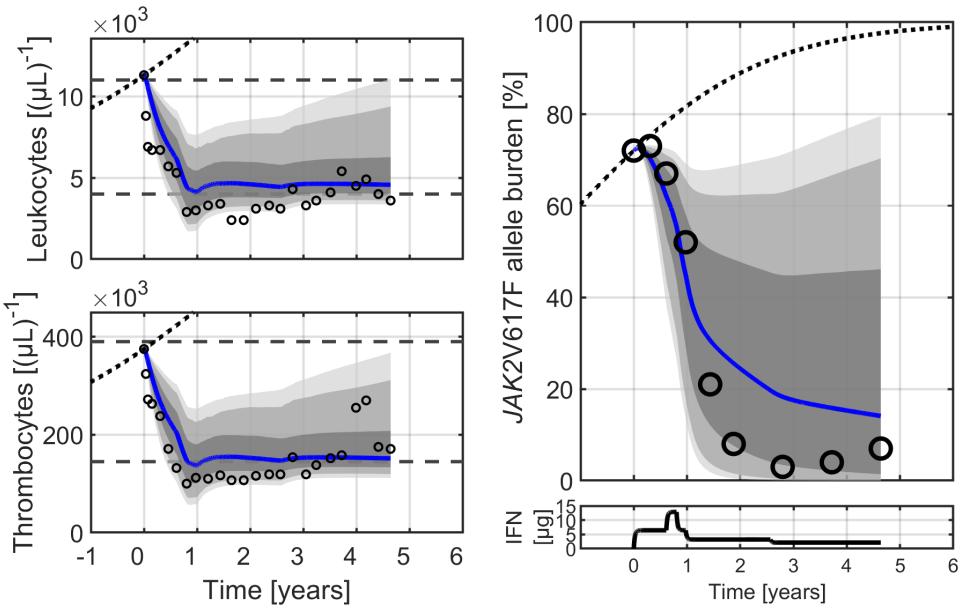


Figure 11.10: Virtual patient responses based on baseline measurement and IFN dosing shows good agreement with real patient-response in some cases. Patient data for patient “P082” is shown as black circles \circ . 1000 virtual patients were simulated and the sum of mature cells were scaled to agree with the baseline data-point for either leukocytes or thrombocytes. The blue curve shows the median response-curve. The shaded grey areas displays the distribution, with the darkest grey showing the interval from 25% to 75% of values at the given time-points, the next-darkest interval shows from 10% to 90% while the final interval from 5% to 95% of virtual patient-responses is shown in light grey. The bottom right panel displays the IFN blood-concentration used for both the real patient and the virtual patients.

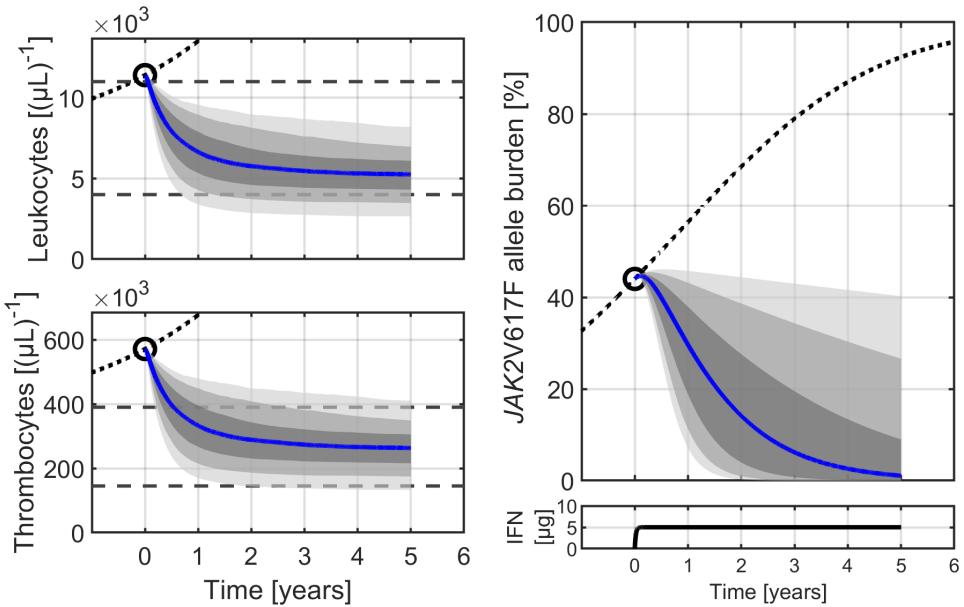


Figure 11.11: Virtual patient responses based on average PV baseline measurements shows the distribution of an idealized IFN response. Based on the average baseline values of PV patients, shown as black circles \circlearrowleft , 1000 virtual patients were simulated and the sum of mature cells were scaled to agree with the baseline data-point for leukocytes and thrombocytes in the top-left and bottom-left panel respectively. The blue curve shows the median response-curve. The shaded grey areas displays the distribution, with the darkest grey showing the interval from 25% to 75% of values at the given time-points, the next-darkest interval shows from 10% to 90% while the final interval from 5% to 95% of virtual patient-responses is shown in light grey. The bottom right panel displays the IFN blood-concentration used, corresponding to a constant $5\mu\text{g}$ IFN dose.

The local stability of the steady states of the combined model allowed us to determine a IFN-dose threshold above which the malignant steady state becomes unstable and the healthy steady state becomes locally stable. This threshold can be interpreted as the minimum dose required for long-term treatment of the patients, and was found to be below $5\mu\text{g}$ for two thirds of the patients, suggesting that the initial dose of IFN typically used is sufficient for long-term treatment for most patients.

Considering the daily average IFN dose required to both reduce the relative frequency of malignant cells by 50% and to normalize total blood-cell counts within a five year treatment period, we found that $5\mu\text{g}$ IFN is sufficient for only about half of the patients. For approximately a fourth of the patients, these two measures would not be attained within five years of treatment with a daily dose of $10\mu\text{g}$ IFN. These findings were based on simulations in which treatment was initiated 20 years after the initial mutation of a malignant stem cell. Thus, while this suggests that doses above $10\mu\text{g}$ IFN daily are necessary to attain successful treatment within a five year period for a significant proportion of patients, earlier diagnosis and treatment could change this outcome. Similarly, changes in dosing such as e.g. intermittent high-dose IFN and/or short-term treatment cessation could affect the dose required to attain both goals. Optimizing dose scheduling to minimize the average dose while attaining both goals could be possible and should be explored in future work.

While the patient-specific fits were (local) minima of the sum of the squared error between model and patient-data and typically gave a good visual agreement with data for most patients, there were some patients for which the model-fit differed from the actual treatment-response observed in data. We hypothesize that some of this is due to heterogeneity of patient-responses, to be discussed in detail in the next chapter. Different methodologies for fitting were tested for comparison. In particular, we initially fitted the model to all data simultaneously. However, due to scaling-issues with the three biomarkers considered (JAK2 allele burden, thrombocytes and leukocytes), our initial tests lead to agreement with one measure but disagreement with the others for the vast majority of patients. This lead us to the presently described methodology which, while sometimes suboptimal for certain patients, gave a satisfying visual fit to the data of most patients.

Based on an adjusted R^2 measure of goodness-of-fit, the 20 patients that the model agreed best with were used for modelling a population-level response to IFN treatment. As the patients chosen represent the 20 patients that the model agreed with the best, there is a risk that the population-level responses found are biased toward a particular type of response. Hence, this final part of the work should primarily be considered a proof-of-concept. Regardless, the collection of responses of the 20 patients were used to determine a distribution of dose-dependent model parameters, which allowed us to do population modelling, showing how simulated patients would respond to a given IFN-dose schedule. Comparing with patient data, as shown in figure 11.10 and in Electronic Supplementary Material 5, the distribution of virtual patient captured the data of the real patients well, based solely on the baseline measurement taken before treatment initiation. For some patients, scaling the blood-cell counts with baseline measurements lead to bad agreement between the patient behaviour and the virtual patients. As an example, patient “P021” shows disagreement between the virtual patients and the thrombocyte-counts. However, the baseline thrombocyte-measurement appears to be much lower than the next measurements taken just weeks later. This could be indicative of an erroneous baseline measurement, which consequently leads to a bad scaling of the virtual patients. Similar problems with the scaling of blood cells can possibly explain discrepancies for other patients, such as e.g. higher leukocytes counts of the virtual patients observed in patient “P028”. Additional samples at or before baseline could help alleviate such problems. In figure 11.11, we showed the predicted distribution of responses of an idealized PV patient, defined as a patient with baseline cell-counts and JAK2

allele burden equal to the average values for all PV patients. This hypothetical patient-response illustrates how the population-level response is distributed on a conceptual level, leading to almost all virtual patients having thrombocyte-counts in the healthy range within five years, as well as a significantly reduced JAK2 allele burden, even without increasing the IFN-dose. However, the figure also shows how the reduction of thrombocyte-counts can be slow. For the median response, shown in blue in the figure, heightened thrombocyte-counts were still observed after the first four months of treatment. As discussed in chapter 3, this was a criteria for dose escalation in the DALIAH trial. Hence, the model predicts that the dose for an idealized PV would be increased after four months, in agreement with the changes in dosing observed for many patients in the DALIAH trial. Thus, while the model suggests that five-years of low-dose treatment is sufficient for achieving the clinical goals of reduced cell-counts and reduced JAK2 allele burden, the model also suggests that the goals are not achieved quickly enough without increasing the IFN-dose.

For the results discussed in this section, we considered perturbations of the parameters ρ_L , ω_H and ω_L . As such, we did not include perturbation of the parameters u_H and u_L , related to the rate of release or activation of niche-bound (or quiescent) HSC. This is regarded as one possible effect of IFN, see section 2.1. Preliminary investigation not presented here showed that including perturbations of both u_H and u_L could improve the fit of the model to data. In particular, an increase in u_C and u_L was found to allow for a short-term increase in the relative frequency of clones, as observed in a small group of patients. However, fitting with two additional parameter introduced uncertainty and ambiguity to the model-behaviour due to the risk of overfitting. For this reason, we only perturbed ρ_L , ω_H and ω_L , with the equal perturbations of ω_H and ω_L , to reduce the degrees of freedom.

In conclusion, we investigated if the IFN-treatment responses observed in the DALIAH trial could be explained by a reduction of the parameters ρ_L , ω_H and ω_L of the combined Cancitis-Niche model, associated with increasing differentiation and/or apoptosis of malignant stem cells. Model-fits to data were determined to be visually satisfactory, suggesting that our hypothetical interpretation of the effect of IFN on a biological level is reasonable. Importantly, we found that from the two types of data considered, JAK2 allele burden and blood-cell counts, two separate effects of IFN could be determined: A slow decrease of the JAK2 allele burden due to changes to HSC competition and a fast decrease of blood-cell counts due to a lowered production of both healthy and malignant mature cells.

Part VI

Discussion and concluding remarks

Chapter 12

General discussion of findings, limitations and future work

In this thesis, we have described our work toward accurate mathematical modelling of blood cancers, particularly MPNs. In this final chapter, the results and overarching themes are discussed and the different findings are related to each other. Finally, we describe some of the future challenges that still await for mathematical modelling of MPNs and give a prediction for how overcoming these challenges could benefit clinicians and patients.

12.1 Implications of the proposed models

A range of mathematical models were investigated and discussed in this thesis: The Cancitis model due to Andersen et al. (2017) (Section 5), the proposed HSC niche model (Section 8) and finally a combined Cancitis-Niche model (Section 10).

The proposed HSC niche model was formulated by considering the biological processes that HSC undergo, with an assumption of a limited pool of HSC niches and a novel idea of post-cell-division exhaustion of the divisional capacity of HSC. Simulating the niche model suggested that significant differences between the relative frequency of a malignant clone in the HSC compartment and in the mature blood-cell compartment could arise naturally due to differences in HSC properties. Additionally, the niche model allowed for a thorough investigation of combinations of HSC mobilization therapy, transplantation and pre-conditioning HSC ablation. We found that short-term HSC mobilization was beneficial in a HSC transplantation scenario, particularly in combination with preconditioning of reduced intensity.

Analysis of the properties of the model showed that a quasi-steady-state approximation was appropriate, resulting in a reduced model. By transforming the variables such that the sum of cells and the relative frequency of the clones involved are modelled, rather than counts of individual clones, an equivalent model arose, allowing for easier comparison with experimental data. In addition, the transformation allowed for global analysis in a two-clone scenario. Simplifying the model further, a simple logistic expression arose, describing just the relative frequency of a particular clone. This is a model of e.g. the chimerism of a transplanted HSC-clone or the relative frequency of a malignant HSC-clone out of all HSC. The dynamics of the logistic approximation was found to agree well with the kinetics of the JAK2 allele burden of a cohort of patients from the DALIAH trial, suggesting that the simple kinetics of the JAK2 allele burden can be explained by changes in the properties of HSC.

The success of using the single-parameter logistic approximation to describe the JAK2 allele burden in data highlights a problem of identifiability of the parameters of the HSC-niche model, as briefly discussed in section 9.5. Any particular value of ϕ that leads to good agreement between the logistic approximation model and the JAK2 allele burden data could arise in a multitude of ways from the parameters of the reduced HSC-niche model and in turn even more ways from the parameters of the full HSC niche-model. As a consequence, if model parameters were to be determined solely on data that the logistic approximation can accurately fit to, such as the examples of JAK2 allele burden shown in figure 9.1, the resulting parameters would be highly uncertain. Hence, any interpretation of results based on the parameters would likewise be uncertain.

The combined Cancitis-niche model retains many of the features of both models, such as the positive feedback on blood-production through immune-system regulation from the Cancitis model and the details of HSC properties of the niche model. Default parameters were based on the default parameters of the Cancitis model and of the reduced HSC niche-model. To determine the parameters related to the malignant stem cells that give rise to MPNs, ρ_L , α_L and u_L , we fitted the disease growth to agree with the Cancitis model and the exponential growth described in (Pedersen et al., 2020). As discussed above, determining three parameters on such limited data results in parameter-values that are uncertain. However, with the limited data available for untreated patients, a more certain estimate was difficult to make. In addition, initial investigations determined that changes to the three parameters had little effect on the overall behaviour of the model, particularly for the later considerations about different treatment-scenarios. We believe that the combined model provides a reasonable middle-ground between oversimplified empirical models and complicated models with excessive biological detail. The balance is, in our opinion, obtained thanks to the combination of model reductions and the increased detail of HSC-biology added through the niche model. Such questions of model complexity are important for successful and useful mathematical modelling and should be considered and evaluated with great care.

The immune system feedback considered in the Cancitis model and in the combined model suggests that inflammation plays an important role in the progression of MPN. The increased external inflammation, due to e.g. smoking, typically only has a small effect in relation to the qualitative behaviour of the system, as observed along the horizontal axes of figure 5.5. However, as illustrated by figures 5.5 and 5.3 for the Cancitis model and figure 10.10a for the combined model, increased inflammation leads to a significant increase of both healthy and malignant mature blood cell-counts in both models. This suggests that patients with heightened external inflammation are at risk of complications due to high cell-counts earlier along the progression of MPN compared to patients with low external inflammatory stimulus. As the models suggest that inflammation increases as the disease progresses, the increased cell-counts related to the malignant clone is heightened further and the positive feedback between malignant cell-production and inflammation provides an explanation for rapid disease progression in later stages of MPN.

All three models suggest that treatment of MPN must act on stem cell level. The reduced Cancitis model suggests that long-term dynamics are decided by expressions of relative reproduction of the two HSC clones considered, as described by the primary and secondary reproduction numbers, \mathcal{R} and \mathcal{S} . In the niche model and the combined model, we determined an expression of fitness for an individual HSC clone. By perturbing parameters related to only the mature blood-cells or to the immune system, it was possible to decrease malignant mature cell-counts and hence reduce the related health-risks. However, such effects were found to be unsustainable if the competition between healthy and malignant HSC were not perturbed as well, in the sense that following treatment cessation, cell-counts would quickly return to dangerous levels. Loosening the hold that the malignant clone has on the bone-marrow microenvironment is thus

necessary for the success of long-term treatment. The notion of fitness of HSC is not a novel idea (see e.g. (Traulsen et al., 2010) or (Watson et al., 2020)), however, the notion appears to arise naturally from the niche model, and is maintained even when the model is reduced, simplified and combined with the Cancitis model. This robustness of a simple expression of stem cell fitness suggests that considering HSC clones as populations competing for resources is not only a useful idea, but in fact an accurate representation of the interaction between stem cells in the bone-marrow microenvironment. This highlights the importance of determining the fitness of a stem cell clone clone, or more importantly, the relative fitness between the population of healthy clones and a malignant clone.

As described in section 10.1, three biological assumptions about progenitor cells implied that considering progenitor cells explicitly was unnecessary in the combined model and a quasi-steady-state approximation of intermediate steps between stem cells and mature cells was appropriate. The argument holds in general, implying that if self-renewal of progenitor is limited and differentiation is faster than the time-scale of HSC dynamics, then the dynamics of progenitors are qualitatively similar to the production of the first step of progenitors and hence on the differentiation of HSC. Additionally, we found evidence that the ratio of HSC that are quiescent could be related to the life-span of mature blood cells. In the model parametrizations, the average life-span of mature cells ($d_{x_1}^{-1}$) was found to be comparable to the average time between HSC release from the niche (u_j^{-1}). This meant that a quasi-steady-state approximation of the dynamics of mature blood cells was not appropriate. In the niche model, the release of HSC from the niche was related to the ratio between the quiescent HSC and the sum of all HSC. Hence, we hypothesize that the fraction of quiescent HSC and the time-scale of HSC behaviour within the bone-marrow could be directly determined by the necessity of replenishing the blood-cells that undergo apoptosis on the comparable time-scale. We emphasize that further investigation into such hypothetical connection is necessary before any conclusions about a connection can be made.

12.2 Limitations of data

Understanding the progression and treatment of MPNs rely on the data available. We here comment on the DALIAH trial data, as discussed in section 3, without explicitly considering the work on relating the mathematical models to the DALIAH trial data (sections 6 and 11). Throughout this thesis, we assumed that the MPN diagnoses of ET, PV and myelofibrosis all exist on a continuum, and that the particular diagnosis did not allow for a meaningful distinction of patient for our purposes of modelling the response to IFN treatment. Although baseline thrombocyte counts and JAK2 allele burden was determined to be different for ET patients compared to other diagnoses in a statistically significant sense, the responses to treatment were visually similar. Additionally, statistical testing of the parameters fitted to data in section 11 showed no significant distinction between groups of patients with different diagnoses.

We hypothesize that one reason behind the lack of discernment between the diagnosis and the variance in fitted parameters is due to heterogeneity in patient-responses to IFN as well as the differences in IFN-dosing that followed as a consequence of patient heterogeneity. Particular examples of the heterogeneity of patients with similar IFN-scheduling is shown in figure 12.1. As previously discussed, differences in patient responses are fairly common in oncology, and some classification of the patient-responses, such as those described by (Lewin et al., 2020), could be possible, as shown in figure 3.6. Unfortunately, the low number of patients ultimately meant that it was difficult to divide patients into groups that were not too small, particularly if the grouping was based only on pre-treatment measurement. However, we emphasize that

our assessment of this difficulty is based primarily on the measurements of JAK2 allele burden, thrombocyte-counts and leukocyte-counts. Additional biological markers were obtained in the DALIAH trial, but were not considered in the present work. Different bio-markers (such as e.g. hemoglobin-levels, considered as part of the PV diagnostic criteria (Arber et al., 2016), or inflammation-related cytokines) could possibly lead to a pre-treatment distinction of patients that respond well to treatment and patients that do not. Such distinction could in turn allow for splitting the distribution of fitted model-parameters discussed below into different types of IFN-responses, and hence possibly a robust classification of patient response-types.

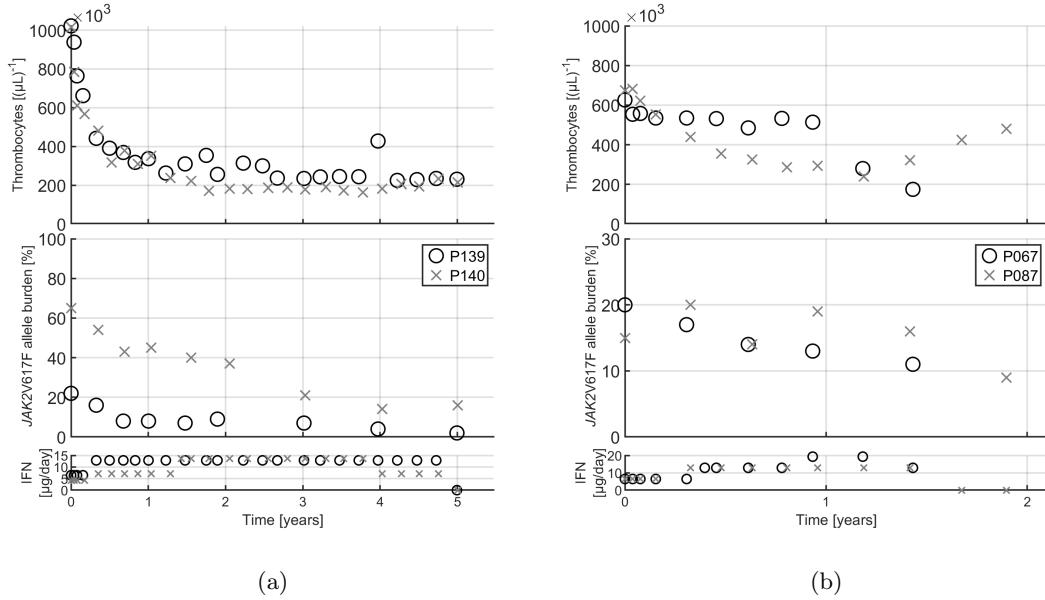


Figure 12.1: Examples of patient heterogeneity in response to IFN treatment. In both panels, thrombocyte-counts, JAK2 allele burden measurements and IFN dosing is shown. IFN dose is given as the future daily average, i.e. the dose prescribed to follow after the given data-point. The left-hand panels displays the data for two patients diagnosed with PV (“P139” and “P140”) for which the thrombocyte-counts decline similarly in response to comparable doses of IFN. While the kinetics of the JAK2 allele burden appear visually similar, the JAK2 allele burden of patient “P140” is approximately three times as high as the JAK2 allele burden of patient “P139” for all data-points. In the right-hand panels, patients “P067” and “P087” are compared. Both patients were diagnosed with ET. The JAK2 allele burden of both patients appear mostly unchanged the first 1.5 year of treatment, while the thrombocyte-counts differed significantly. “P067” experienced a sustained thrombocyte count until around one year after baseline, where it declines rapidly, possibly due to the increase in IFN dose. In contrast, “P087” experienced a steady decline of thrombocytes, until an increase after 1.2 year. Note that both patient dropped out of the study after the date of the final data-point.

In section 3.1 some cursory statistical investigations of the patient-data at baseline were presented. Some correlation between baseline JAK2 allele burden and leukocyte counts was

observed in the DALIAH trial data, however the JAK2 allele burden did not correlate with thrombocyte counts at baseline. For both the empirical modelling described in section 3 and in (Pedersen et al., 2020) as well as in the model-fits described in sections 6 and 11, we assumed that disease progression followed a common dynamic where both JAK2 allele burden and mature blood-cell counts increased as time passed from the initial $JAK2^{V617F}$ mutation. In particular, we interpreted that relative frequency of mature blood-cells (and hence thrombocytes) as a measure of the JAK2 allele burden. While the lack of correlation between thrombocytes and the JAK2 allele burden could suggest that our interpretation of JAK2 allele burden is not appropriate, there are other possible explanations behind it. In figure 10.10, we showed how changes in external inflammatory stimulus and changes in the number of available HSC-niches can lead to significantly different relative frequencies of malignant blood-cells out of all blood-cells for the same sum of blood cells at diagnosis. Hence, differences in inflammatory stimulus or bone-marrow micro-environment could explain why some DALIAH patients experienced heightened blood-cells counts together with low JAK2 allele burden at baseline and vice versa. Another possible hypothesis for why the thrombocyte-count and the JAK2 allele burden could be due to the specifics of the $JAK2^{V617F}$ mutation of the given patients. As mentioned in section 2.1, the mutation is either heterozygous or homozygous. For a heterozygous JAK2 mutation, the maximal JAK2 allele burden possible is 50%. Evidence suggests that PV typically arise from the expansion of a homozygous clone while ET is from a heterozygous clone (Kjær, 2020; Godfrey et al., 2012). Hence, our interpretation of the JAK2 allele burden as a measure of disease level could have been underestimating the stage of disease for patients with JAK2 mutations that were heterozygous. Whether distinguishing between heterozygous and homozygous clone would lead to better correlation between thrombocyte-count and JAK2 allele burden at diagnosis is unclear, as this information were not available to us at the time of writing.

To understand the IFN-response of MPN patients, it is first necessary to understand how blood-cell counts and bio-markers behave in the absence of treatment. However, since patients are typically treated shortly following the MPN diagnosis, data of patients not receiving treatment is naturally very limited. In (Pedersen et al., 2020), three MPN-patients observed for an extended period of no treatment were presented. All three had previously received treatment with IFN, and hence it is possible that the disease progression of these three patients differ from the progression for patients that had not previously received treatment. This data was used to estimate the growth of the JAK2 allele burden in (Pedersen et al., 2020). This estimate was in turn used to determine parameters of the malignant clone in both the Cancitis model and in the combined model. As such, the specifics of how the malignant clone grows, in both the empirical study and in the model, are very uncertain. Obtaining additional data for the JAK2 allele burden as well as blood-cell counts could result in significant changes to all results described in this thesis. As a “Watch-and-wait” approach is used for MPN-patients in some clinics, it is possible that such data does in fact exist, however, we did not manage to obtain such data for our study.

12.3 Relating data and models

The Cancitis model and the combined model were related and fitted to patient-data, in sections 6 and 11 respectively. This granted insight about not only the applicability of the models but also about the IFN-response of patients in the DALIAH trial.

The DALIAH trial included criteria for dose changes, leading to differences in IFN-dose and -timing for the patients considered. While the empirical modelling of data did not explicitly consider the IFN-dose, it was included in the mechanism-based modelling to correctly interpret how the treatment affects biological mechanisms. As dose- and treatment-timing varied greatly,

we first simplified doses as a daily average dose of IFN. We assumed a constant uptake and clearance of IFN. Initial tests showed that considering a daily average rather than e.g. weekly administration made little difference in model behaviour. Through a simple pharmacokinetic model of IFN blood-concentration, we modelled parameter-perturbations that increased with increasing blood-concentration of IFN. This allowed for effects of treatment where changes in dose directly affect model dynamics and allowed for capturing specific dynamics such as e.g. patients where temporary treatment cessation or reduction of dose lead to increasing blood-cell counts. For both models, the effect of treatment always scaled with the dose. Realistically, it is possible that a certain threshold of IFN must be attained before a notable effect occurs. Similarly, patients could also have an upper bound for how significant the effect of IFN-treatment is. We hypothesize that this is the case for some patients observed in the DALIAH trial data. For a subgroup of patients, the JAK2 allele burden decreased in a way that appeared independent of changes in IFN dose, suggesting that increasing the dose does not increase the effect for these patients. Some of these responses were captured very well by the empirical modelling described in (Pedersen et al., 2020), while the dose-dependent responses of the mechanism-based models suggest that the dose must be increased to see the decrease in the relative frequency of the malignant cells. Our approach with a dose-dependence without upper or lower boundaries on the effect is the simplest possible model for the effect of IFN. However, for future work on investigating the nature of how IFN-dose affects model parameters, and in turn model dynamics, it should first be determined whether considering such thresholds is appropriate.

Fitting the Cancitis model to the DALIAH trial data suggested a possible hypothesis for the IFN-responses observed in data, namely a dose-dependent increase in the apoptosis rate of malignant stem cells, \hat{d}_{y_0} . For a large number of patients, this resulted in good agreement between the Cancitis model and both the kinetics of the JAK2 allele burden and of the mature blood-cells. For approximately two thirds of the patients, a daily average IFN dose of $5\mu\text{g}$ was determined to be sufficient for attaining a decline in the JAK2 allele burden, albeit possibly very slowly. This suggests a biological interpretation that, for most patients, $5\mu\text{g}$ average daily IFN results in a more than 1.8-fold increase in the apoptosis rate of malignant stem cells. While more complex effects of IFN could allow for a more detailed description of how IFN affects MPN-patients, the modelled increase of the apoptosis rate provides a first approximation of a mechanistic explanation, with patient-specific expressions of dose-responses. We observed that most patients experienced a decline of thrombocytes and leukocytes immediately after treatment initiation, occurring faster than the decline of the JAK2 allele burden. Such difference between the time-scales of the decline of blood-cell counts and the decline of the JAK2 allele burden was not captured by the fits of the Cancitis model to data. By fitting the combined Cancitis-Niche model to data, we determined that this behaviour could be captured by letting IFN affect the parameter ρ_L as well as parameters ω_H and ω_L , related to self-renewal of malignant HSC and the cell-count amplification from progenitors to mature cells, respectively. One hypothetical interpretation of such changes is that IFN affects differentiation of malignant stem cells, instead of, or in addition to, affecting the apoptosis rate of malignant stem cells. A decrease of ω_H and ω_L could also be suggestive of increased apoptosis- and/or differentiation rates of progenitor cells. Hence our results suggest that the effect of IFN is two-fold: A long-term effect on malignant stem cells, either through increased apoptosis or differentiation, as well as an effect on the production of mature blood-cells, occurring on a faster time-scale than the effect on stem cells.

Comparing the models to data of individual patients allowed us to determine a lower threshold for IFN-doses that would lead to long-term treatment of the given patient if the treatment was maintained. This was done by determining the local stability of the steady states of the models for a given dose. If a malignancy-free steady state was determined to be the only stable steady state, we interpreted the given dose as sufficient for successful treatment. For most patients, both

models suggested that doses with a daily average $5\mu\text{g}$ of IFN allows for long-term treatment. This is similar to the initial doses of used in the DALIAH trial: $30\mu\text{g}$ weekly for *PegIntron* (Daily average of $4.3\mu\text{g}$) and $45\mu\text{g}$ weekly for *Pegasys* (Daily average of $6.4\mu\text{g}$). Note however that the models suggest that larger doses results in a faster decline of the malignant clone and thus faster treatment. For some patients, significantly higher doses of IFN were determined to be necessary for successful treatment. Identifying such patients early could be a clinical goal, allowing the clinicians to make changes to treatment, either to higher doses of IFN if tolerable or to alternative drugs. This requires sufficient data within short time following treatment initiation to be available. Such data would however also be beneficial for purposes of model calibration and for future identification of MPN diagnosed patients, and hence there is a need for obtaining additional data within e.g. the first year of treatment.

For all patients considered from the DALIAH trial, the first measurement of the JAK2 allele burden was made four months after inclusion. From the baseline-measurement to the first post-baseline measurement, some patient experienced a sustained JAK2 allele burden, or even in a few cases, an increase in JAK2 allele burden. This was the reasoning behind considering the bi-exponential decay described in section 3 and (Pedersen et al., 2020). This four-month gap in data results in great uncertainty about the JAK2 kinetics immediately after initiation of IFN-treatment. As illustrated in figure 6.7 and in Electronic Supplementary Material 3, the first four months could feature a significantly increased JAK2 allele burden due to effects on stem cell differentiation. Similarly, other hypothetical effects of IFN could be considered in both the Cancitis model and the combined model, resulting in vastly different dynamics of the JAK2 allele burden in the time between data-points. Determining the best hypothesis for the effect of IFN on the JAK2 allele burden is therefore highly dependent on the time-resolution at which the JAK2 allele burden is measured. This substantiates the need for additional data for the JAK2 allele burden, particularly during the first year of IFN treatment.

12.4 Future aims of mathematical modelling of MPN

Mathematical modelling of MPN shows great promise for the future, both as a prognostic tool and to further our understanding of the diseases on a conceptual level. While the findings discussed above can be useful on their own, much can be done to further develop mathematical modelling of MPN. We here suggest possible directions of research and model development that, in our opinion, would be most appropriate and beneficial.

As discussed above, heterogeneity of patient-responses is a significant limitation of the present study. Most of the uncertainty observed in fitted parameters is believed to arise because of heterogeneity of patients. If patients were to be grouped in clear groups where a similar response is expected, modelling their IFN-response could allow for a clearer interpretation of the effect of IFN. Such groupings could, as mentioned, be determined through detailed analysis of baseline measurements, or through careful calibration to patient-data based on an initial period of data with a high time-resolution. There is an urgent need for classification of such groups of patient-responses, and our other suggestions for future research rely on good classification.

The field of sensitivity analysis describes a wide range of methods that can be used, in broad terms, to determine the relation between model inputs, e.g. parameters and data used for calibration, and model output, e.g. patient prognosis or optimal dosing. For the present purposes, sensitivity analysis can be used to determine how important parameters or particular features of a model are, in terms of model agreement with data. Similarly it can be used to investigate how important certain effects of treatment are in improving the health of the patient. When comparing model outputs to data, insensitive parameters can be indicative of a overparametrized

model, suggesting that a model reduction is appropriate, or that additional data should be obtained. In this thesis, we did not analyse parameter sensitivity, neither for the niche-model or for the combined model. Preliminary sensitivity analysis of the Cancitis model was included in the original proposal of the model (Andersen et al., 2017). One challenge of sensitivity analysis is to decide which model output to compare to what data. As model calibration of the treatment-free disease progression was based on a very limited set of data, sensitivity analysis of the parameters related to disease progression would be difficult as a result. This in turn affects sensitivity analysis of treatment-related parameters as well. Hence, as already discussed above, obtaining more data for disease progression is important, to guarantee the robustness of our results and to determine the sensitivity of model parameters. The methods of the related field of uncertainty quantification could also allow for determining how uncertainty of data relate to uncertainty of parameter values. In the work presented, all estimated parameters were point estimates, that is, singular values that minimize the difference between model and data in a least-square sense. Important tools from uncertainty quantification such as the Markov chain Monte Carlo method of delayed rejection adaptive Metropolis sampling (DRAM) could be used to determine the distribution of model parameters. This would more clearly describe how IFN affects the related biological processes for any given patient: If a treatment-perturbed parameters is very uncertain, the biological interpretation of the effect of IFN is likewise uncertain. Similarly, uncertainty of parameters would allow for a way to communicate to clinicians the probability that a certain patient is responding well to treatment against the probability that the patient is not responding. For patients where model calibration and uncertainty quantification suggests a high probability for a treatment-response with a future increase in disease burden, clinicians can make changes to treatment and use the formulation of the uncertainty to determine their trust in the predictions of the mathematical models.

On a population level, uncertainty of the estimated parameters can also be beneficial to understand how different treatment-responses relate to one another. Indeed, grouping of patients based on the ranges of parameter uncertainty could solve the patient-response classification-challenge discussed above. We considered the distribution of fit-parameters for a cohort of the 20 patients that the model was determined to most accurately agree with. The fit-parameters of these patients were found to be approximately log-normally distributed, allowing us to describe a two-dimensional distribution of fit-parameters on a population level. As a proof-of-concept, we simulated 1000 virtual patient responses, where the fit-parameters of the virtual patients were randomly chosen from the population-level distribution. By having the virtual patients follow the same IFN-dose schedule as a given patient and start from the same baseline blood-cell measurements, we were able to simulate 1000 different outcomes of the same treatment. The results were shown in Electronic Supplementary Material 5. For most of the 20 patients considered, the patient data was found to fall within the distribution of the virtual patients, suggesting that the given patient would be accurately captured by the model using only the population-level distribution of parameters and measurement at diagnosis. This illustrates that using only information available at diagnosis, the population-level relation between IFN-dose and response in the model can accurately predict how changes in dose will affect the leukocyte-counts, thrombocyte-counts and the JAK2 allele burden. Collecting additional data and calibrating the model to the response of a particular patient ad-hoc is hypothesized to improve such predictions. We emphasize that these results should only be considered proof-of-concept. The virtual patient responses were based on the response of the 20 patients that the model was determined to fit best with, out of the 63 patient considered. While some of the excluded 43 had responses that also agreed with the virtual patients responses (not shown), additional work should be done to ensure that the model more accurately agrees with more types of patient-responses. In addition, not all of the 20 patients were in perfect agreement with the virtual patient responses, and hence

further investigation must be made before model predictions using a cohort of virtual patient can be used with confidence e.g. in the clinic.

We considered the treatment-response of an idealized PV-diagnosed patients, by simulating virtual patients as above, with baseline values corresponding to the average baseline values of patients diagnosed with PV. The dose was maintained at $5\mu\text{g}$ average daily IFN, similar to the starting dose used in the DALIAH trial. The simulations suggested that more than half of patients would not attain thrombocyte-counts below $400 \cdot 10^3 (\mu\text{L})^{-1}$ until after five months of treatment. In the DALIAH trial, heightened thrombocyte-counts four months after initiation of treatment was a criteria for increasing the IFN dose. Such an increase after four months was observed in a significant proportion of patients, such as e.g. patient "P198" shown in figure 11.2b and patient "P086" shown in figure 6.6. Our results suggest that increasing the IFN dose is necessary if a rapid decline of the thrombocyte-counts is desired. Whether such higher initial doses of IFN introduce a risk of drug toxicity was not part of our considerations, however the doses used in the DALIAH trial are considered low-dose IFN. A possible suggestion could thus be the use of initial low-dose IFN to determine possible toxicity, before increasing the dose to attain the clinical criteria of rapid decline in blood-cells.

Mathematical models can be used to predict how changes in treatment affects individual patient. We showed two particular types of model predictions. In addition to the discussion above about the virtual patient responses based on baseline measurements of an individual patient, figure 6.5 displayed the predicted model behaviour that would arise if treatment was ceased at a given point in time, on a patient-specific level using the Cancitis model. We believe that both types of predictions, i.e. pre-treatment initiation and illustrations of the effect of treatment cessation, could be useful in a clinic setting, to help clinicians communicate important ideas to the patient about e.g. the reasoning behind changes in dosing. It is also possible that visual representations of disease marker predictions could have a positive psychological effect for the patient. The predicting of individualized patient-responses is discussed on a conceptual level by Brady and Enderling (2019), highlighting the importance of distinguishing between models that are of academic interest and models that allow for robust translation of predictions into the clinic. The predictions of the models discussed in this thesis are shrouded in great uncertainty due to the limited data used to calibrate parameters. While many of the improvements and further developments discussed above would result in a more reliable model, it is instructive to relate the work presented in this thesis with a six-step framework for reliable predictions presented by Brady and Enderling (2019). The first step consists of identifying a biomarker to focus on. For our work, primary focus was given to the JAK2 allele burden and to a lesser degree also thrombocyte and leukocyte counts. In future work, a measure of the immune system load could also be considered. In the work of (Andersen et al., 2017), the Cancitis model was related to specific cytokines (e.g. *IL-1 β* , *IL-6* and *IL-10*), considered to be indicative of immune system load. These cytokines, or other biomarkers, could similarly be considered for the combined Cancitis-Niche model. Having decided on one or more biomarkers, the second step consists of developing mechanistic models that describe the biomarker(s). The Cancitis model and the combined Cancitis-Niche model, both describe mature cell counts. The interpretation of specific cell-lines (e.g. thrombocytes or leukocytes) as a scaling of the mature cell count in the model allows us to relate the model output to the biomarker. Similarly, we interpreted the relative frequency of malignant mature cells in the model as indicative of the JAK2 allele burden. Calibrating the model with existing data, the third step in the pipeline described by Brady and Enderling (2019), was the primary purpose of the work described in sections 6 and 11. While the model themselves were previously validated with existing data from the literature (In (Andersen et al., 2017) and (Pedersen et al., nd1), for the Cancitis model and the Niche-model respectively), this validation was not specific to the MPN diseases. In addition, treatment with IFN was not considered in previous validation of the

Part VI Discussion and concluding remarks

models. The model calibration to data provided population-level estimates that were difficult to evaluate the reliability of. The heterogeneity of patient-responses resulted in a relation between model input (IFN-treatment timing) and model output (the changes in the chosen biomarkers) that differed greatly from patient to patient. Again, classification of patient could allow for separate calibration to different populations, leading to improved agreement between patient-data and model predictions. The remaining steps described by Brady and Enderling (2019) consists of validating the model with untrained data, evaluating the predictive performance of the model and then finally simulating and predicting novel or optimized treatment schemes. While different methods such as e.g. leave-one-out or k-fold cross validation could be used to validate the model on the available data, making predictions for a completely separate set of patient-data is preferable.

In conclusion, our work suggests a path toward mathematical modelling of MPN-treatment with IFN; (1) Additional data should be considered and included in the mathematical analysis. This would allow for statistically significant criteria for distinguishing between patients groupings, both preceding and during treatment. For future patients, collecting data on a higher time-resolution would be beneficial, particularly for the JAK2 allele burden. (2) The models must be calibrated independently to patient data for different patient classifications, and if necessary, model modifications must be considered to take into account possible differences between classifications. (3) Validation of new data (or cross-validation within the same data-set) must be carried out and the performance of the model predictions must be evaluated.

While this appears a tremendous task, it is not unlikely that considering additional patient-data already collected within the DALIAH trial (e.g. cytokine data, patient age, etc.), such patient classification could be possible, and translational predictions could be made using the models presented here, albeit on a possibly uncertain basis due to the relative low number of patients that each classification would contain. In addition, accurate classifications of patients is useful on its own, as it could lead to improvement of MPN diagnostic criteria and possibly new diagnoses. Hence, patient classification is an important goal by itself, and should be the focus of future work, both mathematical and otherwise.

Clarke and Fisher (2020) discuss the concept of executable models to help guide clinical decisions. We believe that our work is a step in this direction and that personalized mathematical modelling of individual MPN-diagnosed patients is a realistic goal to be reached in the near future. Such mathematical modelling could be a useful tool for clinicians and the inclusion into clinical practice could ultimately benefit the lives of patients suffering from MPN.

Bibliography

- Abkowitz, J. L., Catlin, S. N., and Guttorp, P. (1996). Evidence that hematopoiesis may be a stochastic process in vivo. *Nature Medicine*, 2(2):190–197.
- Abkowitz, J. L., Catlin, S. N., McCallie, M. T., and Guttorp, P. (2002). Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood*, 100(7):2665–2667.
- Altrock, P. M., Liu, L. L., and Michor, F. (2015). The mathematics of cancer: Integrating quantitative models. *Nature Reviews Cancer*, 15(12):730–745.
- Andersen, M., Sajid, Z., Pedersen, R. K., Gudmand-Hoeyer, J., Ellervik, C., Skov, V., Kjær, L., Pallisgaard, N., Kruse, T. A., Thomassen, M., Troelsen, J., Hasselbalch, H. C., and Ottesen, J. T. (2017). Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. *PLOS ONE*, 12(8).
- Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M., and Vardiman, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, 127(20):2391–405.
- Ashcroft, P., Manz, M. G., and Bonhoeffer, S. (2017). Clonal dominance and transplantation dynamics in hematopoietic stem cell compartments. *PLOS Computational Biology*, 13(10).
- Austin, R. J., Straube, J., Bruedigam, C., Pali, G., Jacquelin, S., Vu, T., Green, J., Gräsel, J., Lansink, L., Cooper, L., Lee, S. J., Chen, N. T., Lee, C. W., Haque, A., Heidel, F. H., D’Andrea, R., Hill, G. R., Mullally, A., Milsom, M. D., Bywater, M., and Lane, S. W. (2020). Distinct effects of ruxolitinib and interferon-alpha on murine JAK2V617F myeloproliferative neoplasm hematopoietic stem cell populations. *Leukemia*, 34(4):1075–1089.
- Bangsgaard, K. O., Andersen, M., Skov, V., Kjær, L., Hasselbalch, H. C., and Ottesen, J. T. (n.d.). Dynamics of competing heterogeneous clones in blood cancers explains multiple observations - a mathematical modeling approach. *In review*.
- Barosi, G., Birgegard, G., Finazzi, G., Griesshammer, M., Harrison, C., Hasselbalch, H. C., Kiadjian, J.-J., Lengfelder, E., McMullin, M. F., Passamonti, F., Reilly, J. T., Vannucchi, A. M., and Barbui, T. (2009). Response criteria for essential thrombocythemia and polycythemia vera: result of a European LeukemiaNet consensus conference. *Blood*, 113(20):4829–4833.
- Barosi, G., Mesa, R., Finazzi, G., Harrison, C., Kiladjian, J.-J., Lengfelder, E., McMullin, M. F., Passamonti, F., Vannucchi, A. M., Besses, C., Gisslinger, H., Samuelsson, J., Verstovsek, S., Hoffman, R., Pardanani, A., Cervantes, F., Tefferi, A., and Barbui, T. (2013). Revised response criteria for polycythemia vera and essential thrombocythemia: an ELN and IWG-MRT consensus project. *Blood*, 121(23):4778–4781.

Part VI Discussion and concluding remarks

- Becker, N. B., Günther, M., Li, C., Jolly, A., and Höfer, T. (2019). Stem cell homeostasis by integral feedback through the niche. *Journal of Theoretical Biology*, 481:100–109.
- Besse, A., Lepoutre, T., and Bernard, S. (2017). Long-term treatment effects in chronic myeloid leukemia. *Journal of Mathematical Biology*, 75(3):733–758.
- Bhattacharya, D., Czechowicz, A., Ooi, A. L., Rossi, D. J., Bryder, D., and Weissman, I. L. (2009). Niche recycling through division-independent egress of hematopoietic stem cells. *The Journal of Experimental Medicine*, 206(12):2837–2850.
- Brady, R. and Enderling, H. (2019). Mathematical Models of Cancer: When to Predict Novel Therapies, and When Not to. *Bulletin of Mathematical Biology*, 81(10):3722–3731.
- Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P., and Abkowitz, J. L. (2011). The replication rate of human hematopoietic stem cells in vivo. *Blood*, 117(17):4460–4466.
- Clapp, G. and Levy, D. (2015). A review of mathematical models for leukemia and lymphoma. *Drug Discovery Today: Disease Models*, 16(1):1–6.
- Clapp, G. D., Lepoutre, T., El Cheikh, R., Bernard, S., Ruby, J., Labussiere-Wallet, H., Nicolini, F. E., and Levy, D. (2015). Implication of the Autologous Immune System in BCR-ABL Transcript Variations in Chronic Myelogenous Leukemia Patients Treated with Imatinib. *Cancer Research*, 75(19):4053–4062.
- Clarke, M. A. and Fisher, J. (2020). Executable cancer models: successes and challenges. *Nature Reviews Cancer*, 20(6):343–354.
- Colijn, C. and Mackey, M. C. (2005a). A mathematical model of hematopoiesis: II. Cyclical neutropenia. *Journal of Theoretical Biology*, 237(2):133–146.
- Colijn, C. and Mackey, M. C. (2005b). A mathematical model of hematopoiesis—I. Periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, 237(2):117–132.
- Colotta, F., Allavena, P., Sica, A., Garlanda, C., and Mantovani, A. (2009). Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*, 30(7):1073–1081.
- Cordua, S., Kjaer, L., Skov, V., Pallisgaard, N., Hasselbalch, H. C., and Ellervik, C. (2019). Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. *Blood*, 134(5):469–479.
- Dale, D. C. and Mackey, M. C. (2015). Understanding, Treating and Avoiding Hematological Disease: Better Medicine Through Mathematics? *Bulletin of Mathematical Biology*, 77(5):739–757.
- Dam, M. J. B., Andersen, M., Pedersen, R. K., Kjær, L., Skov, V., Hasselbalch, H. C., and Ottesen, J. T. (n.d.). Data-Driven Analysis of the Kinetics of the JAK2V617F Allele Burden and Blood Cell Counts During Hydroxyurea Treatment of Patients with Polycythemia Vera, Essential Thrombocythemia and Primary Myelofibrosis. *Submitted*.
- Dameshek, W. (1951). Some speculations on the myeloproliferative syndromes. *Blood*, 6(4):372–5.
- Dick, J. E. (2008). Stem cell concepts renew cancer research. *Blood*, 112(13):4793–4807.

Bibliography

- Dingli, D. and Michor, F. (2006). Successful Therapy Must Eradicate Cancer Stem Cells. *Stem Cells*, 24(12):2603–2610.
- Dingli, D. and Pacheco, J. M. (2010). Modeling the architecture and dynamics of hematopoiesis. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(2):235–244.
- Enderling, H., Kim, S., and Pilon-Thomas, S. (2018). The accelerating quest for optimal radiation and immunotherapy combinations for local and systemic tumor control. *Therapeutic Radiology and Oncology*, 2:33–33.
- Estey, E. and Döhner, H. (2006). Acute myeloid leukaemia. *The Lancet*, 368(9550):1894–1907.
- Frederiksen, H., Farkas, D. K., Christiansen, C. F., Hasselbalch, H. C., and Sorensen, H. T. (2011). Chronic myeloproliferative neoplasms and subsequent cancer risk: a Danish population-based cohort study. *Blood*, 118(25):6515–6520.
- Gatenby, R. A. and Brown, J. S. (2020). Integrating evolutionary dynamics into cancer therapy. *Nature Reviews Clinical Oncology*.
- Gerlee, P. and Anderson, A. R. A. (2015). The evolution of carrying capacity in constrained and expanding tumour cell populations. *Physical Biology*, 12(5):056001.
- Glauche, I., Horn, K., Horn, M., Thielecke, L., Essers, M. A. G., Trumpp, A., and Roeder, I. (2012). Therapy of chronic myeloid leukaemia can benefit from the activation of stem cells: Simulation studies of different treatment combinations. *British Journal of Cancer*, 106(11):1742–1752.
- Godfrey, A. L., Chen, E., Pagano, F., Ortmann, C. A., Silber, Y., Bellosillo, B., Guglielmelli, P., Harrison, C. N., Reilly, J. T., Stegelmann, F., Bijou, F., Lippert, E., McMullin, M. F., Boiron, J.-M., Döhner, K., Vannucchi, A. M., Besses, C., Campbell, P. J., and Green, A. R. (2012). JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. *Blood*, 120(13):2704–2707.
- Goldman, J. M. and Melo, J. V. (2003). Chronic Myeloid Leukemia — Advances in Biology and New Approaches to Treatment. *New England Journal of Medicine*, 349(15):1451–1464.
- Hasselbalch, H. C. (2011). A new era for IFN- α in the treatment of Philadelphia-negative chronic myeloproliferative neoplasms. *Expert Review of Hematology*, 4(6):637–655.
- Hasselbalch, H. C. and Bjørn, M. E. (2015). MPNs as Inflammatory Diseases: The Evidence, Consequences, and Perspectives. *Mediators of Inflammation*, 2015.
- Hasselbalch, H. C. and Holmström, M. O. (2018). Perspectives on interferon-alpha in the treatment of Polycythemia vera and related Myeloproliferative neoplasms: minimal residual disease and cure? *Seminars in Immunopathology*.
- Holyoake, T. L. and Vetrie, D. (2017). The chronic myeloid leukemia stem cell: stemming the tide of persistence. *Blood*, 129(12):1595–1606.
- Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N., Fukata, M., Miyamoto, T., Lyons, B., Ohshima, K., Uchida, N., Taniguchi, S., Ohara, O., Akashi, K., Harada, M., and Shultz, L. D. (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nature Biotechnology*, 25(11):1315.

Part VI Discussion and concluding remarks

- Jaiswal, S. and Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science*, 366(6465).
- Jarrett, A. M., Faghihi, D., Hormuth, D. A., Lima, E. A. B. F., Virostko, J., Biros, G., Patt, D., and Yankeelov, T. E. (2020). Optimal Control Theory for Personalized Therapeutic Regimens in Oncology: Background, History, Challenges, and Opportunities. *Journal of Clinical Medicine*, 9(5):1314.
- Jilkine, A. and Gutenkunst, R. N. (2014). Effect of Dedifferentiation on Time to Mutation Acquisition in Stem Cell-Driven Cancers. *PLoS Computational Biology*, 10(3):e1003481.
- Kim, P. S., Lee, P. P., and Levy, D. (2008a). A PDE model for imatinib-treated chronic myelogenous leukemia. *Bulletin of Mathematical Biology*, 70(7):1994–2016.
- Kim, P. S., Lee, P. P., and Levy, D. (2008b). Modeling imatinib-treated chronic myelogenous leukemia: Reducing the complexity of agent-based models. *Bulletin of Mathematical Biology*, 70(3):728–744.
- Kjær, L. (2020). Clonal Hematopoiesis and Mutations of Myeloproliferative Neoplasms. *Cancers*, 12(8):2100.
- Komarova, N. L., Burger, J. A., and Wodarz, D. (2014). Evolution of ibrutinib resistance in chronic lymphocytic leukemia (CLL). *Proceedings of the National Academy of Sciences*, 111(38):13906–13911.
- Komarova, N. L. and Wodarz, D. (2005). Drug resistance in cancer: Principles of emergence and prevention. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9714–9719.
- Komarova, N. L. and Wodarz, D. (2007). Effect of cellular quiescence on the success of targeted CML therapy. *PLoS ONE*, 2(10).
- Kumar, S. and Geiger, H. (2017). HSC Niche Biology and HSC Expansion Ex Vivo. *Trends in Molecular Medicine*, 23(9):799–819.
- Larsen, T. S., Pallisgaard, N., Møller, M. B., and Hasselbalch, H. C. (2007). The JAK2 V617F allele burden in essential thrombocythemia, polycythemia vera and primary myelofibrosis – impact on disease phenotype. *European Journal of Haematology*, 79(6):508–515.
- Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J. P., Martincorena, I., Anderson, E., O'Neill, L., Stratton, M. R., Laurenti, E., Green, A. R., Kent, D. G., and Campbell, P. J. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*, 561(7724):473–478.
- Lewin, T. D., Byrne, H. M., Maini, P. K., Caudell, J. J., Moros, E. G., and Enderling, H. (2020). The importance of dead material within a tumour on the dynamics in response to radiotherapy. *Physics in Medicine and Biology*, 65(1).
- Mackey, M. C. (1978). Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis. *Blood*, 51(5):941–956.
- Mackey, M. C. (2001). Cell kinetic status of haematopoietic stem cells. *Cell Proliferation*, 34(2):71–83.

- Mackey, M. C. and Maini, P. K. (2015). What Has Mathematics Done for Biology? *Bulletin of Mathematical Biology*, 77(5):735–738.
- Marciniak-Czochra, A., Stiehl, T., Ho, A. D., Jäger, W., and Wagner, W. (2009). Modeling of Asymmetric Cell Division in Hematopoietic Stem Cells—Regulation of Self-Renewal Is Essential for Efficient Repopulation. *Stem Cells and Development*, 18(3):377–386.
- Michor, F., Hughes, T. P., Iwasa, Y., Branford, S., Shah, N. P., Sawyers, C. L., and Nowak, M. A. (2005). Dynamics of chronic myeloid leukaemia. *Nature*, 435(7046):1267–1270.
- Mullally, A., Bruedigam, C., Poveromo, L., Heidel, F. H., Purdon, A., Vu, T., Austin, R., Heckl, D., Breyfogle, L. J., Kuhn, C. P., Kalaitzidis, D., Armstrong, S. A., Williams, D. A., Hill, G. R., Ebert, B. L., and Lane, S. W. (2013). Depletion of Jak2V617F myeloproliferative neoplasm-propagating stem cells by interferon- α in a murine model of polycythaemia vera. *Blood*, 121(18):3692–702.
- Nielsen, C., Bojesen, S. E., Nordestgaard, B. G., Kofoed, K. F., and Birgens, H. S. (2014). JAK2V617F somatic mutation in the general population: Myeloproliferative neoplasm development and progression rate. *Haematologica*, 99(9):1448–1455.
- Oguro, H., Ding, L., and Morrison, S. J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*, 13(1):102–116.
- Ottesen, J. T., Pedersen, R. K., Dam, M. J. B., Knudsen, T. A., Skov, V., Kjær, L., and Andersen, M. (2020). Mathematical Modeling of MPNs Offers Understanding and Decision Support for Personalized Treatment. *Cancers*, 12(8):2119.
- Ottesen, J. T., Pedersen, R. K., Sajid, Z., Gudmand-Hoeyer, J., Bangsgaard, K. O., Skov, V., Kjær, L., Knudsen, T. A., Pallisgaard, N., Hasselbalch, H. C., and Andersen, M. (2019). Bridging blood cancers and inflammation: The reduced Cancitis model. *Journal of Theoretical Biology*, 465:90–108.
- Park, D. S., Akuffo, A. A., Muench, D. E., Grimes, H. L., Epling-Burnette, P. K., Maini, P. K., Anderson, A. R. A., and Bonsall, M. B. (2019). Clonal hematopoiesis of indeterminate potential and its impact on patient trajectories after stem cell transplantation. *PLOS Computational Biology*, 15(4).
- Pedersen, R. K., Andersen, M., Knudsen, T. A., Sajid, Z., Gudmand-Hoeyer, J., Dam, M. J. B., Skov, V., Kjær, L., Ellervik, C., Larsen, T. S., Hansen, D., Pallisgaard, N., Hasselbalch, H. C., and Ottesen, J. T. (2020). Data-driven analysis of JAK2 V617F kinetics during interferon-alpha2 treatment of patients with polycythaemia vera and related neoplasms. *Cancer Medicine*, 9(6):2039–2051.
- Pedersen, R. K., Andersen, M., Skov, V., Kjær, L., Hasselbalch, H. C., Ottesen, J. T., and Stiehl, T. (n.d.,1). Combining stem cell mobilization with preconditioning: Insights about competition in the stem cell niche from mathematical modelling. *Submitted*.
- Pedersen, R. K., Andersen, M., Stiehl, T., and Ottesen, J. T. (n.d.,2). Mathematical Modelling of the Hematopoietic Stem Cell–Niche System: Clonal Dominance based on Stem Cell Fitness. *Submitted*.
- Pinho, S. and Frenette, P. S. (2019). Haematopoietic stem cell activity and interactions with the niche. *Nature reviews. Molecular cell biology*, 20(5):303–320.

Part VI Discussion and concluding remarks

- Poleszczuk, J. and Enderling, H. (2018). The Optimal Radiation Dose to Induce Robust Systemic Anti-Tumor Immunity. *International Journal of Molecular Sciences*, 19(11):3377.
- Quintás-Cardama, A., Abdel-Wahab, O., Mansouri, T., Kilpivaara, O., Cortes, J., Roupie, A. L., Zhang, S. J., Harris, D., Estrov, Z., Kantarjian, H., Levine, R. L., and Verstovsek, S. (2013). Molecular analysis of patients with polycythemia vera or essential thrombocythemia receiving pegylated interferon α -2a. *Blood*, 122(6):893–901.
- Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859):105–111.
- Rockne, R. C., Hawkins-Daarud, A., Swanson, K. R., Sluka, J. P., Glazier, J. A., Macklin, P., Hormuth, D. A., Jarrett, A. M., Lima, E. A. B. F., Tinsley Oden, J., Biros, G., Yankeelov, T. E., Curtius, K., Al Bakir, I., Wodarz, D., Komarova, N., Aparicio, L., Bordyuh, M., Rabadan, R., Finley, S. D., Enderling, H., Caudell, J., Moros, E. G., Anderson, A. R. A., Gatenby, R. A., Kaznatcheev, A., Jeavons, P., Krishnan, N., Pelesko, J., Wadhwa, R. R., Yoon, N., Nichol, D., Marusyk, A., Hinczewski, M., and Scott, J. G. (2019). The 2019 mathematical oncology roadmap. *Physical Biology*, 16(4):041005.
- Roeder, I., Horn, M., Glauche, I., Hochhaus, A., Mueller, M. C., and Loeffler, M. (2006). Dynamic modeling of imatinib-treated chronic myeloid leukemia: Functional insights and clinical implications. *Nature Medicine*, 12(10):1181–1184.
- Roeder, I., Kamminga, L. M., Braesel, K., Dontje, B., De Haan, G., and Loeffler, M. (2005). Competitive clonal hematopoiesis in mouse chimeras explained by a stochastic model of stem cell organization. *Blood*, 105(2):609–616.
- Saito, T., Iida, S., and Kawanishi, T. (2012). Population Pharmacokinetic-Pharmacodynamic Modeling and Simulation of Platelet Decrease Induced by Peg-interferon-alpha 2a. *Drug Metabolism and Pharmacokinetics*, 27(6):614–620.
- Sajid, Z., Andersen, M., and T. Ottesen, J. (2019). Mathematical analysis of the Cancitis model and the role of inflammation in blood cancer progression. *Mathematical Biosciences and Engineering*, 16(6):8268–8289.
- Simonsson, B., Gedde-Dahl, T., Markevärn, B., Remes, K., Stentoft, J., Almqvist, A., Björeman, M., Floegård, M., Koskenvesa, P., Lindblom, A., Malm, C., Mustjoki, S., Myhr-Eriksson, K., Ohm, L., Räsänen, A., Sinisalo, M., Själander, A., Strömberg, U., Bjerrum, O. W., Ehrencrona, H., Gruber, F., Kairisto, V., Olsson, K., Sandin, F., Nagler, A., Nielsen, J. L., Hjorth-Hansen, H., and Porkka, K. (2011). Combination of pegylated IFN- α 2b with imatinib increases molecular response rates in patients with low- or intermediate-risk chronic myeloid leukemia. *Blood*, 118(12):3228–3235.
- Singh, H. and Singh, I. (2018). *Fundamentals of Medical Physiology*. Elsevier India, Philadelphia.
- Skov, V., Larsen, T. S., Thomassen, M., Riley, C., Jensen, M. K., Bjerrum, O. W., Kruse, T. A., and Hasselbalch, H. C. (2010). Gene Expression Profiling with Principal Component Analysis Depicts the Biological Continuum From Essential Thrombocythemia Over Polycythemia Vera to Myelofibrosis. *Blood*, 116(21):4115–4115.
- Stauffer Larsen, T., Iversen, K. F., Hansen, E., Mathiasen, A. B., Marcher, C., Frederiksen, M., Larsen, H., Helleberg, I., Riley, C. H., Bjerrum, O. W., Rønnov-Jessen, D., Møller, M. B., de Stricker, K., Vestergaard, H., and Hasselbalch, H. C. (2013). Long term molecular responses

- in a cohort of Danish patients with essential thrombocythemia, polycythemia vera and myelofibrosis treated with recombinant interferon alpha. *Leukemia Research*, 37(9):1041–1045.
- Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P., and Ebert, B. L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*, 126(1):9–16.
- Stiehl, T., Baran, N., Ho, A. D., and Marciniak-Czochra, A. (2015). Cell division patterns in acute myeloid leukemia stem-like cells determine clinical course: A model to predict patient survival. *Cancer Research*, 75(6):940–949.
- Stiehl, T., Ho, A. D., and Marciniak-Czochra, A. (2014). The impact of CD34+ cell dose on engraftment after SCTs: Personalized estimates based on mathematical modeling. *Bone Marrow Transplantation*, 49(1):30–37.
- Stiehl, T., Ho, A. D., and Marciniak-Czochra, A. (2018). Mathematical modeling of the impact of cytokine response of acute myeloid leukemia cells on patient prognosis. *Scientific Reports*, (January):1–11.
- Stiehl, T. and Marciniak-Czochra, A. (2012). Mathematical Modeling of Leukemogenesis and Cancer Stem Cell Dynamics. *Mathematical Modelling of Natural Phenomena*, 7(1):166–202.
- Stiehl, T., Wang, W., Lutz, C., and Marciniak-Czochra, A. (2020). Mathematical modeling provides evidence for niche competition in human AML and serves as a tool to improve risk stratification. *Cancer Research*, canres.0283.2020.
- Takahashi, K., Patel, K. P., Kantarjian, H., Luthra, R., Pierce, S., Cortes, J., and Verstovsek, S. (2013). JAK2 p.V617F detection and allele burden measurement in peripheral blood and bone marrow aspirates in patients with myeloproliferative neoplasms. *Blood*, 122(23):3784–3786.
- Talpaz, M., Mercer, J., and Hehlmann, R. (2015). The interferon-alpha revival in CML. *Annals of Hematology*, 94(S2):195–207.
- Tan, B. T., Park, C. Y., Ailles, L. E., and Weissman, I. L. (2006). The cancer stem cell hypothesis: a work in progress. *Laboratory Investigation*, 86(12):1203–1207.
- Tefferi, A. and Elliott, M. (2007). Thrombosis in Myeloproliferative Disorders: Prevalence, Prognostic Factors, and the Role of Leukocytes and JAK2 V617F. *Seminars in Thrombosis and Hemostasis*, 33(4):313–320.
- Titmarsh, G. J., Duncombe, A. S., McMullin, M. F., O'Rorke, M., Mesa, R., De Vocht, F., Horan, S., Fritschi, L., Clarke, M., and Anderson, L. A. (2014). How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *American Journal of Hematology*, 89(6):581–587.
- Traulsen, A., Lenaerts, T., Pacheco, J. M., and Dingli, D. (2013). On the dynamics of neutral mutations in a mathematical model for a homogeneous stem cell population. *Journal of the Royal Society Interface*, 10(79).
- Traulsen, A., Pacheco, J. M., and Dingli, D. (2010). Reproductive fitness advantage of BCR-ABL expressing leukemia cells. *Cancer Letters*, 294(1):43–48.
- Trumpp, A., Essers, M., and Wilson, A. (2010). Awakening dormant haematopoietic stem cells. *Nature Reviews Immunology*, 10(3):201–209.

Part VI Discussion and concluding remarks

- Vaidya, A. and Kale, V. (2015). Hematopoietic Stem Cells, Their Niche, and the Concept of Co-Culture Systems: A Critical Review. *Journal of stem cells*, 10(1):13–31.
- Vandekerckhove, J. (2008). General simulated annealing algorithm. <https://www.mathworks.com/matlabcentral/fileexchange/10548-general-simulated-annealing-algorithm>, accessed August 31, 2020.
- Walter, D., Lier, A., Geiselhart, A., Thalheimer, F. B., Huntscha, S., Sobotta, M. C., Moehrle, B., Brocks, D., Bayindir, I., Kaschutnig, P., Muedder, K., Klein, C., Jauch, A., Schroeder, T., Geiger, H., Dick, T. P., Holland-Letz, T., Schmezer, P., Lane, S. W., Rieger, M. A., Essers, M. A. G., Williams, D. A., Trumpp, A., and Milsom, M. D. (2015). Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature*, 520(7548):549–552.
- Wang, W., Stiehl, T., Raffel, S., Hoang, V. T., Hoffmann, I., Poisa-Beiro, L., Saeed, B. R., Blume, R., Manta, L., Eckstein, V., Bochtler, T., Wuchter, P., Essers, M., Jauch, A., Trumpp, A., Marciniak-Czochra, A., Ho, A. D., and Lutz, C. (2017). Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia. *Haematologica*, 102(9):1567–1577.
- Watson, C. J., Papula, A. L., Poon, G. Y. P., Wong, W. H., Young, A. L., Druley, T. E., Fisher, D. S., and Blundell, J. R. (2020). The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*, 367(6485):1449–1454.
- Wilson, A. and Trumpp, A. (2006). Bone-marrow haematopoietic-stem-cell niches. *Nature Reviews Immunology*, 6(2):93–106.
- Yang, S.-H., Wang, T.-F., Tsai, H.-H., Lin, T.-Y., Wen, S.-H., and Chen, S.-H. (2009). Pre-harvest hematopoietic progenitor cell counts predict CD34+ cell yields in granulocyte-colony-stimulating factor-mobilized peripheral blood stem cell harvest in healthy donors. *Transfusion*, 50(5):1088–1095.
- Zhang, Y. and Gao, Y. (2016). Novel chemical attempts at ex vivo hematopoietic stem cell expansion. *International Journal of Hematology*, 103(5):519–529.

Part VII

Appendix

Appendix A

Reprints of articles

Data-driven analysis of *JAK2V617F* kinetics during interferon-alpha2 treatment of patients with polycythemia vera and related neoplasms

Rasmus K. Pedersen¹  | Morten Andersen¹  | Trine A. Knudsen²  | Zamra Sajid¹  | Johanne Gudmand-Hoeyer¹  | Marc J. B. Dam¹  | Vibe Skov²  | Lasse Kjær²  | Christina Ellervik^{3,4,5,6}  | Thomas S. Larsen⁷  | Dennis Hansen⁷  | Niels Pallisgaard⁸ | Hans C. Hasselbalch²  | Johnny T. Ottesen¹ 

¹Department of Science and Environment, Roskilde University, Roskilde, Denmark

²Department of Haematology, Zealand University Hospital, Roskilde, Denmark

³Department of Production, Research, and Innovation, Region Zealand, Sorø, Denmark

⁴Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

⁵Department of Pathology, Harvard Medical School, Boston, FL, USA

⁶Department of Laboratory Medicine, Boston Children's Hospital, Boston, FL, USA

⁷Department of Hematology, Odense University Hospital, Odense, Denmark

⁸Department of Surgical Pathology, Zealand University Hospital, Roskilde, Denmark

Correspondence

Johnny T. Ottesen, Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark.
Email: johnny@ruc.dk

Funding information

OUH Frie Forskningsmidler. OUH-Region Sjælland Fælles Forskningspulje. Fonden til Lægevidenskabens Fremme. Ellen og Aage Fausbølls Helsefond af 1975. Swedish Orphan. Den Forskningsfremmende Pulje for Kliniske Professorer ansat i Region Sjælland og ved Institut for Klinisk Medicin, Københavns Universitet. Regions Sjællands Sundhedsvidenskabelige Forskningsfond.

Abstract

Treatment with PEGylated interferon-alpha2 (IFN) of patients with essential thrombocythemia and polycythemia vera induces major molecular remissions with a reduction in the *JAK2V617F* allele burden to undetectable levels in a subset of patients. A favorable response to IFN has been argued to depend upon the tumor burden, implying that institution of treatment with IFN should be as early as possible after the diagnosis. However, evidence for this statement is not available. We present a thorough analysis of unique serial *JAK2V617F* measurements in 66 IFN-treated patients and in 6 untreated patients. Without IFN treatment, the *JAK2V617F* allele burden increased exponentially with a period of doubling of 1.4 year. During monotherapy with IFN, the *JAK2V617F* allele burden decreased mono- or bi-exponentially for 33 responders of which 28 patients satisfied both descriptions. Bi-exponential description improved the fits in 19 cases being associated with late *JAK2V617F* responses. The decay of the *JAK2V617F* allele burden during IFN treatment was estimated to have half-lives of 1.6 year for the monoexponential response and 1.0 year in the long term for the bi-exponential response. In conclusion, through data-driven analysis of the *JAK2V617F* allele burden, we provide novel information regarding the *JAK2V617F* kinetics during IFN-treatment, arguing for early intervention.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

KEY WORDS

early treatment, essential thrombocythemia, interferon-alpha2, JAK2V617F kinetics, myeloproliferative neoplasms, polycythemia vera, primary myelofibrosis

1 | INTRODUCTION

The classic Philadelphia chromosome-negative chronic myeloproliferative neoplasms (MPNs) encompass essential thrombocythemia (ET), polycythemia vera (PV), and primary myelofibrosis (PMF), including early prefibrotic myelofibrosis. These neoplasms arise due to an acquired stem cell insult with ensuing clonal myeloproliferation in the biological continuum from the early cancer stages (ET and PV) to the advanced myelofibrosis stage¹ and ultimately leukemic transformation.^{2,3}

Molecular markers for MPNs include *JAK2V617F*, *CALR*- and *MPL*-mutations. These mutations are the so-called driver mutations whereas additional mutations (e.g., *ASXL1*, *TET2*) are frequently recorded in the more advanced disease stages with severe myelofibrosis.⁴⁻¹² The *JAK2V617F* mutation associates with laboratory (hemoglobin level, leukocyte count, platelet count, CD34+ count, serum lactic dehydrogenase, and in vivo granulocyte and platelet activation) and clinical (pruritus, thrombosis, spleen size) outcomes^{8,13-25} and prognosis.^{13,17} Accordingly, a new concept of these diseases as a biological continuum from ET over PV to PMF has emerged, implying the *JAK2V617F* mutational load to reflect the tumor burden as assessed by a rising leukocyte count and increasing splenomegaly during disease progression toward myelofibrotic and leukemic transformation.²⁶ However, this hypothesis on the biological continuum is still being debated.

Interferon-alpha2 (IFN) has been used in the treatment of MPNs for about 30 years and several studies have convincingly demonstrated that this agent is safe and highly efficacious in normalizing elevated cell counts.²⁷⁻⁵² Indeed, prolonged treatment (about 5 years) may be followed by polyclonal hematopoiesis, normalization of the bone marrow and low-burden *JAK2V617F* in a subset of patients, even being sustained for 2-3 years after discontinuation of IFN.^{36,48}

These highly encouraging results have been the focus of increasing interest, since we may enter a new era with “Minimal Residual Disease” (MRD) as a novel treatment goal.^{36,44,48}

Early treatment to reduce or eradicate the malignant clone is of paramount importance for achievement of MRD or cure in all cancers. However, in MPNs a “watch and

wait” strategy is used in “low-risk” patients allowing the malignant clone to expand with the inherent risk of increasing genomic instability, sub-clone formation, resistance to treatment and disease progression from the early cancer stages (ET and PV) to the advanced metastatic cancer stage—myelofibrosis with bone marrow failure and ultimately leukemic transformation.

In recent years, the “watch and wait” strategy has been challenged by reports demonstrating the potential of IFN to induce MRD in an increasing number of patients.^{36,44,48} Furthermore, these studies also indicate that early treatment with IFN increases the chance of sustained hematological and molecular remissions. However, evidence for this statement is lacking.

As noted above, changes in the *JAK2V617F* allele burden before and during IFN-treatment have been studied extensively, whereas the time-scale kinetics of these changes still remain to be identified and described in detail.

In this study, we predict the *JAK2V617F* kinetics during IFN-treatment through data-driven analysis of serial *JAK2V617F* measurements in MPN patients receiving cytoreductive treatment with IFN only and patients being observed without cytoreduction. Evidence for tumor burden reduction through early intervention with IFN is provided, thereby challenging the “watch and wait” strategy commonly applied to low-risk MPN patients.

2 | METHODS

2.1 | Study design

2.1.1 | Prospective study

Data were obtained from two different study populations. *JAK2V617F* observations during PEGylated r-IFN α (IFN) monotherapy were obtained from 120 patients enrolled in the DALIAH trial (#EudraCT 2011-001919-31), which is an ongoing Danish multicenter prospective randomized open-label phase III clinical trial comparing IFN with hydroxyurea in MPN patients. Enrollment began in February 2012 and was completed in July 2015. Patients are followed for five years. The DALIAH trial was approved by the Danish Regional Science Ethics Committee and by the Danish Medicines Agency.

The clinical characteristics of the patients used in later analysis is shown in Table 1.

TABLE 1 Prospective study. Baseline demographics and clinical characteristics of *JAK2V617F* positive patients from the DALIAH trial randomized to IFN. Only patients with four or more measurements of the *JAK2V617F* allele burden are included

Characteristics	ET (n = 15)	PV (n = 39)	Pre-MF (n = 5)	PMF (n = 7)	Total (n = 66)
IFN type (IFN α -2a/IFN α -2b)	10/5	21/18	3/2	2/5	36/30
Age (y)	53 (43-64)	64 (52-69)	62 (59-65)	64 (51-65)	62 (51-67)
Gender, male	6 (40)	20 (51)	3 (60)	6 (86)	35 (53)
History of major thrombotic event	0 (0)	12 (31)	1 (20)	2 (29)	15 (23)
<i>JAK2V617F</i> allele burden (%)	15 (10-21)	44 (22-62)	35 (19-40)	51 (50-88)	37 (19-51)
Haematocrit (vol%)	44 (39-47)	55 (48-59)	45 (41-45)	49 (37-51)	49 (45-55)
Haemoglobin (mmol/L)	9.0 (8.3-9.7)	11.4 (10.0-12.3)	9.2 (8.6-9.5)	9.6 (7.2-10.2)	9.9 (9.1-11.8)
Platelets ($\times 10^9$ /L)	730 (626-887)	538 (343-670)	681 (667-776)	460 (351-611)	611 (413-739)
White blood cells ($\times 10^9$ /L)	8.9 (7.6-11.9)	9.9 (8.4-13.2)	9.2 (8.6-9.5)	10.8 (5.5-17.4)	9.7 (8.2-12.9)
Plasma lactate dehydrogenase (U/L)	193 (164-210)	229 (199-304)	281 (181-342)	367 (284-643)	222 (189-308)
Splenomegaly (≥ 13 cm by US)	2/9 (22)	11/26 (42)	1/4 (25)	7/7 (100)	21/46 (46)
Disease-related symptoms ^a	8 (53)	24 (62)	1 (20)	3 (43)	36 (55)
Phlebotomy before enrolment	2 (13)	35 (90)	1 (20)	4 (57)	42 (64)
Low-risk disease ^b	10 (67)	13 (33)	1 (20)	3 (43)	27 (41)

Note: Data are median (IQR) and n (%)

^aConstitutional symptoms, microvascular disturbances and pruritus

^bAge ≤ 60 y of age, platelets $\leq 1500 (\times 10^9/\text{L})$ and no prior major thrombosis

2.1.2 | Retrospective study

We retrospectively obtained information on *JAK2V617F* kinetics in six untreated (i.e., no cytoreductive therapy) MPN patients followed at the outpatient clinic at the Department of Haematology, Zealand University Hospital, Denmark. Four patients had previously received cytoreductive therapy with either r-IFN α -2a (Pegasys[®]) or r-IFN α -2b (PegIntron[®]) (n = 2) or monotherapy with both HU and r-IFN α -2a (n = 2) according to standard care, but had discontinued therapy due to intolerance (r-IFN α -2a: n = 2, HU: n = 2) and/or hematologic response in concert with a low *JAK2V617F* allele burden (r-IFN α -2a: n = 1, r-IFN α -2b: n = 1). Patient (A) discontinued treatment due to *JAK2V617F* < 1% for more than 1 year after r-IFN α -2b exposure and patient C discontinued r-IFN α -2a due to complete hematologic response in concert with a low *JAK2V617F* allele burden (6%). Two patients had not received any prior cytoreductive treatment. At the time of inclusion in the study all untreated patients had been off

cytoreductive treatment for at least 0.5 months (median: 1.2 month; range 0.5-5.3 months).

Patients were evaluated for enrollment when attending regular appointments between 1st of May 2018 and 15th of December 2018.

Written informed consent was provided from all patients according to the Declaration of Helsinki.

The clinical characteristics of the patients in the retrospective study are shown in Table 2.

2.2 | MPN diagnosis and eligibility

Eligibility criteria were age ≥ 18 years and a diagnosis of *JAK2V617F* positive Philadelphia chromosome negative MPN according the World Health Organization criteria.⁵³ Patients from the DALIAH trial were all newly diagnosed or previously phlebotomized only, and all had evidence of active disease at enrolment. Active disease was defined by a requirement for phlebotomy, WBC $> 10 \times 10^9/\text{L}$ or

Characteristics	ET (n = 1)	PV (n = 4)	Post-PV MF (n = 1)	Total (n = 6)
Age (y)	68	66 (59-70)	75	68 (59-75)
Gender, male	0 (0)	4 (80)	1 (100)	5 (83)
History of major thrombohemorrhagic event	0 (0)	3 (75)	0 (0)	3 (50)
Prior cytoreductive therapy	1 (100)	2 (50)	1 (100)	4 (67)
Hydroxyurea	1 (100)	0 (0)	1 (100)	2 (33)
r-IFN α -2a	1 (100)	1 (100)	1 (100)	3 (50)
r-IFN α -2b	0 (0)	1 (100)	0 (0)	1 (17)
Time off cytoreductive therapy before first <i>JAK2V617F</i> measurement (months)	0.5	1.2 (0.6-1.8)	5.3	1.2 (0.5-5.3)
<i>JAK2V617F</i> allele (%)	6	6 (0.7-27)	93	8.5 (0.7-96)

platelets $> 400 \times 10^9/L$ in the absence of infection or inflammation, hypermetabolic symptoms ie weight loss $> 10\%$ within 6 months, night sweats, low-grade fever for more than 2 weeks without signs of infection, pruritus, splenomegaly with symptoms, or previous thrombosis.

Eligible patients from the outpatient clinic studied retrospectively off cytoreductive treatment have been described above.

2.3 | Intervention

DALIAH patients received monotherapy with either IFN α -2a or r-IFN α -2b subcutaneously once weekly at a starting dose of 45 and 35 μ g, respectively. Dose escalation was performed in a stepwise manner at pre-defined time points in the absence of a complete hematological response (i.e., WBC $> 10 \times 10^9/L$ or platelets $> 400 \times 10^9/L$) after 4 and 12 months and in the absence of a partial or complete molecular response according to the 2009 European LeukaemiaNet (ELN) criteria⁵⁴ after 8 and 18 months. However, the IFN dose was de-escalated by the treating physician to the highest tolerable dose in the event of drug-related toxicity. Toxicity was graded according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 and IFN was discontinued in the event of grade 4 events or recurrent grade 3 events (See Supplementary Material E).

2.4 | Molecular diagnosis and *JAK2V617F*

The *JAK2V617F* allele burden was accessed by a highly sensitive quantitative real-time polymerase chain reaction (qPCR) on DNA from peripheral blood,⁵⁵ which has been assessed as the European Reference Assay.⁵⁶ In the DALIAH-trial, *JAK2V617F* measurements were performed every three

TABLE 2 Retrospective study.

Baseline demographics and clinical characteristics of *JAK2V617F* patients from the outpatient clinic at the time of first *JAK2V617F* measurement

months the first year, once every six months the second year, and yearly thereafter until the end of the study. For patients followed in the outpatient clinic, the *JAK2V617F* allele burden was assessed according to the physician's decision. All patients with four or more *JAK2V617F* samples were eligible for data-driven analysis.

2.5 | *JAK2V617F* allele burden development in untreated patients

Due to the expanding nature of malignant cells, the *JAK2V617F* allele burden in untreated patients is expected to increase exponentially. Hence, it is reasonable to assume that an exponentially increasing function can be fitted to data and an exponential growth-rate for specific patients can be found.

To generalize from multiple patient-specific growth-rates, we calculated the mean of the growth-rates, resulting in an expression of the growth on a population level. The expression describes the expected growth of the *JAK2V617F* allele burden in a larger population across multiple orders of magnitude, even if the estimate is made on a small sample of the population. Since the *JAK2V617F* allele burden varies at diagnosis, we shifted data in time such that the individual patient-specific fits coalesced with the population-level growth-curve at the mean time of these observations. Data were then pooled into a single data-set, and the exponential growth was estimated, see Table S5. All growth rates were found with MATLAB R2018a, using the least square fitting method *fit* included in the *curve fit* toolbox.

Of the six patients available for this analysis, one was excluded since all measurements were $\geq 90\%$, and thus above the point at which the growth is expected to be exponential. Note that this patient is the single post-PV MF patient shown in Table 2. For two of the remaining patients, a low number of *JAK2V617F* measurements were available, due to which these

were excluded in the initial part of the analysis. This leaves three remaining patients, referred to as patient A, B, and C.

2.6 | JAK2V617F allele burden development during IFN monotherapy

To describe the *JAK2V617F* development in IFN treated patients, two simple descriptive models were fitted to the time series measurements. The first model is that of a monoexponential decay, i.e., of the form $Ae^{-\alpha t}$ (with A and α being positive constants and t the time from treatment onset), while the second model features a bi-exponential decay $Be^{-\beta t} - Ce^{-\gamma t}$ (With B , C , β and γ being positive constants and t the time from treatment onset) with the requirement that the slope at treatment onset is equal to the slope found for the monoexponential growth. This reduces the number of independent parameters from four to three. The bi-exponential model can be considered an extension of the mono-exponential model with the bi-exponential model describing a similar response but allowing for an initial *JAK2V617F* increase before the decay starts.

Since the bi-exponential model has three parameters, patients with three or less *JAK2V617F* measurements were excluded. This reduced the number of datasets available from 120 to 66. We identified how well the models fit data by the adjusted R^2 -value, as described in supplementary material A. Thus, fits with an adjusted R^2 -value below a threshold of 0.6 were excluded from further analysis. The analysis for each of the two models was done independently.

3 | RESULTS

3.1 | JAK2V617F allele burden development in untreated patients

Raw data used to estimate the *JAK2V617F* allele burden development in untreated patients are depicted in Figures S72-S76.

Three patients (A, B and C) received IFN treatment up to the date of the initial measurement.

After IFN discontinuation, there was an initial delayed response in the *JAK2V617F* allele burden of approximately 200 days. Only data after the initial delayed response were used to determine the growth-rates. Data are illustrated in Figure 1 as well as the least-square fits of mono-exponential growth to the data, with 95% prediction intervals.

An exponential growth-rate implies a constant period of doubling. The periods of doubling found for patients A, B, and C are shown in Table S5, along with confidence intervals.

By pooling the data together, a new fit was made, yielding a general expression for the growth at population level. The resulting population level expression is shown in Figure 2 along with the pooled data and data for the two additional datasets which had only few *JAK2V617F* measurements. The additional datasets were timeshifted such that the mean of the *JAK2V617F* allele burden coalesced with the growth curve at the mean day of the dataset. While the two additional patients were not included in the fitting procedure, the data did not falsify the population level growth of the pooled data.

The period of doubling of the pooled data was found as 1.4 years (CI: 1.2 to 1.7 years). This implies that the allele burden grows from 0.01% to 1% in 9.3 years, while the growth from 1% to 33% takes 7.1 years. Therefore, detecting the *JAK2V617F* allele burden $\leq 1\%$ allows for a much longer time-window for detection and early therapeutic intervention before symptoms arise.

3.2 | JAK2V617F allele burden development during IFN monotherapy

All 66 eligible data-sets are depicted in the Figure S6 through S71.

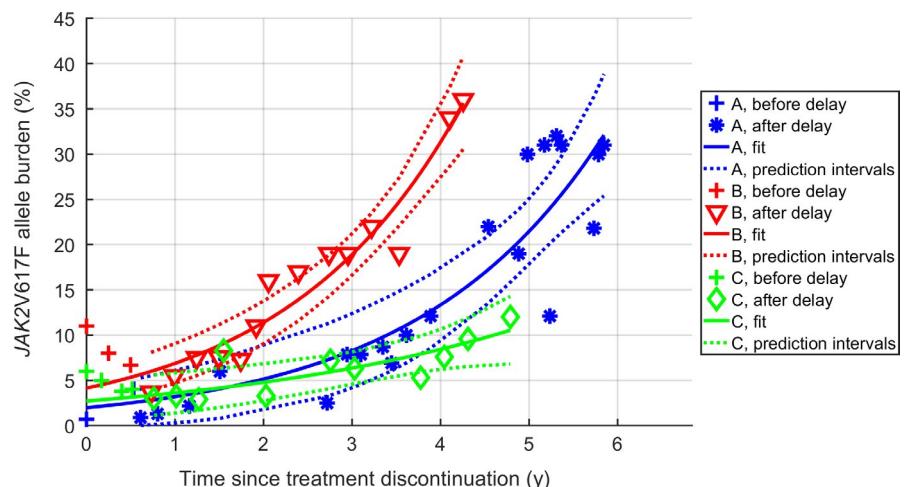


FIGURE 1 *JAK2V617F* allele burden development in untreated patients. Serial measurements for patients A, B and C with least-square fits of exponential growth with 95% prediction intervals. Fits were based only on data after the initial delayed response

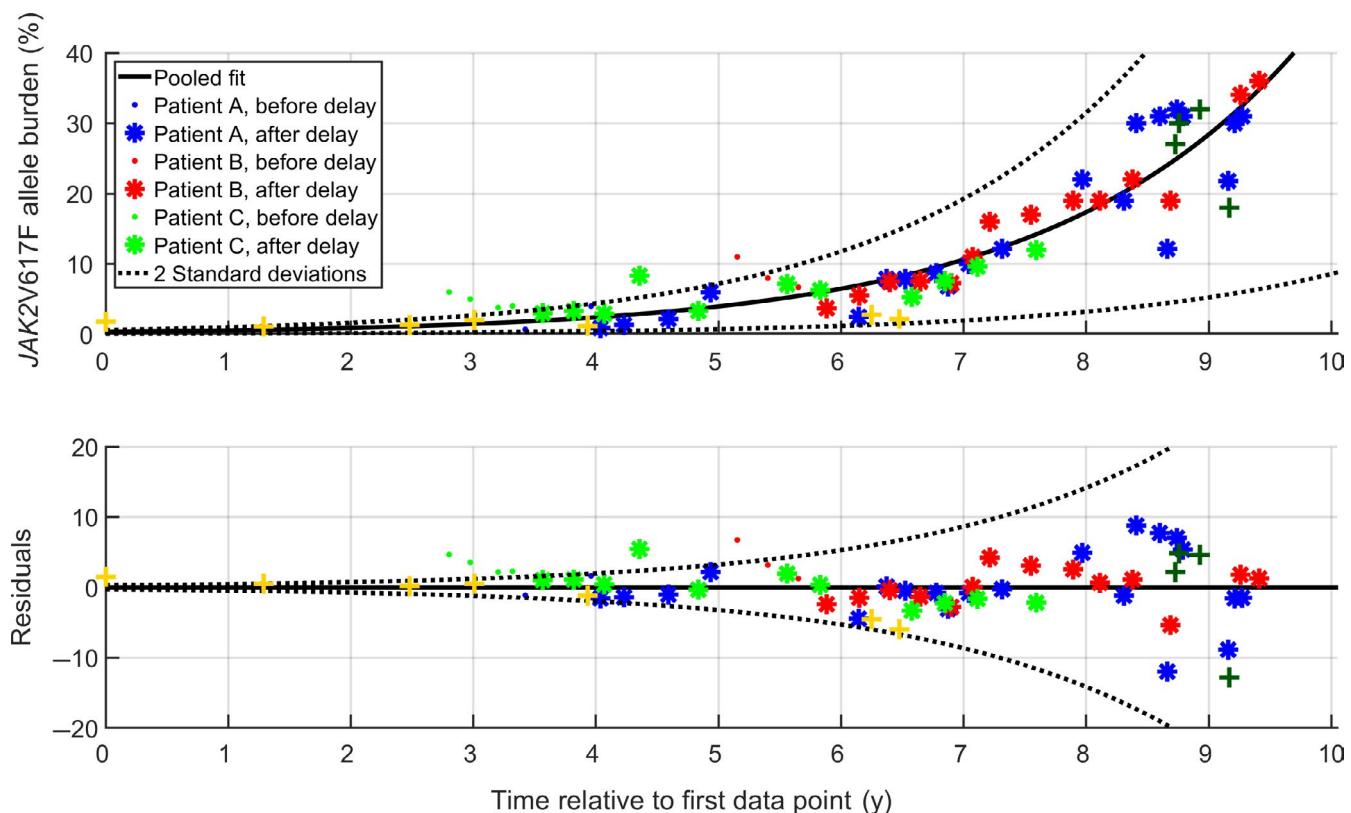


FIGURE 2 Exponential growth of the *JAK2V617F* allele burden. The top displays the exponential growth fit of the pooled data, while the bottom displays the relative residuals scaled with the size of allele burden. +-marks indicate the two patients with very few *JAK2V617F* measurements. The initial delayed response is included as dots. Each subject is depicted with a specific color. Two standard deviations around the mean of the scaled residuals are shown

Patient-specific parameters for the fits, the goodness of the fits, and which model was the better descriptor are shown in Supplementary Material A.

To generalize the *JAK2V617F* kinetics, a threshold for the goodness-of-fit was chosen and estimates for the population-level parameters were found, see Supplementary Material B.

Fits of 28 patients were deemed satisfactory for both response types (i.e., both mono- and bi-exponential). For 14 of these patients, the bi-exponential response was found to be the better fit for the patient data. Note that the goodness-of-fit measure chosen takes the complexity of the models into account, and as such the simpler monoexponential response was preferred when the models yielded almost identical results. Five additional patients had satisfactory fits for the bi-exponential response type but not for the monoexponential. Thus, the bi-exponential model was the best fit for 19 patients.

The datasets for the remaining 33 patients either featured responses not following any of the models or had no decay in the *JAK2V617F* allele burden and are hence considered nonresponders in the context of molecular response.

The monoexponential model using the population decay-rate is shown in Figure 3, along with patient-data for the patients for whom the goodness-of-fits were above the threshold.

The population decay rates found for the monoexponential model corresponded to a half-life of 575 days, or 1.6 years. (95% CI of the decay rates yields half-lives between 0.8 and 11.6 years).

For the monoexponential decay, the development was the same across all orders of magnitude, allowing for a single representative figure, Figure 3. Although the bi-exponential model features an initial growth depending on the starting level, it has a long-term behavior which is approximately monoexponential. This long-term behavior corresponds to a half-life of 1 year (CI: 0.2-4.3 years).

3.3 | Comparison of *JAK2V617F* allele burden development during early and late IFN monotherapy

In silico treatment schemes can be considered using the population level models and the growth rates found. Figure 4 displays the increase and decrease in *JAK2V617F* allele burden, with initiation of IFN treatment 7, 8 and 9 years after the allele burden reached 1%. The figure shows both the monoexponential response as well as the bi-exponential response at population level.

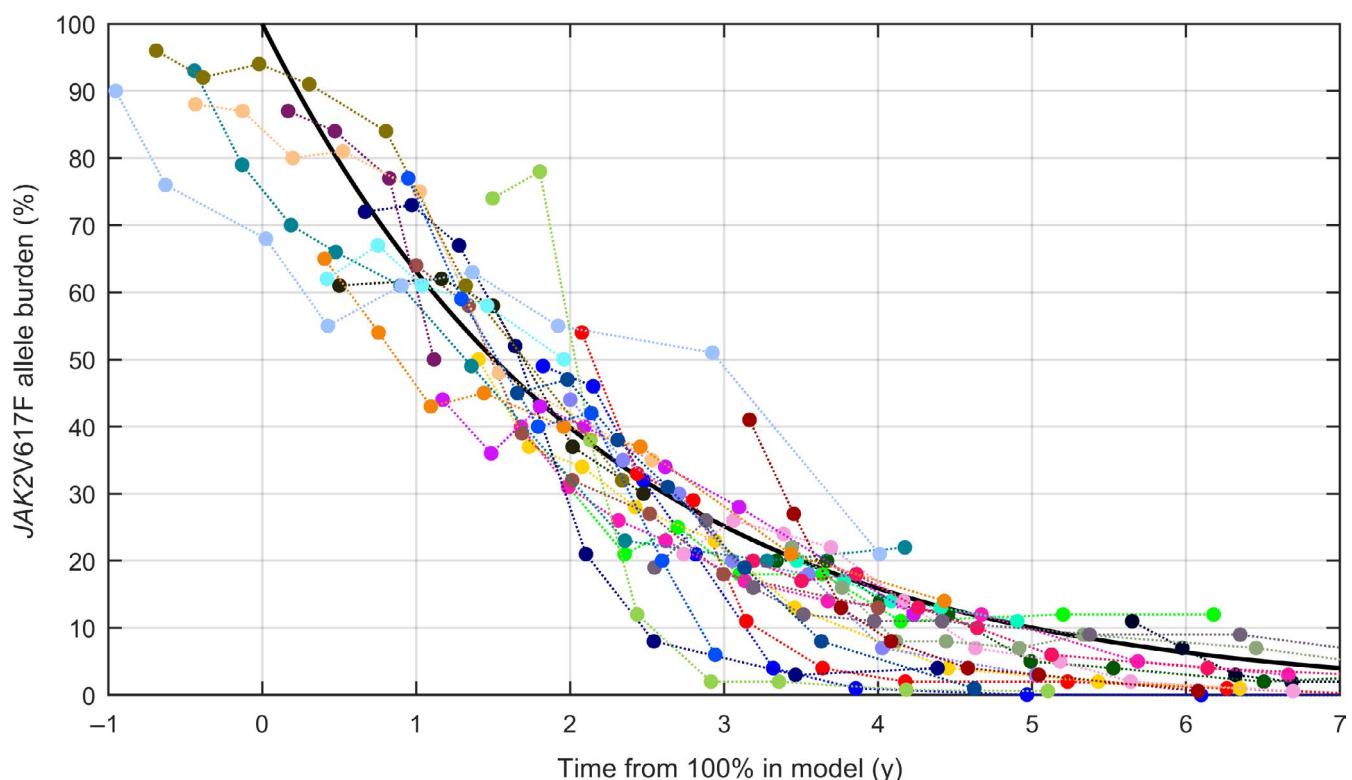
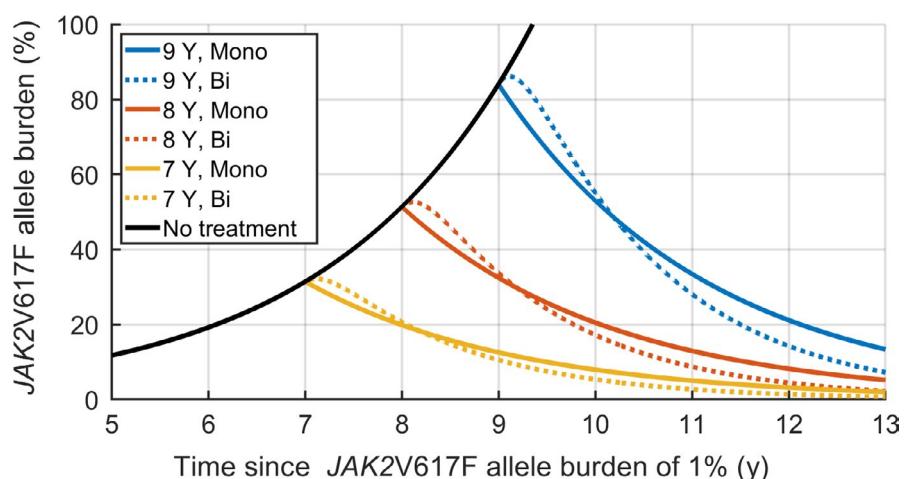


FIGURE 3 Population-level decay rate. The population-level decay rate of the mono-exponential model is shown in black, with data for the 29 patients with the best goodness-of-fits. Note that data were shifted in time such that the mean of the *JAK2V617F* allele burden coincides with the main curve at the mean day of measurement

FIGURE 4 Prediction of *JAK2V617F* development during IFN treatment.

JAK2V617F development over time estimated by either the mono-exponential or bi-exponential response (full lines: Mono-exponential, dotted lines: bi-exponential), with simulated treatment starting at 7, 8 and 9 y after 1% was reached



The figure suggests that the initial 6 months of IFN treatment are associated with minor or no decrease in the *JAK2V617F* allele burden for the bi-exponential response. However, after a year, the *JAK2V617F* allele burden decreases faster than for the monoexponential response. As such, the efficacy of the treatment in decreasing the *JAK2V617F* allele burden may be difficult to determine within the first year.

From a specific baseline *JAK2V617F* allele burden, the population-level responses can be used to estimate the duration of treatment necessary to achieve a *JAK2V617F* allele

burden of 1%. In Figure 5 the duration of treatment necessary is illustrated, when treatment is initiated at a given *JAK2V617F* allele burden.

The monoexponential response features a longer treatment period necessary than the bi-exponential response, when the initial *JAK2V617F* allele burden is greater than a few percent. As such, even if a patient has an initial increase in *JAK2V617F* allele burden as is the case for the bi-exponential response, 1% will be reached faster than if the response was monoexponential.

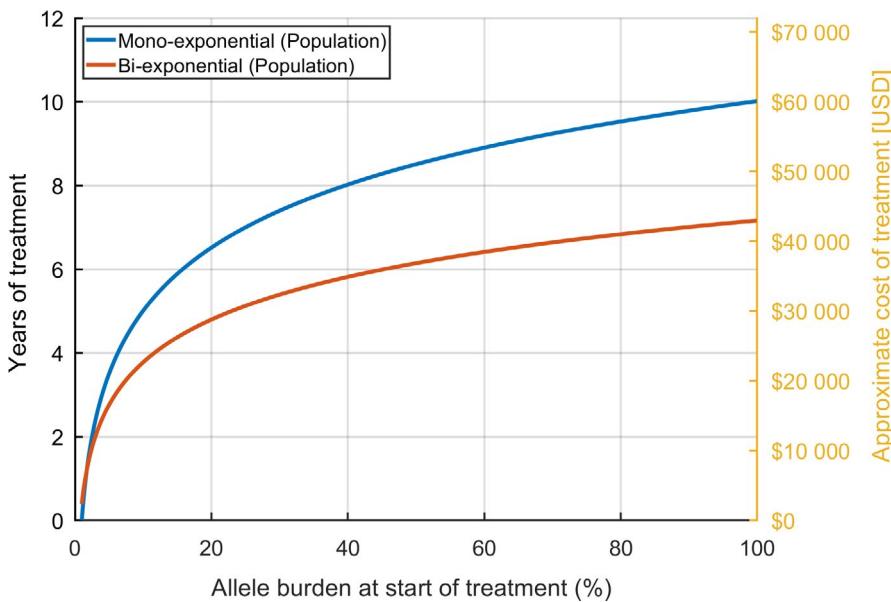


FIGURE 5 Estimated cost of IFN treatment. Years of treatment necessary to reach a *JAK2V617F* allele burden of 1%, if treatment is initiated at the allele burden shown on the first axis. The approximate monetary cost of IFN treatment (estimated price: \$500 USD/month) is also included

Figure 5 also includes a rough estimate of the total medical expenses to IFN treatment (approximately \$500 USD/month) based on the *JAK2V617F* allele burden at treatment onset. While relapse after treatment is still possible, the figure illustrates that the total cost of IFN treatment increases with the *JAK2V617F* allele burden at treatment onset due to the need for longer treatment duration to reach a specific *JAK2V617F* target value.

3.4 | Time span of the *JAK2V617F* allele burden development

The growth of the *JAK2V617F* allele burden from 0.01% to 1% was found to span almost a decade, while further growth to 33% required approximately 7 years. If patients are screened for the *JAK2V617F* mutation on a regular basis, eg once every 10 years, a detection limit of 1% might miss the disease onset, since the *JAK2V617F* allele burden would exceed 33% before the next screening. Conversely, a detection limit of 0.01% could be expected to identify an allele burden below 1% during the 9.3-year period, or it may grow to 1.4% 10 years after an allele burden of 0.01%. This also emphasizes the importance of methods to quantify the *JAK2V617F* allele burden down to low levels, in particular below 1%.

Although exponential growth may not be expected for low numbers of cells, extrapolating back to the time at which the initial mutation appeared (i.e., an allele burden corresponding to a single cell) yields a conservative estimate of the time span of the development before symptoms arise and MPN is diagnosed. The *JAK2V617F* allele burden for patients diagnosed with PV has previously been estimated as 33% (CI: 20-40).⁵⁷ We assume that the

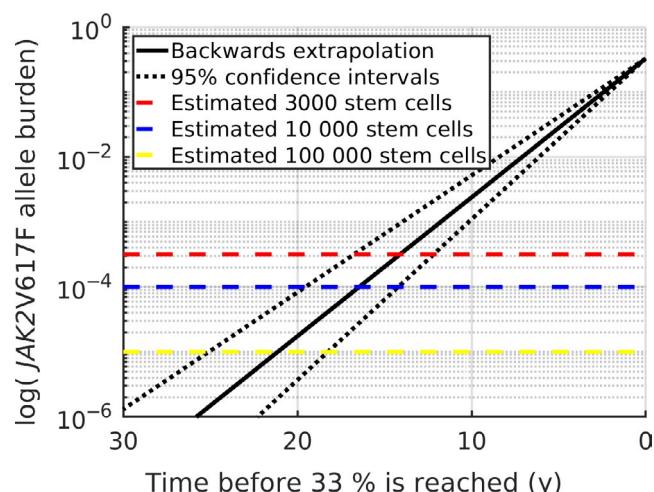


FIGURE 6 Backwards extrapolation to onset of initial mutation. The growth rate (and confidence intervals in dotted black) extrapolated back in time from 33%. The dashed lines mark the limits of $10^{-3.5}$, 10^{-4} and 10^{-5} as discussed in the text, in red, blue and yellow, respectively. Note the logarithmic y-axis

JAK2V617F allele burden in the peripheral blood and of the stem cells are similar. The total number of stem cells has been estimated to be between 11 200 and 22 400 cells,⁵⁸ which is of order 10^4 . As such, an estimate of the allele burden at initial mutation is 10^{-4} , ie one out of 10^4 . This implies a total time from initial mutation to allele burden of 33% of around 16.5 years (CI: 14.2-19.5 years). This estimate of the timespan of MPN development is in agreement with the literature.⁵⁹

Figure 6 shows the extrapolation from an allele burden of 33% backwards in time, with the growth-rate found. The estimate is highly dependent on the total number of stem cells. If another estimate of the number of stem cells is

used, the period before 33% is reached can be found from the graph, as well as the confidence intervals. Estimating a lower stem cell count of 3000⁶⁰ leads to an allele burden of $10^{-3.5}$ at initial mutation, and the time of development would be approximately 14 years. Similarly, estimating a stem cell count between 50 000 and 200 000 as in a most recent paper⁶¹ leads to an initial allele burden around 10^{-5} , and consequently a timespan of 21 years for the development of MPNs.

4 | DISCUSSION

In recent years, the interest of using IFN in the treatment of MPNs has increased due to studies reporting long-term treatment with IFN to be associated with MRD in a subset of patients as defined by sustained complete hematological remissions in concert with induction of low-burden *JAK2V617F* and normalization of the bone marrow.^{36,44,48} Even in the IFN era of 2019 where several safety and efficacy studies have enrolled > 1000 patients during the last 30 years, the MPN scientific community still recommends patients with low risk disease to be observed without any cytoreductive treatment. Using this “watch and wait” strategy translation of common knowledge on cancer biology to MPNs is neglected, implying progression of any cancer without treatment. Early treatment with IFN has been claimed to be a prerequisite for obtaining remarkable results to prohibit clonal evolution before subclones and additive mutations evolve.^{38,40,41,62-67}

The present study delivers novel information regarding the *JAK2V617F* kinetics during IFN-treatment based upon unique serial *JAK2V617F* measurements from the DALIAH trial. Through this description of the *JAK2V617F* kinetics, predictions about the development of disease for specific patients can be made. In untreated patients (i.e., without cytoreductive treatment), the *JAK2V617F* allele burden was demonstrated to grow exponentially with doubling time of 1.4 years (CI: 1.2 to 1.7 years). During IFN treatment, the *JAK2V617F* development followed either a monoexponential or a bi-exponential decay for a significant proportion of patients, with both models describing the development to a satisfactory extent for 28 patients. In a total of 33 patients, the bi-exponential response was found to be the better descriptor of the development when adjusting for model complexity.

The individual *JAK2V617F* developments were combined into two models of the development on a population level. The population-level model of the monoexponential decay showed a half-life of the *JAK2V617F* allele burden of 1.6 years (CI: 0.8 to 11.6 years), while the long-term behavior of the bi-exponential decay displayed a half-life of 1 year (CI: 0.2-4.3 years).

We emphasize that since these population-level models are based on a subset of patients who respond well to treatment, the models do not necessarily generalize to all patients. Determining the development for the poor responders remains an open problem as well as determining why there is a difference in the treatment response.

Analysis of the population-level responses suggests that treatment schemes should extend longer than one year, as the efficacy of IFN treatment on the *JAK2V617F* allele burden cannot be determined after just one year. By comparing early and late treatment modalities, our results suggest that treatment outcome will improve if IFN therapy is initiated early, or will in any case lead to early identification of patient response type. Additionally, some patients had a significantly slower response to treatment compared to other patients. For these slow-responders, our results show, that treatment should be initiated as early as possible, since a “watch and wait” strategy increases the time needed to obtain responses drastically for each day spent watching and waiting.

We and others have argued against the “watch and wait” strategy in low-risk patients.^{38,40,41,62-67} Our study emphasizes the urgent need to rethink this approach and set new standards for treatment of patients with MPNs, implying normalization of cell counts in all patients using IFN from the time of diagnosis. In addition to the rationales provided by the results in our present study, several others are supportive of the early-IFN-intervention concept. Thus, cancer biology in general dictates that any cancer steadily evolves over time with expansion of the malignant clone, increasing genomic instability, subclone formation and ultimately metastasis. Fortunately, MPNs are slowly growing neoplasms which have several transitional stages in the biological continuum from the earliest cancer stages (eg *JAK2V617F* and *CALR* mutations in the background population as clonal hematopoiesis of indeterminate potential (CHIP)) to ET, PV, and the advanced cancer stage with myelofibrosis, bone marrow failure, and huge splenomegaly before terminal leukemic transformation. During this MPN-biological continuum, the *JAK2V617F* mutation will steadily increase in those individuals who develop overt MPN in concert with an increase in the chronic inflammatory load that drives the malignant clone and likely fuels the development of additional mutations as well.⁶⁴⁻⁶⁷ Thus, in this context, it is tempting to suggest that the total number of mutated cells present at any given time may be the driver of additive mutations. Indeed, the sum of the *JAK2V617F* allele burden over a period of time may provide a measure for the risk of additive mutations as it correlates with the number of mutated cells in the given the period. As shown in Supplementary Material C, this measure—as assessed by the *JAK2V617F* allele burden—increases exponentially with the time spent without

treatment. Although speculative, this suggests that, after enough time has passed above a certain threshold, the potential risk of additive mutations may have the same exponential growth-rate (and thus the same period of doubling) as the growth of the *JAK2V617F* allele burden before treatment is initiated. As such, although the exact risk cannot be determined, postponing treatment by 1.4 years will double the risk of an additive mutation. Further data-driven mathematical studies on this potential association are needed, including results from next generation sequencing in IFN-treated *JAK2V617F* positive MPN-cohorts.

Treatment with IFN is associated with drop-out rates of 20%-30% and in some studies even up to 40% due to toxic side effects.³⁷⁻⁴³ It has been speculated whether intolerance to IFN is also dependent upon disease stage, implying more toxic side effects in the advanced myelofibrosis stage and less so in the early disease stages. If so, evidence for exponential growth with doubling time of 1.4 years further undermines the "watch and wait" strategy and adds to the rationales of early treatment with IFN.

Our results deliver important information about treatment duration with IFN to obtain deep molecular remissions and long-lasting sustained remissions after drug discontinuation. The mathematical models presented serve as a novel platform for predicting IFN response in individual patients. More advanced mechanism-based mathematical models are foreseen to allow for improved prediction and insight into IFN response.

Our study dictates that institution of IFN at the earliest time point possible may have important socio-economic implications as well. By minimizing the risk of complications (thrombosis, hemorrhages or cancer), a huge economic burden due to hospitalizations is likely markedly reduced. Importantly, costs concerning rehabilitation after these complications are reduced as well. Another important consequence of early treatment with IFN is the outlook to achieve MRD and at this time point the possibility of discontinuation of IFN for several years (up to 3-5 years) when the patient is feeling healthy with normal cell counts. Of note, our study also suggests that early treatment with IFN is cost-effective, implying a shorter treatment period with IFN if treatment is instituted at the earliest time point possible.

Although our data are supportive of early intervention with IFN it is important to underscore that our data do not deliver the clinical proof for this early intervention recommendation. This proof can only be delivered by the demonstration of reduction in clinically relevant end-points such as thrombotic events, rate of transformation to myelofibrosis and acute leukemia. Indeed, the demonstration of these hard end-points would require long-term follow up of large cohorts of patients treated with IFN and a well-designed control group. This study has never been reported and will

likely never be reported in these orphan diseases, where randomized studies are so difficult to conduct - in particular with follow-up times of decades rather than for instance 5 years, when taking into account that leukemic transformation in general is a late event, developing in the advanced myelofibrosis stage of MPNs. Importantly, the randomized Proud/Continuation-PV Phase III Trials showed RopergIFN (Besremi®) to be associated with a clear benefit over control (ie hydroxyurea) in achieving significant higher maintenance rates of complete hematological remission (CHR) over the course of treatment and in showing a significant lower risk of losing CHR. Since CHR can be considered as surrogate for risk of thrombosis, RopergIFN may be an optimal treatment modality for managing risk of thrombosis.⁷⁰

Our study has some limitations. Firstly, our estimate of the development of the *JAK2V617F* allele burden was based specifically on the data from three patients only (A, B, and C) and longitudinal *JAK2V617F* allele burden measurements in more patients might have substantiated and strengthened our findings of an exponentially growing pattern, which has not previously been described mathematically. Secondly, we did not include serial measurements of the *JAK2V617F* allele burden after discontinuation of hydroxyurea. Unfortunately, we have not such data in our cohort of patients. However, based upon current knowledge on the kinetics of the leukocyte and platelet counts after a few days off hydroxyurea treatment, displaying rapid increases in the cell counts to pretreatment levels, it is reasonable to assume that the *JAK2V617F* allele burden might similarly increase after hydroxyurea discontinuation. An increase in the *JAK2V617F* allele burden when treatment with hydroxyurea is terminated has been previously demonstrated.⁷¹

Third, although the findings in our study are supportive of early treatment with IFN, it does not deliver the definite proof, which would require a study, showing that early IFN-treatment from the time of diagnosis influences hard clinical end-points, such as risk of thrombosis and major bleeding, transformation to myelofibrosis and acute myeloid leukemia and ultimately survival. Hopefully, our DALIAH trial, from which our data in the present study have been retrieved, may provide such data within the next 10-20 years.

In conclusion, data-driven analysis is a novel tool for providing further evidence for the concept of early intervention with IFN in MPNs. Our observations also emphasize that starting treatment early allows for identification of patient responses. Understanding the kinetics of the *JAK2V617F* allele burden is highly valuable in guiding future clinical decisions about IFN-treatment of the individual patient. In this context, our findings substantiate and put in perspective the urgent need of personalized medicine with IFN in MPN-patients.

ACKNOWLEDGMENTS

The authors thank the patients and their families, as well as the research coordinators and site staff for their major effort. Our thanks are also extended to the clinician who helped identify eligible patients. Additionally the authors are grateful for the following grants which helped fund this study: “OUH Frie Forskningsmidler”, “OUH-Region Sjælland Fælles Forskningspulje”, “Fonden til Lægevidenskabens Fremme”, “Ellen og Aage Fausbølls Helsefond af 1975”, “Swedish Orphan”, “Den Forskningsfremmende Pulje for Kliniske Professorer ansat i Region Sjælland og ved Institut for Klinisk Medicin, Københavns Universitet”, “Region Sjællands Sundhedsvidenkabelige Forskningsfond (RSSF) 2018”, “Region Sjælland – Produktion, Forskning og Innovation (PFI)” and “Roskilde University Research Grant.”

CONFLICT OF INTEREST

HH has received a research grant from Novartis Oncology.

AUTHOR CONTRIBUTIONS

Rasmus K. Pedersen: Formal analysis (lead), project administration (supporting), methodology (supporting), software (lead), visualization, writing – original draft, and writing – review and editing. Morten Andersen: Formal analysis (supporting), methodology (supporting), software (supporting), supervision (supporting), and writing – review and editing. Trine A. Knudsen: Investigation (lead), and writing – review and editing. Zamra Sajid: Writing – review and editing. Johanne Gudmand-Hooyer: Writing – review and editing. Marc J. B. Dam: Formal analysis (supporting), software (supporting), writing – review and editing. Vibe Skov: Investigation (supporting), methodology (supporting), writing – review and editing. Lasse Kjær: Investigation (supporting), methodology (supporting), writing – review and editing. Christina Ellervik: Writing – review and editing. Thomas S. Larsen: Investigation (supporting). Dennis Hansen: Data curation (lead), investigation (supporting). Niels Pallisgaard: Methodology (supporting). Hans C. Hasselbalch: Conceptualization (lead), funding acquisition (equal), methodology (supporting), investigation (supporting), supervision (supporting), validation (equal), and writing – review and editing. Johnny T. Ottesen: Conceptualization, funding acquisition (equal), methodology (lead), software (supporting), supervision (lead), validation (equal), and writing – review and editing.

ORCID

Rasmus K. Pedersen  <https://orcid.org/0000-0001-5946-8220>
 Morten Andersen  <https://orcid.org/0000-0001-7582-0685>
 Trine A. Knudsen  <https://orcid.org/0000-0001-9829-3099>
 Zamra Sajid  <https://orcid.org/0000-0002-9143-3092>
 Johanne Gudmand-Hooyer  <https://orcid.org/0000-0002-4972-0653>

Marc J. B. Dam  <https://orcid.org/0000-0003-0096-9690>
 Vibe Skov  <https://orcid.org/0000-0003-0097-7826>
 Lasse Kjær  <https://orcid.org/0000-0001-6767-0226>
 Christina Ellervik  <https://orcid.org/0000-0002-3088-4375>
 Thomas S. Larsen  <https://orcid.org/0000-0002-8418-6708>
 Dennis Hansen  <https://orcid.org/0000-0002-4478-1297>
 Hans C. Hasselbalch  <https://orcid.org/0000-0003-3936-8032>
 Johnny T. Ottesen  <https://orcid.org/0000-0002-9786-0300>

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

- Campbell PJ, Green AR. The myeloproliferative disorders. *N Engl J Med.* 2006;355(23):2452-2466.
- Marchioli R, Finazzi G, Landolfi R, et al. Vascular and neoplastic risk in a large cohort of patients with polycythemia vera. *J Clin Oncol.* 2005;23(10):2224-2232.
- Björkholm M, Derolf ÅR, Hultcrantz M, et al. Treatment-related risk factors for transformation to acute myeloid leukemia and myelodysplastic syndromes in myeloproliferative neoplasms. *J Clin Oncol.* 2011;29(17):2410-2415.
- James C, Ugo V, Le Couédic J-P, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature.* 2005;434(7037):1144-1148.
- Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005;352(17):1779-1790.
- Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet.* 2005;365(9464):1054-1061.
- Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell.* 2005;7(4):387-397.
- Larsen TS, Pallisgaard N, Møller MB, Hasselbalch HC. The JAK2 V617F allele burden in essential thrombocythemia, polycythemia vera and primary myelofibrosis - Impact on disease phenotype. *Eur J Haematol.* 2007;79(6):508-515.
- Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of Calreticulin in myeloproliferative neoplasms. *N Engl J Med.* 2013;369(25):2379-2390.
- Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med.* 2013;369(25):2391-2405.
- Cazzola M, Kralovics R. From Janus kinase 2 to calreticulin : the clinically relevant genomic landscape of myeloproliferative neoplasms. *Blood.* 2016;123(24):3714-3720.
- Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood.* 2014;123(14):2220-2228.
- Wolanskyj AP, Lasho TL, Schwager SM, et al. JAK2V617F mutation in essential thrombocythaemia: clinical associations and

- long-term prognostic relevance. *Br J Haematol.* 2005;131(2):208-213.
14. Carobbio A, Finazzi G, Guerini V, et al. Leukocytosis is a risk factor for thrombosis in essential thrombocythemia: Interaction with treatment, standard risk factors, and Jak2 mutation status. *Blood.* 2007;109(6):2310-2313.
 15. Tefferi A, Strand JJ, Lasho TL, et al. Bone marrow JAK2V617F allele burden and clinical correlates in polycythemia vera. *Leukemia.* 2007;21(9):2074-2075.
 16. Vannucchi AM, Antonioli E, Guglielmelli P, et al. Clinical profile of homozygous JAK2 617V> F mutation in patients with polycythemia vera or essential thrombocythemia. *Blood.* 2007;110(3):840-846.
 17. Vannucchi AM, Antonioli E, Guglielmelli P, et al. Prospective identification of high-risk polycythemia vera patients based on JAK2V617F allele burden. *Leukemia.* 2007;21(9):1952-1959.
 18. Cheung B, Radia D, Pantelidis P, Yadegarfar G, Harrison C. The presence of the JAK2 V617F mutation is associated with a higher haemoglobin and increased risk of thrombosis in essential thrombocythaemia. *Br J Haematol.* 2006;132(2):244-245.
 19. Tefferi A, Lasho TL, Schwager SM, et al. The clinical phenotype of wild-type, heterozygous, and homozygous JAK2V617F in polycythemia vera. *Cancer.* 2006;106(3):631-635.
 20. Passamonti F, Rumi E, Pietra D, et al. Relation between JAK2(V617F) mutation status, granulocyte activation, and constitutive mobilization of CD34+ cells into peripheral blood in myeloproliferative disorders. *Blood.* 2006;107(9):3676-3682.
 21. Arellano-Rodrigo E, Alvarez-Larrán A, Reverter JC, et al. Increased platelet and leukocyte activation as contributing mechanisms for thrombosis in essential thrombocythemia and correlation with the JAK2 mutational status. *Haematologica.* 2006;91(2):169-175.
 22. Bellucci S, Michiels JJ. The role of JAK2 V617F mutation, spontaneous erythropoiesis and megakaryocytopoiesis, hypersensitive platelets, activated leukocytes, and endothelial cells in the etiology of thrombotic manifestations in polycythemia vera and essential thrombocythemia. *Semin Thromb Hemost.* 2006;32(4):381-398.
 23. Falanga A, Marchetti M, Vignoli A, et al. V617F JAK-2 mutation in patients with essential thrombocythemia: relation to platelet, granulocyte, and plasma hemostatic and inflammatory molecules. *Exp Hematol.* 2007;35(5):702-711.
 24. Kittur J, Knudson RA, Lasho TL, et al. Clinical correlates of JAK2V617F allele burden in essential thrombocythemia. *Cancer.* 2007;109(11):2279-2284.
 25. Hsiao HH, Yang MY, Liu YC, et al. The association of JAK2V617F mutation and leukocytosis with thrombotic events in essential thrombocythemia. *Exp Hematol.* 2007;35(11):1704-1707.
 26. Barosi G, Rosti V, Bonetti E, et al. Evidence that prefibrotic myelofibrosis is aligned along a clinical and biological continuum featuring primary myelofibrosis. *PLoS ONE.* 2012;7(4):e35631.
 27. Kiladjian J-J. High molecular response rate of polycythemia vera patients treated with pegylated interferon -2a. *Blood.* 2006;108(6):2037-2040.
 28. Silver RT. Long-term effects of the treatment of polycythemia vera with recombinant interferon- α . *Cancer.* 2006;107(3):451-458.
 29. Kiladjian JJ, Giraudier S, Cassinat B. Interferon-alpha for the therapy of myeloproliferative neoplasms: targeting the malignant clone. *Leukemia.* 2016;30(4):776-781.
 30. Kjær L, Cordua S, Holmström MO, et al. Differential dynamics of CALR mutant allele burden in myeloproliferative neoplasms during interferon alfa treatment. *PLoS ONE.* 2016;11(10):e0165336.
 31. Masarova L, Yin CC, Cortes JE, et al. Histomorphological responses after therapy with pegylated interferon α -2a in patients with essential thrombocythemia (ET) and polycythemia vera (PV). *Exp Hematol Oncol.* 2017;6(1):1-13.
 32. Crisà E, Cerrano M, Beggiato E, et al. Can pegylated interferon improve the outcome of polycythemia vera patients? *J Hematol Oncol.* 2017;10(1). <https://doi.org/10.1186/s13045-017-0395-1>
 33. Foucar CE, Stein BL. Contemporary use of interferon therapy in the myeloproliferative neoplasms. *Curr Hematol Malig Rep.* 2017;12(5):406-414.
 34. Gowin K, Jain T, Kosiorek H, et al. Pegylated interferon alpha—2a is clinically effective and tolerable in myeloproliferative neoplasm patients treated off clinical trial. *Leuk Res.* 2017;54:73-77.
 35. Tashi T, Swierczek S, Kim SJ, et al. Pegylated interferon Alfa-2a and hydroxyurea in polycythemia vera and essential thrombocythemia: differential cellular and molecular responses. *Leukemia.* 2018;32(8):1830-1833.
 36. Larsen TS, Bjerrum OW, Pallisgaard N, et al. Sustained major molecular response on interferon alpha-2b in two patients with polycythemia vera. *Ann Hematol.* 2008;87(10):847-850.
 37. Hasselbalch HC, Larsen TS, Riley CH, Jensen MK, Kiladjian J. Interferon-alpha in the treatment of Philadelphia-negative chronic myeloproliferative neoplasms. Status and perspectives. *Curr Drug Targets.* 2011;12(3):392-419.
 38. Hasselbalch HC. A new era for IFN-alpha in the treatment of Philadelphia-negative chronic myeloproliferative neoplasms. *Expert Rev Hematol.* 2011;4(6):637-655.
 39. Samuelsson J, Hasselbalch H, Bruserud O, et al. A phase II trial of pegylated interferon α -2b therapy for polycythemia vera and essential thrombocythemia. *Cancer.* 2006;106(11):2397-2405.
 40. Silver RT, Hasselbalch HC, Kiladjian JJ. Interferon and the treatment of polycythemia vera, essential thrombocythemia and myelofibrosis. *Expert Rev Hematol.* 2013;6(1):49-58.
 41. Hasselbalch HC, Silver RT. Interferon in polycythemia vera and related neoplasms. Can it become the treatment of choice without a randomized trial? *Expert Rev Hematol.* 2015;8(4):439-445.
 42. Stauffer Larsen T, Iversen KF, Hansen E, et al. Long term molecular responses in a cohort of Danish patients with essential thrombocythemia, polycythemia vera and myelofibrosis treated with recombinant interferon alpha. *Leuk Res.* 2013;37(9):1041-1045.
 43. Stein BL, Tiu RV. Biological rationale and clinical use of interferon in the classical BCR-ABL-negative myeloproliferative neoplasms. *J Interf Cytokine Res.* 2013;33(4):145-153.
 44. Utke Rank C, Weis Bjerrum O, Larsen TS, et al. Minimal residual disease after long-term interferon-alpha2 treatment: A report on hematological, molecular and histomorphological response patterns in 10 patients with essential thrombocythemia and polycythemia vera. *Leuk Lymphoma.* 2016;57(2):348-354.
 45. Masarova L, Patel KP, Newberry KJ, et al. Pegylated interferon alfa-2a in patients with essential thrombocythaemia or polycythaemia vera: a post-hoc, median 83 month follow-up of an open-label, phase 2 trial. *Lancet Haematol.* 2017;4(4):e165-e175.
 46. Kiladjian J, Cassinat B, Chevret S, et al. Pegylated Interferon-alfa-2a induces complete hematological and molecular responses with low toxicity in Polycythemia Vera. *Hematology.* 2008;112(8):3065-3073.
 47. Quintás-Cardama A, Kantarjian H, Mansouri T, et al. Pegylated interferon alfa-2a yields high rates of hematologic and molecular response in patients with advanced essential thrombocythemia and polycythemia vera. *J Clin Oncol.* 2009;27(32):5418-5424.

48. Larsen TS, Møller MB, de Stricker K, et al. Minimal residual disease and normalization of the bone marrow after long-term treatment with alpha-interferon2b in polycythemia vera. A report on molecular response patterns in seven patients in sustained complete hematological remission. *Hematology*. 2009;14(6):331-334.
49. Quintás-Cardama A, Abdel-Wahab O, Mansouri T, et al. Molecular analysis of patients with polycythemia vera or essential thrombocythemia receiving pegylated interferon α -2a. *Blood*. 2013;122(6):893-901.
50. Verger E, Cassinat B, Dosquet C, Giraudier S. Clinical and molecular response to interferon- α therapy in essential thrombocythemia patients with CALR mutations. *Blood*. 2015;126(24):2585-2592.
51. Gisslinger H, Zagrijtschuk O, Buxhofer-ausch V, et al. Ropeginerferon alfa-2b, a novel IFN α -2b, induces high response rates with low toxicity in patients with polycythemia vera. *Blood*. 2015;126(15):1762-1770.
52. Them NCC, Bagienski K, Berg T, et al. Molecular responses and chromosomal aberrations in patients with polycythemia vera treated with peg-proline-interferon alpha-2b. *Am J Hematol*. 2015;90(4):288-294.
53. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*. 2011;117(19):5019-5032.
54. Barosi G, Birgegard G, Finazzi G, et al. Response criteria for essential thrombocythemia and polycythemia vera: result of a European LeukemiaNet consensus conference. *Blood*. 2009;113(20):4829-4833.
55. Larsen TS, Christensen JH, Hasselbalch HC, Pallisgaard N. The JAK2 V617F mutation involves B- and T-lymphocyte lineages in a subgroup of patients with Philadelphia-chromosome negative chronic myeloproliferative disorders. *Br J Haematol*. 2007;136(5):745-751.
56. Jovanovic JV, Ivey A, Vannucchi AM, et al. Establishing optimal quantitative-polymerase chain reaction assays for routine diagnosis and tracking of minimal residual disease in JAK2-V617F-associated myeloproliferative neoplasms: A joint European LeukemiaNet/MPN&MPNr-EuroNet (COST action BM0902) study. *Leukemia*. 2013;27(10):2032-2039.
57. Larsen TS, Pallisgaard N, Møller MB, Hasselbalch HC. The JAK2 V617F allele burden in essential thrombocythemia, polycythemia vera and primary myelofibrosis—impact on disease phenotype. *Eur J Haematol*. 2007;79(6):508-515.
58. Abkowitz JL, Catlin SN, McCallie MT, Guttorp P. Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood*. 2002;100(7):2665-2667.
59. Enblom A, Lindsjö E, Hasselbalch H, et al. High rate of abnormal blood values and vascular complications before diagnosis of myeloproliferative neoplasms. *Eur J Intern Med*. 2015;26(5):344-347.
60. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. 2018;553(7689):418-426.
61. Lee-Six H, Øbro NF, Shepherd MS, et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature*. 2018;561(7724):473-478.
62. Bjørn ME, Hasselbalch HC. Minimal residual disease or cure in MPNs? Rationales and perspectives on combination therapy with interferon-alpha2 and ruxolitinib. *Expert Rev Hematol*. 2017;10(5):393-404.
63. Hasselbalch HC, Holmström MO. Perspectives on interferon-alpha in the treatment of polycythemia vera and related myeloproliferative neoplasms: minimal residual disease and cure? *Semin Immunopathol*. 2019;41(1):5-19.
64. Hasselbalch HC. Perspectives on chronic inflammation in essential thrombocythemia, polycythemia vera, and myelofibrosis: is chronic inflammation a trigger and driver of clonal evolution and development of accelerated atherosclerosis and second cancer? *Blood*. 2012;119(14):3219-3225.
65. Hasselbalch HC. The role of cytokines in the initiation and progression of myelofibrosis. *Cytokine Growth Factor Rev*. 2013;24(2):133-145.
66. Hasselbalch HC. Chronic inflammation as a promotor of mutagenesis in essential thrombocythemia, polycythemia vera and myelofibrosis. A human inflammation model for cancer development? *Leuk Res*. 2013;37(2):214-220.
67. Hasselbalch HC, Bjørn ME. MPNs as inflammatory diseases: the evidence, consequences, and perspectives. *Mediators Inflamm*. 2015;2015:1-16.
68. Fleischman AG. Inflammation as a driver of clonal evolution in myeloproliferative neoplasm. *Mediators Inflamm*. 2015;2015:1-6.
69. Čokić VP, Mitrović-Ajić O, Beleslin-Čokić BB, et al. Proinflammatory cytokine IL-6 and JAK-STAT signaling pathway in myeloproliferative neoplasms. *Mediators Inflamm*. 2015;2015:1-13.
70. Gisslinger H, Klade C, Georgiev P, et al. Maintenance of Response in Long-Term Treatment with RopergInterferon Alfa-2b (Besremi®) vs. Hydroxyurea in Polycythemia Vera Patients (Proud/Continuation - PV Phase III Trials). Poster Presentation: 1457. EHA June 15, 2019.
71. Theocharides A, Passweg JR, Medinger M, et al. The allele burden of JAK2 mutations remains stable over several years in patients with myeloproliferative disorders. *Haematologica*. 2008;93(12):1890-1893.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pedersen RK, Andersen M, Knudsen TA, et al. Data-driven analysis of JAK2V617F kinetics during interferon-alpha2 treatment of patients with polycythemia vera and related neoplasms. *Cancer Med*. 2020;00:1–13.

<https://doi.org/10.1002/cam4.2741>

HSC niche dynamics in regeneration, pre-malignancy and cancer: Insights from

mathematical modeling

Abstract

The hematopoietic stem cell niche plays a key role in regeneration and malignancy. Due to the high complexity of the involved processes and their limited experimental accessibility, our quantitative understanding is still limited. We propose a mathematical model that links processes in the hematopoietic stem cell (HSC) niche to observable cell population dynamics. The model is based on the assumption that HSC bind reversibly to the niche. Upon detachment HSC are activated and prone to division or differentiation. The model is calibrated to experimental data. We use a combination of computer simulations and mathematical analysis to investigate how detachment of HSC from the niche, attachment to the niche, cell proliferation and cell differentiation impact blood regeneration, clonal selection and progression of malignancies. Our study suggests that monitoring of cells in peripheral blood provides only limited insights into dynamics and clonal composition of the stem cell niche. We discuss possible implications of this finding for disease monitoring approaches that are based on peripheral blood sampling. We use the developed model to evaluate how niche dynamics can be exploited to eradicate malignant clones using low-toxicity chemotherapy and how genetically engineered HSC can colonize the niche sustainability.

1. Introduction

Over the last years hematopoiesis has emerged as a paradigm of stem-cell-driven tissue maintenance, regeneration and malignancy^{1,2}. Hematopoietic stem cell (HSC) function crucially depends on specialized microenvironments, referred to as stem cell niches^{3,4}. Through complex signaling networks and cell-cell contacts, the niche contributes to HSC maintenance and regulation⁴.

A broad spectrum of diseases and clinical interventions is closely intertwined with dynamics in the stem cell niche. These include acute myeloid leukemia (AML), chronic myeloid leukemia (CML) or the philadelphia-negative myeloproliferative neoplasms (MPN) which are prototypic stem cell driven diseases^{5,9}. There is

Running title: "HSC niche in regeneration, pre-malignancy & cancer"

5

Rasmus Kristoffer Pedersen¹, Morten Andersen¹, Vibe Skov³, Lasse Kjær³, Hans C.

Hasselbalch³, Johnny T. Ottesen¹, and Thomas Stiehl^{2,4#}

¹IMFUFA, Department of Science and Environment, Roskilde University, Denmark

²Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany

³Department of Haematology, Zealand University Hospital, Denmark

#corresponding author

Corresponding author:

Thomas Stiehl, MD, PhD,

Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany.

Email: Thomas.Stiehl@iwr.uni-heidelberg.de

40 evidence that malignant stem cells either outcompete HSC from the niche or impair niche function¹⁰⁻¹⁶.

Similarly, clonal hematopoiesis of indeterminate potential (CHIP) is related to the expansion of mutated HSC in the niche^{17,18}. As cancer stem cells (CSC) are the target of curative treatment, a quantitative understanding of their interaction with the stem cell niche is of great clinical value.

Therapy monitoring of many hematologic malignancies depends on minimal residual disease (MRD)¹⁹. In clinical practice MRD is measured mostly in peripheral blood or in bulk marrow samples. To understand how far such measurements allow for inference of the CSC burden, a quantitative understanding of processes in the stem cell niche and their link to population dynamics of more mature cell types is required¹⁰⁻¹⁸.

A range of clinical interventions interferes with processes in the stem cell niche. Examples are HSC mobilization²⁰⁻²², stem cell transplantation²³⁻²⁵ and HSC-based gene therapy²⁶⁻²⁹. In the latter case genetically modified HSC have to be sustainably introduced in the niche.

Due to the limited accessibility of human HSC and the impossibility to maintain them in culture³⁰⁻³², the knowledge about *in vivo* stem cell dynamics is limited. Time-lapse imaging³³⁻³⁴, transplantation experiments^{12,35}, cell-tracking³⁶ and co-culture studies³⁷ have contributed to what we currently know about the stem cell niche. This knowledge can be complemented by mathematical models that help relate processes in the stem cell niche to observable dynamics^{31,35,38-40}.

The use of mathematical models has a long tradition in hematology⁴¹⁻⁴⁶. Mathematical modeling is a powerful tool to explore the dynamics of biological processes and to rigorously compare different hypotheses^{24,44,47}. Mathematical models have provided valuable insights in stem cell dynamics in health^{38,40,45,46,48-50} and disease^{51-64,86,87}.

60 In this work, we develop a novel mathematical model of the hematopoietic stem cell niche. The model is calibrated to data from literature and its predictions are compared to experimental results. Our aim is to

study how key processes such as attachment / detachment of stem cells from the niche, stem cell proliferation, self-renewal and differentiation impact population dynamics of healthy and malignant cells. We will use the model to provide insights into the following questions which are closely related to the above mentioned clinical scenarios.

- 65 1. What determines if a stem cell population can persist in the niche? Which mechanisms confer competitive advantages to stem cell clones?
 2. Under which conditions can multiple stem cell populations coexist in the niche?
 3. How can new cell populations be introduced and sustained in the niche?
 4. How can cell populations be eradicated from the niche?
 - 70 5. Are peripheral blood samples sufficient to infer which clones dominate in the stem cell niche?
- We demonstrate that the impact of processes in the stem cell niche on cell population dynamics is complex and partially counterintuitive. We point out pitfalls arising from intuitive interpretation of experimental results.
- 75 In particular, our model implies that the fraction of mutated cells in peripheral blood does not necessarily reflect the fraction of mutated cells in the stem cell niche. This has considerable consequences for the conclusions drawn based on peripheral blood sampling or unsorted bone marrow aspirates. The model allows for quantification of stem cell properties linked to selective advantages and provides ideas how it could be assessed experimentally. Furthermore, we use the proposed model to explore under which conditions a combination of mobilizing agents, stem cell transplantation and low-dose chemotherapy is a suitable intervention for elderly AML patients not eligible for high dose chemotherapy.

2 Methods

2.1 Mathematical Modeling

We have developed a mathematical model describing cellular interactions in the HSC niche. The model allows us to study how processes such as stem cell detachment from the niche, stem cell attachment to the niche, stem cell proliferation or stem cell differentiation impact the dynamics of blood cell formation in health and disease. The model is summarized in Figure 1. A detailed model description is provided in section A.1 of the

supplementary material.

In our model, HSC bind reversibly to specialized niche cells. We assume that the niche has a fixed capacity⁶⁵. The niche capacity corresponds to the maximal number of HSC that can bind to the niche. Upon detachment stem cells become activated⁶⁶⁻⁷¹ and either divide and give rise to more stem cells or they irreversibly differentiate into progenitor cells. Unlike stem cells, progenitor cells possess only limited self-renewal capacity and become increasingly committed towards a specific blood cell lineage.

In the model, the stem cells arising from division of a stem cell enter an inactive state. The inactive stem cells can either differentiate or rebind to the niche. The distinction between active (in the sense of HSC just detached from the niche with the potential to divide) and inactive stem cells (i.e., HSC originating from

division which either differentiate or have to reattach to the niche to maintain stemness) is new and has

The considered processes are quantified by a set of model parameters. The detachment rate describes which fraction of HSC detaches from the niche per unit of time. Analogously, the attachment rate quantifies the attachment of HSC to the niche. Proliferation of active HSC and differentiation are quantified by the proliferation and differentiation rates. The model takes into account active stem cells (outside the niche),

inactive stem cells (outside the niche), niche-bound stem cells and empty niche spaces. Model parameters

of the healthy system have been chosen based on literature and by fitting the model to experiments. Details

卷之三

A version of the model that takes into account the competition between two different stem cell clones is depicted in Figure 1b. A clone consists of stem cells that have identical genome and hence identical properties

3 Results

110 3.1 Increased attachment of HSC to the niche does not reduce circulating HSC cell counts in

homeostasis

Cell counts in homeostasis arise from a complex interplay between the considered processes. The model allows for a systematic study of how each process impacts the steady state cell counts. The results are summarized in Table 1 and in Figure 2. Many of the observed relations are as intuitively expected: Increased stem cell proliferation leads to an increase of HSC inside and outside the niche. Increased differentiation leads to a reduction. Furthermore, an increase in the number of niche spaces leads, as experimentally observed,⁶⁵ to an increase of stem cell numbers.

conclusions from experiments of Bhattacharya et al⁷². Furthermore, in the model a certain percentage of HSC is detached from the niche under homeostatic conditions, as observed in healthy individuals⁷³⁻⁷⁵. The system is in a dynamic equilibrium between attachment of stem cells to the niche and detachment of stem cells from the niche. The higher the attachment rate of HSC, the lower the number of free niches in the equilibrium. The ratio between active and inactive stem cells is independent of the detachment rate.

The model reveals that an increased attachment rate of stem cells to the niche not only increases the equilibrium number of HSC outside and inside the niche. This results from the following mechanism, see

Figure 3: An increased binding rate leads to an increase of niche-bound HSC, which is accompanied by a temporary decrease of free HSC. However, if the detachment rate is unchanged, the number of cells detaching from the niche increases with the number of cells in the niche. For this reason, the number of free stem cells also increase on the long time scale. This implies that increased attachment rates of HSC do not lead to a decrease of free HSC counts in homeostasis. The detailed mathematical analysis leading to these conclusions is presented in section A.9.

3.2 The homeostatic number of circulating HSC does not reflect the number of niche-bound HSC

The model reveals that the rate of HSC detachment from the niche does not impact the number of free niches in homeostasis. The reason for this is that detached stem cells are prone to divide and thus increase the number of stem cells outside the niche, which, in turn, leads to an increased number of cells binding to the niche per unit of time. This counterbalances the increased number of cells detaching from the niche per unit of time, see Figure 4. For this reason, measurements of unbound stem cells alone are not sufficient to monitor stem cell counts in the niche. In particular, if increased homeostatic counts of unbound stem cells are detected this can result either from an increased detachment rate or from an increased attachment rate of HSC to the niche, see Figures 3 and 4. In the first case the equilibrium number of HSC bound to the niche increases in the latter is remains constant.

3.3 Homeostatic progenitor production decreases with increasing stem cell differentiation rate

To maintain the homeostatic cell counts the hematopoietic system has to replace cells cleared from the blood stream. Due to the limited self-renewal capacity of progenitors, the mature cell output scales with the influx to the progenitor compartment, i.e. with the number of differentiating HSC.

Parameter-changes that increase the homeostatic stem cell count, i.e., increased niche capacity or increased stem cell proliferation, also increase the steady state production of progenitors, see Table 1 and Figure 2.

150 Furthermore the production of progenitors increases for increasing detachment rates of stem cells, since this is linked to an increase of actively dividing and differentiating HSC.

Notably, homeostatic progenitor production decreases for increasing HSC differentiation rates. This occurs because an increase of differentiation leads to a reduction of the total stem cell pool in homeostasis, see Figure 5. Transiently the output of progenitors increases, however due to depletion of HSC, it decreases over time. These findings are in line with previously reported results⁴².

3.4 The competitive advantage of a stem cell clone depends on proliferation rate, differentiation rate and niche binding rate

Diseases such as MPN and AML involve expansion of a mutated clone at the expense of healthy stem cells^{1,13,15}.

160 The model allows us to systematically study which cell properties help a stem cell clone to out-compete another stem cell clone from the niche. Our model implies that an increase of the stem cell proliferation rate, a decrease of the stem cell differentiation rate or an increased niche binding rate of the mutated HSC clone compared to the healthy HSC result in a competitive advantage of the mutated clone. The respective analysis is presented in section A.7 of the supplement.

165 However, disadvantageous changes of one cell property can be counterbalanced by advantageous changes of other cell properties. To systematically study how each of the considered cell properties (proliferation rate, differentiation rate, binding rate, detachment rate) contributes to the competitive advantage of a clone, we derived an expression that allows for quantification of a clone's fitness. We define fitness as the ability of a clone to invade the stem cell niche and out-compete other stem cell clones. The fitness is given as

- 190 of free niche spaces in a monoclonal steady state is a measure for the fitness of the respective stem cell clone. Less empty niche spaces in homeostasis of a monoclonal stem cell population, increase the ability of the respective stem cell clone to invade the stem niche in a competitive setting. The higher the number of vacant spaces, the lower the fitness of the respective clone. Due to this relation, the fitness of a clone can be determined from the free niche spaces in a reference scenario in absence of other competing clones.
- 195 Potentially, co-culture experiments of stem and niche cells could allow for quantifying the fitness of a stem cell clone and thus to predict the long-term outcome of clonal competition.
- 3.6 Chimerism in the peripheral blood does not necessarily reflect chimerism in the stem cell niche
- 200 The HSC detachment rate from the niche is closely related to stem cell proliferation and differentiation. Therefore, it is an important determinant of the production of progenitors. Since, as argued above, the stem cell detachment rate has no impact on the competitive advantages of a clone, clones with equal fitness can have very different detachment rates and thus contribute very differently to the production of progenitors. This implies that the chimerism (e.g., after allogeneic transplantation) in the stem cell niche is not reflected by the chimerism in the blood stream. Similarly MRD measurements in peripheral blood or bulk bone marrow samples are not representative for the CSC burden in the stem cell niche.
- 205 In particular, a mutant clone that is more likely to detach from the niche but has identical proliferation, binding and differentiation rates as the healthy HSC can lead to high numbers of mutated circulating cells, without out-competing the healthy clone. In Figure 7, this is illustrated by comparing two clones with equal fitness but different detachment rates. This finding might be relevant in the context of clonal hematopoiesis where measurements from peripheral blood, such as *JAK2V617F* or *CALR* burden, are used to draw conclusions about the composition of the stem cell compartment^{76,77}. These examples illustrate that an
- 175 supplement. When two clones compete for the niche, the clone with the inferior fitness will be out-competed from the niche by the clone with the superior fitness. The above relation shows how a disadvantage in one property can be compensated by an advantage in another property.
- In case of a clinically overt leukemia, this implies that the fitness $F_L = \frac{b_L}{d_{I_L}} \frac{(r_L - d_{A_L})}{(r_L + d_{A_L})}$ of a leukemic clone is larger than the fitness $F_H = \frac{b_H}{d_{I_H}} \frac{(r_H - d_{A_H})}{(r_H + d_{A_H})}$ of the HSC.
- 180 Long-term coexistence of two clones is possible if and only if their fitness is the same. We note that it does not depend on the niche detachment rate whether a given clone possesses the ability to out-compete another clone from the niche. Consequently, treatment approaches solely relying on mobilization of malignant cells from the niche cannot be successful. However, if a given clone has the ability to out-compete another clone from the niche, the time to extinction of the inferior clone decreases with increasing detachment rate of the superior clone, see Figure 6.
- 185 3.5 The number of empty niche spaces in monoclonal homeostasis is a measure for the stem cell fitness
- If the niche is colonized by a single stem cell clone, the number of free niche spaces in homeostasis is given as $1/F$, where F is the fitness introduced above (see supplementary section A.6). This implies that the number

analysis of the primitive bone marrow cells can be superior to quantifications based on peripheral blood samples. Furthermore, this finding suggests that mutations increasing the progenitor or mature cell output of a clone do not necessarily increase the aggressiveness of the disease in terms of stem cell fitness.

The model leads to an analogous conclusion in the context of gene therapy. HSC mobilization prior to transplantation of engineered cells can increase the homing of the transplanted cells to the niche in the nonpreconditioned but not in the preconditioned host. Details are described in supplementary section A.12.

4 Discussion

3.7 Multiple doses of stem cell transplantation and mobilizing agents can be beneficial in the unpreconditioned host

Processes in the stem cell niche significantly contribute to engraftment after stem cell transplantation. Mouse experiments suggest that in unpreconditioned hosts the amount of transplanted stem cells homing to the niche increases if the transplant is split into multiple doses administered at subsequent days. This finding can be reproduced by our model. Model simulations suggest that this effect decreases for increasing doses of preconditioning. Therefore, this approach does not improve the outcome of stem cell transplants after high dose chemotherapy. A similar result has been obtained in a clinical trial⁷⁸. A detailed description is provided in supplementary section A.11.

Stem cell mobilizing agents such as pegfilgrastim or AMD3100 can perturb processes in the stem cell niche⁷⁹.

Our model allows for developing hypotheses in which settings mobilizing agents could be used to further optimize clinical procedures and under which conditions they may lead to unfavorable outcomes. High dose chemotherapy is a major limitation in the treatment of elderly patients. For this reason we asked whether the combination of stem cell mobilization from the niche (using mobilizing agents) followed by the transplantation of HSC could at least partially remove leukemic cells from the niche. The model suggests that this could be possible; however, the reduction of malignant cells due to administration of mobilization agents might be relatively small (less than 10%). Furthermore, if mobilization agents are administered for a longer period, malignant cells mobilized from the niche could be triggered to proliferate which results in an increase of disease burden.

240 The stem cell niche is involved in key processes during regeneration and malignancy. However, due to its limited experimental accessibility our quantitative understanding of the underlying processes remains incomplete. To close this gap, we have developed a novel mathematical model describing interactions of HSC with their niche. We assume that HSC can reversibly bind to a bone marrow niche of fixed capacity. HSC bound to the niche are considered quiescent^{66,67,81-83}. Upon detachment from the niche HSC become activated and are prone to divide or to differentiate⁶⁶⁻⁷¹. We assume that HSC can perform only a limited number of divisions before they either differentiate or have to reattach to the niche. We fitted the model to murine data from the literature and compared simulations to experimental observations.

245 The model suggests that homeostatic cell counts result from the complex interplay of multiple processes, i.e., HSC detachment from the niche, attachment to the niche, proliferation and differentiation. Our work has led to the following new insights.

Changes of cell properties as they occur in disease or experiments may have opposite effects on the short time scale than they have on the long time scale. For example, if the attachment rate of HSC to the niche is increased while all other process parameters remain unchanged, we first observe a decrease of circulating HSC counts, since more HSC attach to the niche. An increased attachment to the niche acts against HSC loss due to differentiation or cell death. For this reason the HSC populations expands over time. If the HSC detachment rate remains unchanged the number of HSC detaching per unit of time increases with the number of niche bound cells. This leads to a new steady state of the system with an increase of circulating HSC counts in homeostasis. Similarly an increased HSC detachment rate from the niche leads to an increase

of circulating and actively cycling HSC in homeostasis, however the homeostatic count of niche-bound HSC does not change. These findings imply that it is not possible to estimate the amount of niche bound stem cells based on circulating HSC in homeostasis.

In a multi-clonal setting the numbers of progenitor or mature cells that are derived from a stem cell clone do not necessarily correlate with the number of stem cells belonging to the clone. Depending on the respective attachment, detachment, proliferation and differentiation rates a small number of stem cells can give rise to a large amount of circulating cells or vice versa. This conclusion is in line with recent results^{15,36}. It is important to note that for some specific mutations, such as CAIR or JAK2/617F, a correlation between clonal composition of the stem cells and the circulating cells has been observed^{76,84}, however, depending on the presence of other mutations this may not be true in general.

In line with previous studies the model suggests that high niche attachment and small stem cell differentiation rates confer a competitive advantage^{39,54,57,55}. Interestingly, the detachment rate from the niche has no impact on the competitive fitness of a clone.

Our analytical results suggest that in case of a monoclonal HSC population, the percentage of niche spaces which are occupied in homeostasis is a measure for the fitness of the clone, i.e. the higher the number of occupied niche spaces, the higher the competitive advantage of the clone in a multi-clonal environment. This implies that the competitive advantage could be experimentally quantified in absence of clonal competition. This is a novel and experimentally testable hypothesis. If the hypothesis holds, the competitive advantage of a clone could be assessed by letting it colonize an irradiated host, a humanized mouse⁸⁵, or an appropriate *in vitro* niche.

From the clinical point of view it is of interest to introduce stem cell clones to the niche, e.g., in case of HSC gene therapy and to eradicate clones from the niche, e.g., during cancer treatment. A testable hypothesis generated by our model is that careful administration of stem cell mobilizing agents could potentially help

with eradication of malignant clones from the niche or with introducing new, genetically modified clones in the niche.

A limitation of the model in its present form is that it considers only one type of stem cell niches, neglecting 260 cells that HSC reside at different locations in the bone marrow (endosteal or perivascular) that might differ with respect to their properties. Furthermore, we neglect turnover of niche cells. If experimental data on these processes become available, they can be easily incorporated in the model. It is important to note that the presented calibration has to be understood as a proof-of-principle example. Depending on the specific experimental protocols, measurements of a given quantity can vary considerably^{72,80}.

265 Taken together, our work contributes to the mechanistic understanding of the stem cell niche in health and disease. It provides a computational framework to design experiments and to interpret clinical data.

5 Acknowledgements

Parts of this work have been completed during a visit of RP at Heidelberg University. The authors thank Anna Marciniak-Czochra (Heidelberg University, Germany) for her support in realizing this research stay, for 270 welcoming RP in her group and for providing the required infrastructure.

275

1. Hoggatt J, Kfouri Y, Scadden DT. Hematopoietic Stem Cell Niche in Health and Disease. *Annu. Rev. Pathol. Mech. Dis.* 2016;11(1):555–581.
2. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001;414(6859):105–111.
3. Pinho S, Frenette PS. Haematopoietic stem cell activity and interactions with the niche. *Nat. Rev. Mol. Cell Biol.* 2019;20(5):303–320.
4. Scadden DT. The stem-cell niche as an entity of action. *Nature*. 2006;441(7097):1075.
5. Lutz C, Hoang VT, Buss E, Ho AD. Identifying leukemia stem cells – Is it feasible and does it matter? *Cancer Lett.* 2013;338(1):10–14.
6. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* 1997;3(7):730.
7. Holyoake TL, Vetrici D. The chronic myeloid leukemia stem cell: stemming the tide of persistence. *Blood*. 2017;129(12):1595–1606.
8. Mead AJ, Mullally A. Myeloproliferative neoplasm stem cells. *Blood*. 2017;129(12):1607–1616.
9. Spivak JL. Myeloproliferative Neoplasms. *N. Engl. J. Med.* 2017;376(22):2168–2181.
10. Plaks V, Kong N, Werb Z. The Cancer Stem Cell Niche: How Essential Is the Niche in Regulating Stemness of Tumor Cells? *Cell Stem Cell*. 2015;16(3):225–238.
11. Wang W, Stiehl T, Raffel S, et al. Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia. *Haematologica*. 2017;102(9):1567–1577.
12. Ishikawa F, Yoshida S, Saito Y, et al. Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nat. Biotechnol.* 2007;25(11):1315.
13. Boyd AL, Campbell CJ V, Hopkins CJ, et al. Niche displacement of human leukemic stem cells uniquely allows their competitive replacement with healthy HSPCs. *J. Exp. Med.* 2014;211(10):1925–1935.
- 320 14. Colmone A, Amorim M, Pontier AL, et al. Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science*. 2008;322(5909):1861–1865.
15. Stiehl T, Wang W, Lutz C, Marciniak-Czochra A. Mathematical modeling provides evidence for niche competition in human AML and serves as a tool to improve risk stratification. *Cancer Res.* doi: 10.1158/0008-5472.CAN-20-0283. Epub ahead of print.
- 325 16. Ghobrial IM, Detappe A, Anderson KC, Steensma DP. The bone-marrow niche in MDS and MGUS: implications for AML and MM. *Nat. Rev. Clin. Oncol.* 2018;15(4):17. Jan M, Ebert BL, Jaiswal S. Clonal hematopoiesis. *Semin. Hematol.* 2017;54(1):43–50.
18. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9–16.
- 330 19. Buccisano F, Maurillo L, Schuurhuis GJ, et al. The emerging role of measurable residual disease detection in AML in morphologic remission. *Semin. Hematol.* 2019;56(2):125–130.
20. To LB, Levesque JP, Herbert KE. How I treat patients who mobilize hematopoietic stem cells poorly. *Blood*. 2011;118(17):4530–4540.
21. Ferraro F, Lympéri S, Méndez-Ferrer S, et al. Diabetes impairs hematopoietic stem cell mobilization by altering niche function. *Sci. Transl. Med.* 2011;3(104):104ra101–104ra101.
- 335 22. Domingues MJ, Nilsson SK, Cao B. New agents in HSC mobilization. *Int. J. Hematol.* 2017;105(2):141.
23. Pedrazzoli P, Ledermann JA, Lotz J-P, et al. High dose chemotherapy with autologous hematopoietic stem cell support for solid tumors other than breast cancer in adults. *Ann. Oncol.* 2006;17(10):1479–

- 1488.
- 340 24. Das M. Stem cell transplantation in multiple myeloma. *Lancet Oncol.* 2017;18(5):e252–e252.
- 350 25. Kasim AA, Savani BN. Hematopoietic stem cell transplantation for acute myeloid leukemia: A review. *Hematol. Oncol. Stem Cell Ther.* 2017;10(4):245–251.
- 360 26. Morgan RA, Gray D, Lomova A, Kohn DB. Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned. *Cell Stem Cell.* 2017;21(5):574–590.
- 365 27. Auti A, Slavin S, Aker M, et al. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science.* 2002;296(5577):2410–2413.
- 370 28. Cowan MJ, Dvorak CC, Long-Boyle J. Opening Marrow Niches in Patients Undergoing Autologous Hematopoietic Stem Cell Gene Therapy. *Hematol. Oncol. Clin. North Am.* 2017;31(5):809–822.
- 375 29. Bernardo ME, Auti A. The Role of Conditioning in Hematopoietic Stem-Cell Gene Therapy. *Hum. Gene Ther.* 2016;27(10):741–748.
- 380 30. Derakhshani M, Abbaszadeh H, Movassaghpoor AA, et al. Strategies for elevating hematopoietic stem cells expansion and engraftment capacity. *Life Sci.* 2019;232::
- 390 31. Kumar S, Geiger H. HSC Niche Biology and HSC Expansion Ex Vivo. *Trends Mol. Med.* 2017;23(9):799–819.
- 395 32. Zhang Y, Gao Y. Novel chemical attempts at ex vivo hematopoietic stem cell expansion. *Int. J. Hematol.* 2016;103(5):519–529.
- 400 33. Christodoulou C, Spencer JA, Yeh S-CA, et al. Live-animal imaging of native haematopoietic stem and progenitor cells. *Nature.* 2020;578(7794):278–283.
- 405 34. Kim S, Lin L, Brown GAJ, Hosaka K, Scott EW. Extended time-lapse in vivo imaging of tibia bone marrow transplantation)(Report). *Leukemia.* 2017;31(7):1582.
- 410 35. Perlin JR, Sporri A, Zon LI. Blood on the tracks: hematopoietic stem cell-endothelial cell interactions in homing and engraftment. *J Mol Med (Berl).* 2017;95(8):809–819.
- 415 36. Lee-Six H, Kent DG. Tracking hematopoietic stem cells and their progeny using whole-genome sequencing. *Exp. Hematol.* 2020;83:12–24.
- 420 37. Vaidya A, Kale V. Hematopoietic Stem Cells, Their Niche, and the Concept of Co-Culture Systems: A Critical Review. *J. Stem Cells.* 2015;10(1):13–31.
- 425 38. Ashcroft P, Manz MG, Bonhoeffer S. Clonal dominance and transplantation dynamics in hematopoietic stem cell compartments. *PLoS Comput. Biol.* 2017;13(10):e1005803.
- 430 39. Walenda T, Stiehl T, Braun H, et al. Feedback Signals in Myelodysplastic Syndromes: Increased Self-Renewal of the Malignant Clone Suppresses Normal Hematopoiesis (Feedback Signals in Myelodysplastic Syndromes). 2014;10(4):e1003599.
- 435 40. Becker NB, Günther M, Li C, Jolly A, Höfer T. Stem cell homeostasis by integral feedback through the niche. *J. Theor. Biol.* 2019;481:100–109.
- 440 41. Mackey MC. Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis. *Blood.* 1978;51(5):941–956.
- 445 42. Mardiniak-Czochra A, Stiehl T, Ho AD, Jager W, Wagner W. Modeling of asymmetric cell division in hematopoietic stem cells—regulation of self-renewal is essential for efficient repopulation.(Report). *Stem Cells Dev.* 2009;18(3):377.
- 450 43. Bélair J, Mackey MC, Mahaffy JM. Age-structured and two-delay models for erythropoiesis. *Math.*

- Biosci.* 1995;128(1–2):317–346.
44. Foley C, Mackey M. Dynamic hematological disease: a review. *J. Math. Biol.* 2009;58(1–2):285–322.
45. Stiehl T, Ho AD, Marciniak-Czochra A. The impact of CD34+ cell dose on engraftment after SCTs: Personalized estimates based on mathematical modeling. *Bone Marrow Transplant.* 2014;49(1):30–37.
- 385
46. Stiehl T, Marciniak-Czochra A. Characterization of stem cells using mathematical models of multistage cell lineages. *Math. Comput. Model.* 2011;53(7–8):1505–1517.
47. Stiehl T, Marciniak-Czochra A. How to Characterize Stem Cells? Contributions from Mathematical Modeling. *Curr. Stem Cell Reports.* 2019;5(2):57–65.
- 390
48. Werner B, Beier F, Hummel S, et al. Reconstructing the *in vivo* dynamics of hematopoietic stem cells from telomere length distributions. *Elife.* 2015;4(2015):..
49. Catlin SN, Busque L, Gale RE, Guttorp P, Abkowitz JL. The replication rate of human hematopoietic stem cells *in vivo*. *Blood.* 2011;117(17):4460–4466.
- 395
50. Manesso E, Teles J, Bryder D, Peterson C. Dynamical modelling of haematopoiesis: an integrated view over the system in homeostasis and under perturbation. *J. R. Soc. Interface.* 2013;10(80):20120817..
51. Wodarz D, Garg N, Komarova NL, et al. Kinetics of CLL cells in tissues and blood during therapy with the BTK inhibitor ibrutinib. *Blood.* 2014;123(26):4132.
52. Mon Pere N, Lenaerts T, Pacheco JM, Dingli D. Evolutionary dynamics of paroxysmal nocturnal hemoglobinuria.(Research Article). *PLoS Comput. Biol.* 2018;14(6):e1006133.
- 400
53. Olszen A, Tang M, Cortes J, et al. Dynamics of chronic myeloid leukemia response to dasatinib, nilotinib, and high-dose imatinib. *Haematologica.* 2014;99(11):1701–1709.
54. Stiehl T, Marciniak-Czochra A. Mathematical Modeling of Leukemogenesis and Cancer Stem Cell Dynamics. *Math. Model. Nat. Phenom.* 2012;7(4):166–202.
55. Watson CJ, Papula AL, Poon GYP, et al. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science (80-.).* 2020;367(6485):1449–1454.
- 405
56. Stiehl T, Baran N, Ho AD, Marciniak-Czochra A. Cell division patterns in acute myeloid leukemia stem-like cells determine clinical course: A model to predict patient survival. *Cancer Res.* 2015;75(6):940–949.
57. Stiehl T, Baran N, Ho A, Marciniak-Czochra A. Clonal selection and therapy resistance in acute leukemias: Mathematical modelling explains different proliferation patterns at diagnosis and relapse. *J R Soc Interface.* 2014;11(94):20140079.
- 410
58. Stiehl T, Ho AD, Marciniak-Czochra A. Mathematical modeling of the impact of cytokine response of acute myeloid leukemia cells on patient prognosis. *Sci. Rep.* 2018;(January):1–11.
59. Stiehl T, Lutz C, Marciniak-Czochra A. Emergence of heterogeneity in acute leukemias. *Biol. Direct.* 2016;11(1):1–19.
- 415
60. Kim PS, Lee PP, Levy D. Dynamics and potential impact of the immune response to chronic myelogenous leukemia. *PLoS Comput. Biol.* 2008;4(6):..
61. Sajid Z, Andersen M, Ottesen JT. Mathematical analysis of the Cancitis model and the role of inflammation in blood cancer progression. *Math. Biosci. Eng.* 2019;16(6):8268–8289.
- 420
62. Ottesen JT, Pedersen RK, Sajid Z, et al. Bridging blood cancers and inflammation: The reduced Cancitis model. *J. Theor. Biol.* 2019;465:90–108.
63. Andersen M, Sajid Z, Pedersen RK, et al. Mathematical modelling as a proof of concept for MPNs as a

- human inflammation model for cancer development. *PLoS One*. 2017;12(8):e0183620.
64. Clapp GD, Lepoutre T, El Cheikh R, et al. Implication of the Autologous Immune System in BCR-ABL Transcript Variations in Chronic Myelogenous Leukemia Patients Treated with Imatinib. *Cancer Res*. 2015;75(19):4053–4062.
65. Zhang J, Niu C, Ye L, et al. Identification of the haematopoietic stem cell niche and control of the niche size. *Nature*. 2003;425(6960):836–841.
66. Nie Y, Han Y-C, Zou Y-R. CXCR4 is required for the quiescence of primitive hematopoietic cells. *J. Exp. Med.* 2008;205(4):777–783.
67. Sugiyama T, Kohara H, Noda M, Nagasawa T. Maintenance of the Hematopoietic Stem Cell Pool by CXCL12-CXCR4 Chemokine Signaling in Bone Marrow Stromal Cell Niches. *Immunity*. 2006;25(6):977–988.
68. Schroeder MA, DiPersio JF. Mobilization of hematopoietic stem and leukemia cells. *J. Leukoc. Biol.* 2012;91(1):47–57.
69. Shen Z-H, Zeng D-F, Ma Y, et al. Are there any new insights for G-CSF and/or AMD3100 in chemotherapy of haematological malignants? *Med. Oncol.* 2015;32(12):262.
70. Shen Z-H, Zeng D-F, Kong P-Y, Mai Y-Y, Zhang X, AMD3100 and G-CSF disrupt the cross-talk between leukemia cells and the endosteal niche and enhance their sensitivity to chemotherapeutic drugs in biomimetic polystyrene scaffolds. *Blood Cells, Mol. Dis.* 2016;59:16–24.
71. Welschinger R, Liedtke F, Basnett J, et al. Plerixafor (AMD3100) induces prolonged mobilization of acute lymphoblastic leukemia cells and increases the proportion of cycling cells in the blood in mice. *Exp. Hematol.* 2013;41(3):293–302.e1.
72. Bhattacharya D, Czechowicz A, Ooi AGI, et al. Niche recycling through division-independent egress of hematopoietic stem cells. *J. Exp. Med.* 2009;206(12):2837–2850.
- 425 445 73. Sodickoff M. Colony-Forming Units in Rat Parabionts. *Radiat. Res.* 1970;44(3):758.
74. Seno S, Fang CH, Himei S, Hsueh CL, Nakashima Y. Hemopoietic Recovery in Bone Marrow of Lethally Irradiated Rats Following Parabiosis. *Acta Haematol.* 1976;55(6):321–331.
75. Udomsakdi C, Lansdorp P, Hogge D, et al. Characterization of primitive hematopoietic cells in normal human peripheral blood. *Blood*. 1992;80(10):2513–2521.
76. Takahashi K, Patel KP, Kantarjian H, et al. JAK2 p.V617F detection and allele burden measurement in peripheral blood and bone marrow aspirates in patients with myeloproliferative neoplasms. *Blood*. 2013;122(23):3784–3786.
77. Cavalloni C, Rumi E, Ferretti VV, et al. Sequential evaluation of CALR mutant burden in patients with myeloproliferative neoplasms. *Oncotarget*. 2017;8(20):33416–33421.
78. Landau H, Wood K, Chung DJ, Koehne G, Lendvai N, Hassoun H, Lesokhin A, Hoover E, Zheng J, Devlin SM, Giralt S. Fractionated stem cell infusions for patients with plasma cell myeloma undergoing autologous hematopoietic cell transplantation. *Leuk Lymphoma*. 2016;57(8):1781–5.
- 430 450 460 79. Nervi B, Link DC, DiPersio JF. Cytokines and hematopoietic stem cell mobilization. *J. Cell. Biochem.* 2006;99(3):690–705.
80. Shimoto M, Sugiyama T, Nagasawa T. Numerous niches for hematopoietic stem cells remain empty during homeostasis. *Blood*. 2017;129(15):2124–2131.
81. Karppova D, Bonig H. Concise Review: CXCR4/CXCL12 Signaling in Immature Hematopoiesis-Lessons From Pharmacological and Genetic Models. *Stem Cells*. 2015;33(8):2391–2399.

- 465 82. Kunisaki Y, Bruns I, Scheiermann C, et al. Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature*. 2013;502(7473):637–643.
83. Cashman J, Clark-Lewis I, Eaves A, Eaves C. Stromal-derived factor 1 inhibits the cycling of very primitive human hematopoietic cells in vitro and in NOD/SCID mice. *Blood*. 2002;99(3):792–799.
84. Kjaer L, Holmström MO, Cordua S, et al. Sorted peripheral blood cells identify CALR mutations in B-470 and T-lymphocytes. *Leuk. Lymphoma*. 2018;59(4):973–977.
85. Reinisch A, Hernandez DC, Schallmoser K, Majeti R. Generation and use of a humanized bone-marrow-475 ossicle niche for hematopoietic xenotransplantation into mice. *Nat. Protoc.* 2017;12(10):2169–2188.
86. Ottesen, J. T., Pedersen, R. K., Dam, M. J. B., Knudsen, T. A., Skov, V., Kjær, L., & Andersen, M. Mathematical Modeling of MPNs Offers Understanding and Decision Support for Personalized Treatment. *Cancer*, 2020;12(8):2119
- 475 87. Andersen, M., Hasselbalch, H. C., Kjær, L., Skov, V., & Ottesen, J. T. Global dynamics of healthy and cancer cells competing in the hematopoietic system. *Mathematical Biosciences*, 2020;326(108372):1–15.

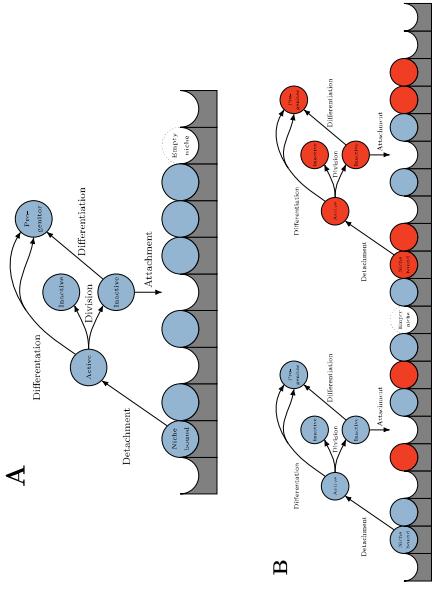


Figure 1: Illustration of the mathematical model. Stem cells are depicted as blue and red circles. Stem cell niches are illustrated by gray blocks. Panel (a) shows a scenario with a homogeneous stem cell population. Panel (b) illustrates a scenario with two stem cell clones competing for spaces in the joined niche. Arrows illustrate the processes considered in the model: Detachment from the niche, attachment to an empty niche space, self-renewing division and differentiation.

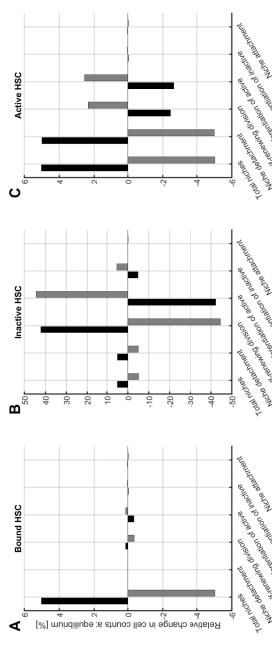


Figure 2: Impact of niche processes on homeostatic stem cell counts. The panels show how equilibrium stem cell counts change if a given parameter is increased or decreased by 5% compared to the fitted parameters provided in appendix A.3. Black boxes denote an increased parameter value while gray boxes denote a decreased parameter value. Note that the impact of some parameter changes is too small to be visible on the depicted scale. The number of niche-bound stem cells is independent from the detachment rate.

Increased parameter	Niche-bound N_H	Inactive I_H	Active A_H	Progenitors i_{D_H}
Niche capacity, K	↑	↑	↑	↑
Detachment from niche, u	0	↑	↑	↑
Division rate, r	↑	↑	*	↑
Differentiation of active, d_A	↓	↓	↓	↓
Differentiation of inactive, d_I	↓	↓	↓	↓
Attachment to the niche, b	↑	↑	↑	↑

Table 1: Summary of the effect of parameter changes on equilibrium cell counts, ↑ indicates that the value of the respective variable increases if the value of the respective parameter is increased, ↓ indicates that the value of the respective variable decreases if the value of the respective parameter is increased. *: Increasing only if division rate of activated stem cells, r , is close to the differentiation rate of activated stem cells, d_A , otherwise decreasing. See supplementary section A.9.

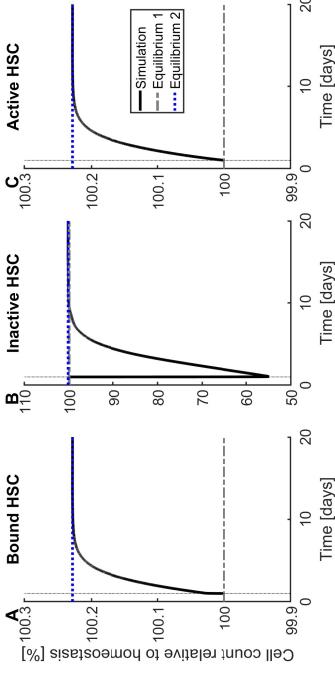


Figure 3: System dynamics in response to changes in HSC attachment rate. Starting at equilibrium, the HSC attachment rate is doubled at day 1. The system converges to a new equilibrium, with both higher free (activated and inactive) and higher niche-bound HSC counts. Equilibrium 1 refers to the homeostatic state and Equilibrium 2 refers to the equilibrium reached after the attachment rate has been increased. The observations can be explained as follows: A higher attachment rate leads to an increased binding of free cells to the niche and a consecutive temporal decrease of unbound HSC. However, the higher HSC number in the niche leads to an increase in the number of HSC detaching per unit of time and thus, on a longer time scale to an increase of free cells.

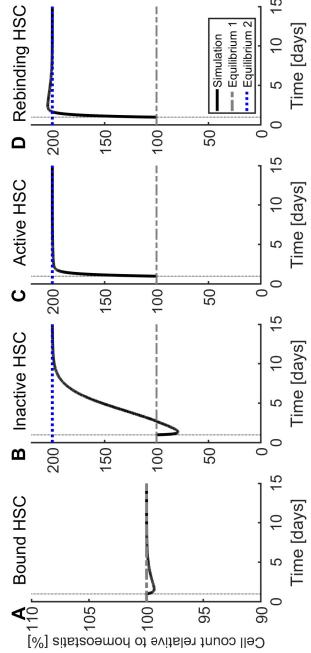


Figure 4: System dynamics in response to changes in HSC detachment rate. Starting at equilibrium, the detachment rate is doubled at day 1. The system converges to a new equilibrium, with both higher activated and inactive HSC, but an unchanged number of niche-bound HSC. Equilibrium 1 refers to the homeostatic state and Equilibrium 2 refers to the equilibrium reached after perturbation of parameter values. Panel D shows the amount of free stem cells binding to the niche per unit of time.

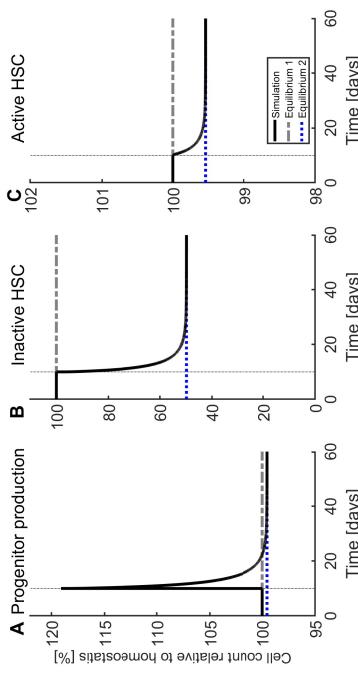
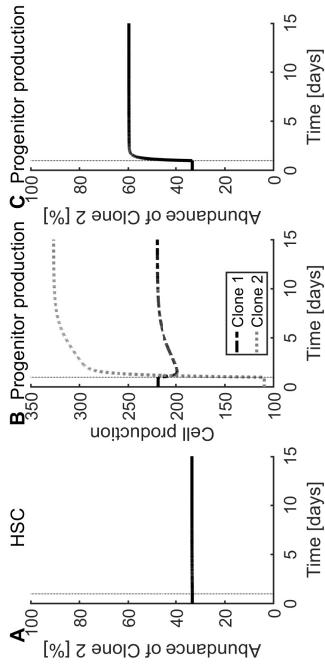
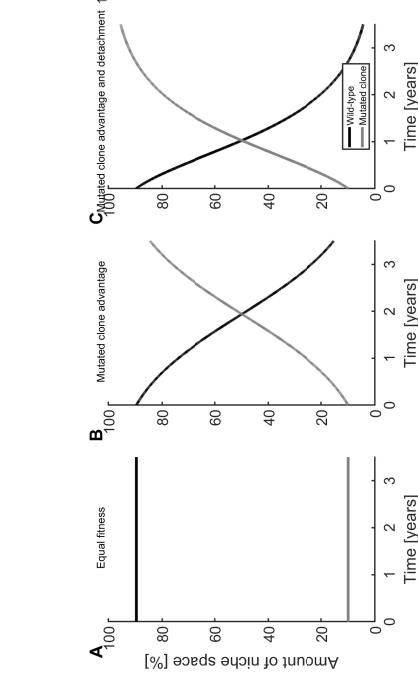


Figure 5: Production of progenitors following changes of the differentiation rate. Starting at equilibrium, the differentiation rate of inactive HSC is doubled at day 10. The system converges to a new equilibrium, with both lower activated and inactive HSC outside the niche. A temporary increase in the production of progenitors is seen, however, after some time a new homeostatic equilibrium with reduced progenitor outputs is approached.



2. Activated stem cells divide and differentiate at rates d_{A_H} and d_{A_L} . This can be written as

$$\begin{aligned} A_H &\xrightarrow{d_{A_H}} 2\mathcal{D}_H & (A.9) \\ A_L &\xrightarrow{d_{A_L}} 2\mathcal{D}_L & (A.10) \end{aligned}$$

corresponding to the loss of stem cells. Here, \mathcal{D}_E denotes a LSC-derived progenitor.

3. Activated cells divide into two inactive cells at rates r_H and r_L respectively. Inactive cells can either differentiate (at rates d_{I_H} and d_{I_L}) or reattach to the niche (at rates b_H and b_L):

$$A_H \xrightarrow{r_H} 2\mathcal{I}_H \quad (A.11)$$

$$\mathcal{I}_H \xrightarrow{d_{I_H}} \mathcal{D}_H \quad (A.12)$$

$$\mathcal{I}_H + \mathcal{N}_E \xrightarrow{b_H} \mathcal{N}_H \quad (A.13)$$

$$A_L \xrightarrow{r_L} 2\mathcal{I}_L \quad (A.14)$$

$$\mathcal{I}_L \xrightarrow{d_{I_L}} \mathcal{D}_L \quad (A.15)$$

$$\mathcal{I}_L + \mathcal{N}_E \xrightarrow{b_L} \mathcal{N}_L \quad (A.16)$$

where \mathcal{I}_E denotes an inactive LSC.

Denoting by A_E the abundance (measured e.g., as cells per kg of body weight) of activated LSC, by I_E that of inactive LSC, by D_E that of LSC-derived progenitors and by N_E that of niches occupied by LSC we obtain the following system of ODEs:

$$\frac{dN_H}{dt} = b_H(K - N_H - N_L)I_H - u_H N_H \quad (A.17a)$$

$$\frac{dI_H}{dt} = -b_H(K - N_H - N_L)I_H + 2r_H A_H - d_{I_H} I_H \quad (A.17b)$$

$$\frac{dA_H}{dt} = u_H N_H - r_H A_H - d_{A_H} A_H \quad (A.17c)$$

$$\frac{dN_L}{dt} = b_L(K - N_H - N_L)I_L - u_L N_L \quad (A.17d)$$

$$\frac{dI_L}{dt} = -b_L(K - N_H - N_L)I_L + 2r_L A_L - d_{I_L} I_L \quad (A.17e)$$

$$\frac{dA_L}{dt} = u_L N_L - r_L A_L - d_{A_L} A_L \quad (A.17f)$$

$$N_E = K - N_H - N_L, \quad (A.17g)$$

$$i_{D_H} = 2d_{A_H} A_H + d_{I_H} I_H, \quad (A.17h)$$

$$i_{D_L} = 2d_{A_L} A_L + d_{I_L} I_L \quad (A.17i)$$

with the initial conditions $N_E(0) \geq 0$, $N_H(0) \geq 0$, $N_L(0) \leq K$, $A_H(0) \geq 0$, $A_L(0) \geq 0$, $I_L(0) \geq 0$.

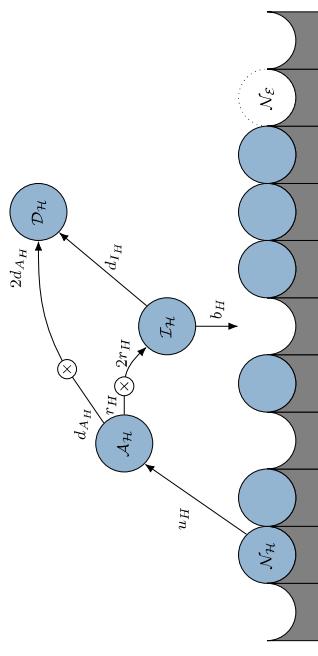
In the model both stem cell clones compete for the same niche spaces. Depending on the cell properties (proliferation rate, attachment rate, detachment rate etc.) one clone can out-compete the other clone from the niche.

A.2 Equilibria

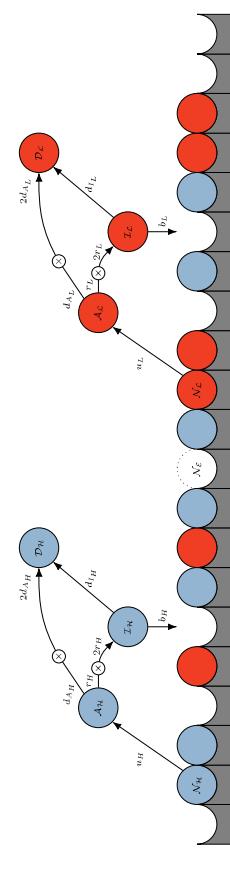
A basic analysis of the model identifies the equilibrium states. This analysis is provided in section A.6 of the supplement. The model possesses the following four equilibrium states.

- **Stem cell free equilibrium:** $N_H = N_L = I_H = A_H = I_L = A_L = 0$, $N_E = K$.

This equilibrium corresponds to the absence of stem cells and will lead to the death of the organism. This equilibrium also corresponds to the state after conditioning for bone marrow transplantation before the stem and progenitor cells are infused or to a situation with graft failure. This equilibrium exists for all parameter values.



(a) **Homogeneous stem cell population.** HSC detachment from the niche occurs at rate u_H , division of activated stem cells followed by differentiation at rate d_{A_H} , differentiation of inactive stem cells at rate d_{I_H} , division of active stem cells at rate r_H and attachment of inactive stem cells to the niche at rate b_H . Counts of niche-bound cells are denoted N_H , counts of inactive cells as I_H and counts of activated stem cells as A_H .



(b) **Competition of two stem cell clones.** Subscript H denotes the first (e.g., healthy) clone, while L denotes the second (e.g., leukemic) clone.
 Figure A1: Schematic representation of the model. (a) Homogeneous stem cell population, corresponding to a healthy scenario. (b) Competition of two stem cell populations, corresponding to AML, MPN, clonal hematopoiesis or following transplantation of donor cells or genetically engineered cells. In panel (b), we use the abbreviations introduced in panel (a). Empty niches are denoted by N_E . The index H refers to healthy cells, the index L to donor-related, mutated or malignant cells.

- **Homeostasis:** $N_E > 0$, $N_H > 0$, $I_H > 0$, $A_H > 0$, $N_L = I_L = A_L = 0$. This equilibrium corresponds to the homeostatic state of a healthy organism. All cells are wild-type cells. As shown in section A.6 of the supplement, this equilibrium only exists if activated stem cells divide more often than they differentiate and if generation of inactive stem cells outweighs their differentiation.
- **Out-competition of wild-type cells:** $N_E > 0$, $N_L > 0$, $I_L > 0$, $A_H > 0$, $N_H = I_H = A_H = 0$. This equilibrium corresponds to a state where all wild-type cells have been out-competed. In case of stem cell transplantation this means that all HSC are derived from the graft. In case of malignancy, this state corresponds to the out-competition of HSC by leukemic cells which leads to the death of the organism. In case of CHIP, this corresponds to monoclonal hematopoiesis. This equilibrium exists only if activated (mutated or respectively transplanted) stem cells divide more often than they differentiate and if generation of inactive stem cells outweighs their differentiation.
- **Coexistence in equilibrium:** In this equilibrium all population counts are positive. This corresponds to coexistence of wild-type and other cells. This type of equilibrium exists only for very specific parameter choices which is biologically unlikely to be realized.

Detailed calculations are provided in supplementary section A.6.

A.3 Parameterization and validation

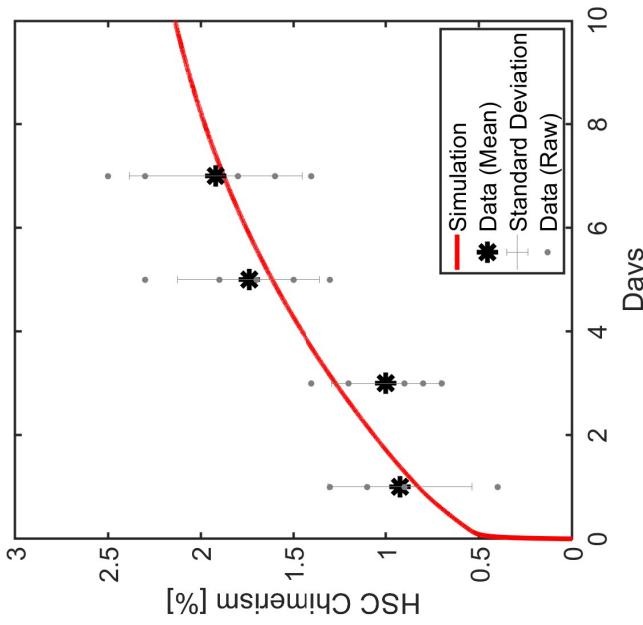
The number of niches is set to $K = 15000$, based on estimations in the work of Catlin et al. [8]. Model dynamics remain qualitatively unchanged if different values for K are chosen.

Stem cell niche dynamics can be experimentally studied based on transplantation of stem cells in unconditioned hosts [4]. The work from [4] provides evidence that consecutive transplantation of small HSC doses (12700/7 cells on 7 consecutive days) leads to an increased chimerism compared to the transplantation of the total cell dose (12700 cells) at a single time point. This observation suggests that transplanted stem cells home to niche spaces that are unoccupied under steady state conditions. To re-equilibrate the system releases random (i.e., donor and host-derived) stem cells from the niche. In case of sequential transplants this process is reiterated and donor-derived stem cells home to the niche at the expense of the randomly released host-derived cells. As a proof of concept, we demonstrate that our model can reproduce these experimental observations.

For this purpose we fit the model to the data of [4], using simulated annealing. The experimental data was extracted from Figure 6 of Bhattacharya et al. [4] based on visual inspection. Simulated annealing was performed using the MATLAB implementation due to Vandekerkhove [27] with the following settings: Initial temperature: 10^3 , Stop temperature: 10^{-3} , Max tries: 30. We assume that host and transplant cells have identical parameters, i.e. we set $r_H = r_L$, $d_{AH} = d_{AL}$, $u_H = u_L$, $b_H = b_L$ and $d_{IH} = d_{IL}$. Cell populations with index H are considered as host cells, cell populations with index L as cells of donor origin. At the time point of transplantation the host cell counts A_H , I_H , N_H are in equilibrium, and a number of free cells A_L and I_L is added, in accordance with the number of cells added in the work of Bhattacharya et al. [4]. As in the work of Stiehl et al. [23], we do not consider all transplanted cells to be stem cells, but reduce the number of cells added by one order of magnitude. Simulations indicate that if the transplanted cell count was much further reduced, the final chimerism in the model was smaller than in the experiments from Bhattacharya et al. [4], regardless of model parameterization. The transplanted stem cells added are assumed to be an equal mixture of activated and inactive cells, i.e., $A_L(0) = L(0)$, where $t = 0$ is the time of transplantation. For comparison with experimental data we calculate the chimerism at time t as $\frac{N_H(t)}{N_H(t) + N_L(t)} \cdot 100\%$.

The parameters obtained from the simulated annealing are given in Table A1. The model can fit the experimentally observed short-term (one week) and long term (16 weeks) dynamics, however with different parameter sets. The results are shown in Figures A2 and A3. Qualitative model dynamics are similar for both parameter sets. All depicted simulations are based on the parameter set that matches long-term dynamics. Using the other parameter set leads to the same qualitative results.

The number of HSC contained in CD34⁺ cell grafts and the fraction that homes to the bone marrow cannot be exactly quantified. For simulations of clinical scenarios we assume, in agreement with the data in Stiehl et al. [23], that a standard human graft contains $3 \cdot 10^3$ HSC per kg of body weight. We assume that HSC are equally distributed between activated and the inactive state. Numerical simulations imply that the qualitative results are robust with respect to changes in the number of transplanted activated and inactive HSC.



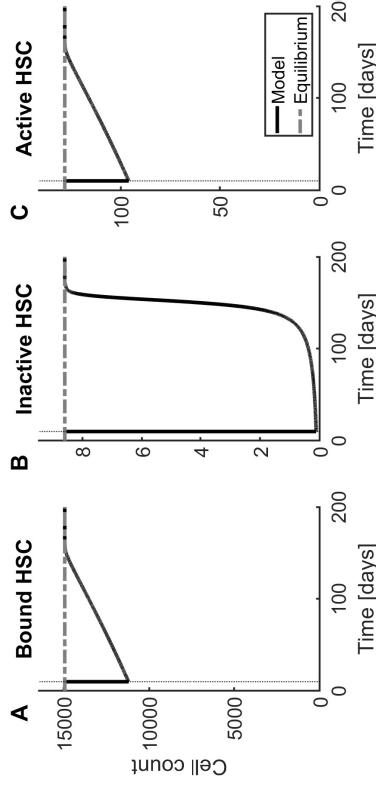
Supplemental figure A1: Simulated HSC chimerism (best fit) after single dose transplant in an unconditioned host. Black points depict the data from Bhattacharya et al. [4]. $R^2 = 0.86$. Parameter values are provided as "Short-term" in table A1.

	Short-term	Long-term		
K	15000 cells	u	0.004 day^{-1}	K
r	0.14 day^{-1}	b_I	0.09 day^{-1}	r
d_A	0.07 day^{-1}	d_I	2.39 day^{-1}	d_A
			2.32 day^{-1}	b_I
			2.06 day^{-1}	d_I
			0.96 day^{-1}	3.77 day^{-1}

A.4 Basic model properties

Perturbation of the homeostatic state

After perturbations, e.g., by blood loss, immune reaction, radiation, chemotherapy or bone marrow transplantation the hematopoietic system comes back to its equilibrium state. This property is captured by our model. When slightly perturbing homeostatic cell counts, the system returns to the homeostatic state. In supplementary section A.7 we prove this property rigorously.



Supplemental figure A1: System dynamics after perturbation.

Starting at the equilibrium state, 25 % of all cells are removed at time $t = 10$ days. The figure shows how the cell-counts in the model return to the equilibrium state. The dashed grey line corresponds to equilibrium cell counts, the black lines show the dynamical reaction of the system to the perturbation. A thin dotted line shows the time $t = 10$ days. Model parameters are specified in appendix A.3

HSC dynamics in absence of niche cells

In the absence of niche cells it holds $K = 0$ implying $N_H(0) = N_E(0) = 0$. Under this assumptions, as shown in Figure A5, A_H and I_H converge to zero. This is in line with experimental observations from cell culture experiments [12, 26, 31].

A.5 Equivalence of models with and without asymmetric HSC divisions

In the model derivation we have assumed that HSC always divide symmetrically. In the following we show that up to a reparametrization, this is equivalent to a model allowing symmetric and asymmetric HSC divisions.

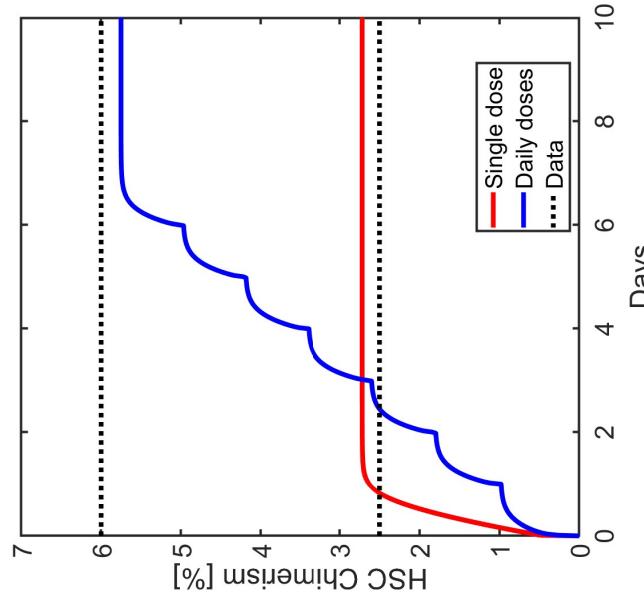
In a scenario where asymmetric HSC divisions are possible, the processes

$$\begin{aligned} A_H &\xrightarrow{r_H} 2T_H & (A.18) \\ A_H &\xrightarrow{a} T_H + D_H & (A.21) \\ A_H &\xrightarrow{d_H} 2D_H & (A.19) \end{aligned}$$

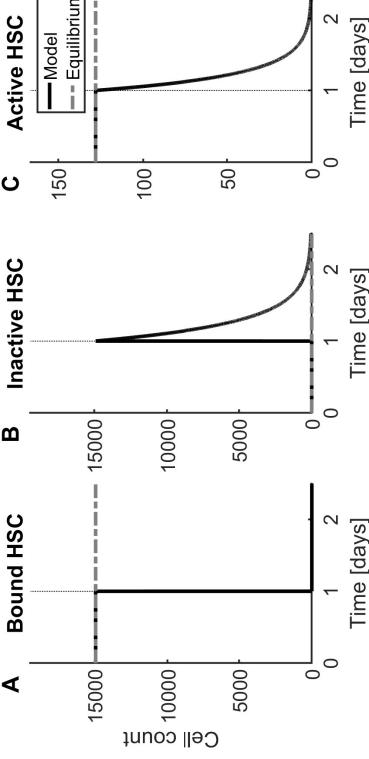
are replaced by

$$\begin{aligned} A_H &\xrightarrow{S_r} 2T_H & (A.20) \\ A_H &\xrightarrow{a} T_H + D_H & (A.21) \\ A_H &\xrightarrow{S_d} 2D_H & (A.22) \end{aligned}$$

where S_r is the rate of symmetric self-renewing divisions, S_d the rate of symmetric divisions resulting in differentiated cells, and a is the rate of asymmetric divisions.



Supplemental figure A3: Simulated HSC chimerism (best fit) after 7 small transplants on consecutive days compared to a single dose transplant. Hosts are unconditioned. Black dotted lines correspond to the chimerism after 16 weeks as measured in Bhattacharya et al. [4]. Parameter values are provided as “Long-term” in table A1.



Supplemental figure A5: System dynamics in absence of niche cells. Starting at the equilibrium state, we set $K = 0$ at time $t = 1$ day. This corresponds to a removal of the stem cell niche. In absence of the niche stem cells become extinct due to differentiation. The dashed grey line corresponds to the homeostatic cell counts. The black lines show model dynamics in absence of the niche. A thin dotted line shows the time $t = 1$ days. Model parameters are specified in appendix A.3.

The full mathematical model with explicit asymmetric divisions is then

$$\begin{aligned} \frac{dN_H}{dt} &= b_H(K - N_H)I_H - u_H N_H \\ \frac{dI_H}{dt} &= -b_H(K - N_H)I_H + (2S_r + a)A_H - d_{I_H} I_H \\ \frac{dA_H}{dt} &= u_H N_H - (S_d + a + S_d)A_H \\ N_E &= K - N_H \\ i_{D_H} &= (2S_d + a)A_H + d_{I_H} I_H \end{aligned} \quad \begin{aligned} (\text{A.23a}) \\ (\text{A.23b}) \\ (\text{A.23c}) \\ (\text{A.23d}) \\ (\text{A.23e}) \end{aligned}$$

Defining $r_H = S_r + \frac{a}{2}$, we have $\frac{dI_H}{dt} = -b_H(K - N_H)I_H + 2r_H A_H - d_{I_H} I_H$ as in equations (A.6). This implies that

$$\begin{aligned} \frac{dA_H}{dt} &= u_H N_H - (S_d + 2r_H - S_r)A_H \\ i_{D_H} &= (2S_d + 2r_H - 2S_r)A_H + d_{I_H} I_H. \end{aligned} \quad \begin{aligned} (\text{A.24}) \\ (\text{A.25}) \end{aligned}$$

Defining $d_{A_H} = S_d + r_H - S_r = S_d + \frac{a}{2}$ yields exactly the system of equations as (A.6). This demonstrates the equivalence of the model with and the model without asymmetric divisions. We note that the inequality $r_H > d_{A_H}$ is equivalent to $S_r > S_d$. Hence, the requirement for positive steady state counts of cells, can be interpreted such that the rate of symmetric self-renewing division, S_r , must exceed the rate of symmetric differentiating division, S_d , regardless of the absence or presence of asymmetric division.

A.6 Calculation of steady states

The existence and uniqueness of solutions follows from the Theorem of Picard-Lindelöf.

The system in equation (A.17a) has three equilibria that always exist, and a fourth which only exists given certain conditions. We present all four equilibria here.

Stem cell free equilibrium

The system always has a trivial steady state $E_0^* := (N_{H,E_0}^*, I_{H,E_0}^*, A_{H,E_0}^*, N_{L,E_0}^*, I_{L,E_0}^*, A_{L,E_0}^*)^T = (0, 0, 0, 0, 0, 0)^T$ implying $N_E^* = K$. This corresponds to the loss of all stem cells.

Monoclonal equilibria

There exist two equilibria at which one clone has become extinct and the other clone is present at positive quantities. We denote these equilibria by $E_H^* := (N_{H,E_H}^*, I_{H,E_H}^*, A_{H,E_H}^*, 0, 0, 0)^T$ and $E_L^* := (0, 0, 0, N_{L,E_L}^*, I_{L,E_L}^*, A_{L,E_L}^*)^T$. In the following we calculate E_H^* , the expressions for E_L^* are obtained by exchanging indices H and L .

Observe that $\dot{N}_H + \dot{I}_H + \dot{A}_H = (r_H - d_{A_H})A_H - d_{I_H} I_H$. Thus, the steady states for I_H and A_H are related through

$$I_H^* = \frac{r_H - d_{A_H}}{d_{I_H}} A_H^*. \quad (\text{A.26})$$

We note that for $I_H^* > 0$ and $A_H^* > 0$ to hold, it is necessary that $r_H > d_{A_H}$. This corresponds to the biological requirement that proliferation rate of activated stem cells has to be larger than their differentiation rate. It is intuitive that this is required for maintenance of the cell population, since in a scenario where differentiation outweighs proliferation the cell number will decline to zero.

Substituting this relation in $0 = \dot{I}_H$ yields $0 = -b_H(K - N_H^*)I_H^* + 2r_H \left(\frac{d_{I_H}}{r_H - d_{A_H}} \right) I_H^* - d_{I_H} I_H^*$. Trivially, this holds for $I_H^* = 0$, which implies $A_H^* = 0$ and in turn $N_H^* = 0$. This is the stem cell free steady state.

Assuming $I_H^* \neq 0$ implies $0 = -b_H K + b_H N_H^* + \frac{2r_H d_{A_H}}{r_H - d_{A_H}} - d_{I_H} \frac{r_H - d_{A_H}}{r_H - d_{A_H}}$, which thus yields an expression for N_H in the non-trivial steady state: $N_H^* = K - \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})}$. This allows us to find the equilibrium of activated stem cells, A_H^* , since:

$$0 = \dot{A}_H \quad (\text{A.27})$$

$$0 = u_H N_H^* - r_H A_H^* - d_{A_H} A_H^* \quad (\text{A.28})$$

$$A_H^* = \frac{u_H - r_H}{r_H + d_{A_H}} N_H^* \quad (\text{A.29})$$

The monoclonal equilibrium E_H^* with positive N_H , A_H and I_H , is then given as:

$$\left(\begin{array}{c} N_{H,E_H}^* \\ I_{H,E_H}^* \\ d_{I_H} \\ A_{H,E_H}^* \\ u_H \\ N_{L,E_H}^* \\ I_{L,E_H}^* \\ A_{L,E_H}^* \end{array} \right) = \left(\begin{array}{c} K - \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} \\ \frac{r_H - d_{A_H}}{d_{I_H}} A_{H,E_H}^* \\ \frac{u_H(r_H - d_{A_H})K - u_H}{d_{I_H}(r_H + d_{A_H})} \\ \frac{u_H}{r_H + d_{A_H}} N_{H,E_H}^* \\ \frac{u_H d_{I_H}}{(r_H + d_{A_H})K - b_H(r_H - d_{A_H})} \\ 0 \\ 0 \\ 0 \end{array} \right) \quad (\text{A.30})$$

Note that the number of empty niches in the steady state is simply given as $N_E^* = \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})}$. This equilibrium has three positive components if and only if $r_H > d_{A_H}$ and $\frac{K b_H}{d_{I_H}(r_H - d_{A_H})} > \frac{r_H + d_{A_H}}{r_H - d_{A_H}}$.

Analogously the monoclonal equilibrium of the second clone (denoted by subscript L) has three positive components if and only if $r_L > d_{A_L}$ and $\frac{K b_L}{d_{I_L}(r_L - d_{A_L})} > \frac{r_L + d_{A_L}}{r_L - d_{A_L}}$.

Coexistence equilibria

Imposing additional conditions on the parameters, both cell clones can coexist in an equilibrium state of the system. The equilibrium will be denoted as E_B^* (with the B denoting Both as both clones are present).

Observe that the relation between the steady states of A_H and I_H shown in equation (A.26) holds by the same argument as in the single clone case. As such we have the relations

$$I_H^* = \frac{r_H - d_{A_H}}{d_{I_H}} A_H^* \quad (\text{A.31})$$

$$I_L^* = \frac{r_L - d_{A_L}}{d_{I_L}} A_L^* \quad (\text{A.32})$$

Substitution of the first relation in $\dot{I}_H = 0$ yields $0 = -b_H(K - N_H^*)I_H^* + 2r_H \left(\frac{d_{I_H}}{r_H - d_{A_H}} \right) I_H^* - d_{I_H}I_H$, implying either $I_H^* = 0$ or

$$N_E^* = K - N_H^* - N_L^* = \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} \quad (\text{A.33})$$

Using the analogous equation for the second clone we find that either $I_L^* = 0$ or

$$N_E^* = K - N_H^* - N_L^* = \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})} \quad (\text{A.34})$$

In the case $\dot{I}_L^* = 0$ we also have $A_L^* = 0$ due to the relation in equation (A.32). This in turn implies that $N_L^* = 0$ since $\dot{A}_L^* = -u_L N_L$ when $I_L = 0$. In this case we either obtain the stem cell free steady state E_0^* or the monoclonal steady state E_L^* . The analogous argument applies to the case $I_H^* = 0$, where we obtain either the stem cell free steady state E_0^* or the monoclonal steady state E_H^* .

Equations (A.33) and (A.34) imply the following necessary condition for coexistence of two clones in an equilibrium state:

$$\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} = \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})} \quad (\text{A.35})$$

Notably this condition does not depend on the rates of detachment from the niche (u_H and u_L).

The condition (A.35) is sufficient for existence of a manifold of coexistence equilibria. Indeed, set

$$N_E^* = \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} = \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}$$

and consider the following manifold for $\alpha \in \mathbb{R}$

$$E_B^* = \begin{pmatrix} N_{H,E_B}^* \\ I_{H,E_B}^* \\ A_{H,E_B}^* \\ N_{L,E_B}^* \\ I_{L,E_B}^* \\ A_{L,E_B}^* \end{pmatrix} = \begin{pmatrix} \alpha \\ \frac{u_H r_H - d_{A_H}}{d_{I_H} r_H + d_{A_H}} \alpha \\ \frac{u_H}{r_H + d_{A_H}} \alpha \\ K - \alpha - N_E^* \\ \frac{u_L r_L - d_{A_L}}{d_{I_L} r_L + d_{A_L}} (K - \alpha - N_E^*) \\ \frac{u_H}{r_L + d_{A_L}} (K - \alpha - N_E^*) \end{pmatrix}. \quad (\text{A.36})$$

It then holds that

$$\frac{d}{dt} N_H = b_H \frac{d_{I_H} r_H + d_{A_H} r_H - d_{A_H} u_H}{b_H r_H - d_{A_H} r_H + d_{A_H} u_H} \alpha - u_H \alpha = 0$$

$$\frac{d}{dt} I_H = -b_H \frac{d_{I_H} r_H + d_{A_H} r_H - d_{A_H} u_H}{b_H r_H - d_{A_H} r_H + d_{A_H} u_H} \alpha + 2r_H \frac{u_H}{r_H + d_{A_H}} \alpha - d_{I_H} \frac{u_H r_H - d_{A_H} u_H}{d_{I_H} r_H + d_{A_H}} \alpha = 0$$

$$\frac{d}{dt} A_H = u_H \alpha - r_H \frac{u_H}{r_H + d_{A_H}} \alpha - d_{A_H} \frac{u_H}{r_H + d_{A_H}} \alpha = 0$$

$$\frac{d}{dt} N_L = b_L \frac{d_{I_L} r_L + d_{A_L} r_L - d_{A_L} u_L}{b_L r_L - d_{A_L} r_L + d_{A_L} u_L} \alpha - u_L \alpha = 0$$

$$\frac{d}{dt} I_L = -b_L \frac{d_{I_L} r_L + d_{A_L} r_L - d_{A_L} u_L}{b_L r_L - d_{A_L} r_L + d_{A_L} u_L} \alpha + 2r_L \frac{u_L r_L - d_{A_L} u_L}{r_L + d_{A_L}} \alpha - d_{I_L} \frac{u_L r_L - d_{A_L} u_L}{d_{I_L} r_L + d_{A_L}} \alpha = 0$$

$$\frac{d}{dt} A_L = u_L \alpha - r_L \frac{u_L}{r_L + d_{A_L}} \alpha - d_{A_L} \frac{u_L}{r_L + d_{A_L}} \alpha = 0$$

$$\text{Using } \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} = \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}, \text{ an analogous calculation reveals that } 0 = \frac{d}{dt} N_L = \frac{d}{dt} I_L = \frac{d}{dt} A_L.$$

With $\alpha \in (0, K - N_E^*)$ all components of E_B^* are positive if and only if $r_H > d_{A_H}$, $r_L > d_{A_L}$ and $N_E^* < K$.

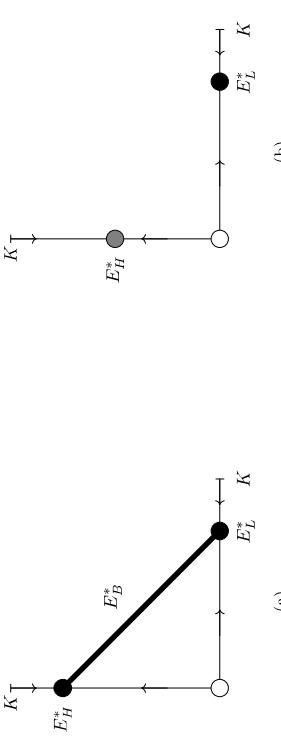


Figure A6: Phase plane illustrations of equilibria. The circles and the thick black line denote equilibria. The white circle corresponds to the stem cell free equilibrium E_0^* , E_H^* and E_L^* correspond to monoclonal equilibria. The thick line E_B^* corresponds to the manifold of equilibria where two clones coexist. The arrows depict system dynamics in the presence of a single clone. If both clones have equal fitness, the system will move asymptotically toward a point on the thick line, which depends on the initial conditions. In panel a, the two clones have equal fitness, whereas panel b depicts a scenario where the L -clone has a competitive advantage.

To see this, consider the linearization of model for one HSC clone around the homeostatic steady state, E_H^* :

$$A_1 = \begin{pmatrix} -b_H I_H^* - u_H & b_H N_{E,E_H}^* & 0 \\ b_H I_H^* & -b_H N_{E,E_H}^* - d_{I_H} & 2r_H \\ u_H & 0 & -r_H - d_{A_H} \end{pmatrix} \quad (\text{A.37})$$

The characteristic polynomial of A_1 is:

$$\begin{aligned} \chi_{A_1}(X) &= \det(A_1 - \mathbb{I}_{3 \times 3} X) \\ &= (-b_H I_H^* - u_H - X)(-b_H N_{E,E_H}^* - d_{I_H} - X) - b_H N_{E,E_H}^* [b_H I_H^* (-r_H - d_{A_H} - X) - 2r_H d_{A_H}] \end{aligned} \quad (\text{A.38}) \quad (\text{A.39})$$

which can be written on the form $\chi_{A_1}(X) = X^3 + a_2 X^2 + a_1 X + a_0$, with factors

$$\begin{aligned} a_2 &= b_H (I_H^* + N_{E,E_H}^*) + d_{A_H} + d_{I_H} + r_H + u_H \\ a_1 &= b_H (r_H + d_{A_H}) (I_H^* + N_{E,E_H}^*) + b_H d_{I_H} I_H^* + b_H u_H N_{E,E_H}^* + u_H (d_{A_H} + d_{I_H} + r_H d_{A_H}) \\ a_0 &= d_{I_H} (r_H + d_{A_H}) (b_H I_H^* + u_H) - b_H u_H (r_H - d_{A_H}) N_{E,E_H}^* = 3u_H d_{I_H} (r_H + d_{A_H}) \end{aligned} \quad (\text{A.40}) \quad (\text{A.41}) \quad (\text{A.42})$$

The Routh-Hurwitz Criterion ensures that the polynomial only has roots with negative real parts if

- (i) $a_2 > 0$ and
- (ii) $a_2 a_1 - a_0 > 0$ and
- (iii) $a_0 > 0$.

(i) and (iii) are fulfilled. (ii) is a consequence of the following calculation.

In absence of a second clone the homeostatic steady state is linearly stable whenever it exists i.e., if $r_H > d_{A_H}$,

$$\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} < K.$$

A.7 Stability analysis

A.7.1 Homeostatic state

In absence of a second clone the homeostatic steady state is linearly stable whenever it exists i.e., if $r_H > d_{A_H}$,

$$\begin{aligned}
& -a_0 + a_2 a_1 \\
& = -3u_H d_{I_H} (r_H + d_{A_H}) + \underbrace{\left[b_H (I_H^* + N_{E,E_H}^*) \right]}_{=:\xi_1 > 0} + \underbrace{\left(r_H + d_{A_H} + d_{I_H} + u_H \right)}_{=:\xi_2 > 0} \times \\
& \quad \left[\underbrace{\left(b_H (r_H + d_{A_H}) (I_H^* + N_{E,E_H}^*) + b_H d_{I_H} I_H^* + b_H u_H N_{E,E_H}^* \right)}_{=:\xi_3 > 0} + \underbrace{\left(u_H (d_{A_H} + d_{I_H} + r_H) + d_{I_H} (d_{A_H} + r_H) \right)}_{=:\xi_4 > 0} \right] \\
& = -3u_H d_{I_H} (r_H + d_{A_H}) + \xi_1 \xi_3 + \xi_2 \xi_4 + \left(r_H + d_{A_H} + d_{I_H} + u_H \right) \left(u_H (d_{A_H} + r_H) + u_H d_{I_H} + d_{I_H} (d_{A_H} + r_H) \right) \\
& \quad + d_{I_H}^2 (r_H + d_{A_H}) + u_H^2 (r_H + d_{A_H})^2 + u_H d_{I_H} (r_H + d_{A_H}) + d_{I_H} (r_H + d_{A_H})^2 + u_H d_{I_H} (r_H + d_{A_H}) + u_H d_{I_H}^2 \\
& \quad + d_{I_H}^2 (r_H + d_{A_H}) + u_H^2 (r_H + d_{A_H}) + u_H d_{I_H} (r_H + d_{A_H}) + u_H d_{I_H}^2 + d_{I_H}^2 (r_H + d_{A_H}) + u_H^2 (r_H + d_{A_H}) + u_H^2 d_{I_H}^2 \\
& > 0
\end{aligned}$$

Hence, all roots of the polynomial $\chi_{A_1}(X)$ have negative real parts. Thus, the eigenvalues of A_1 all have negative real part and the steady state E_H^* is locally stable.

A.7.2 Perturbations of the homeostatic state

In the following we investigate under which conditions the healthy equilibrium can be destabilized by adding a second stem cell clone. We assume that $r_H > D_{A_H}$, $r_L > d_{A_L}$, $\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} < K$, and $\frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})} < K$, which is necessary and sufficient for existence of the two monoclonal steady states E_H^* and E_L^* .

In the following we show that if we add a small number of malignant cells to the homeostatic state these cells will expand if $\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} > \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}$ and die out if $\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} < \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}$. For this purpose we consider the linearization of the model around the healthy equilibrium.

The linearization of the model around the homeostatic equilibrium E_H^* is given by

$$A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_4 \end{pmatrix},$$

where A_1 is the linearization of the model for one HSC clone around the homeostatic steady state, A_2 is a real 3×3 matrix and

$$A_4 = \begin{pmatrix} -u_L & b_L N_{E,E_H}^* & 0 \\ 0 & -b_L N_{E,E_H}^* - d_{A_L} & 2r_L \\ u_L & 0 & -r_L - d_{A_L} \end{pmatrix}. \quad (\text{A.43})$$

The characteristic polynomial of A is

$$\begin{aligned}
\chi_A(X) & = \det(A - \mathbb{I}_{6 \times 6} X) \\
& = \begin{pmatrix} A_1 - \mathbb{I}_{3 \times 3} X & A_2 \\ 0 & A_4 - \mathbb{I}_{3 \times 3} X \end{pmatrix} \\
& = \chi_{A_1}(X) \cdot \chi_{A_4}(X).
\end{aligned}
\quad (\text{A.44}) \quad (\text{A.45})$$

The latter holds since A is a block triangular matrix. The eigenvalues of A_1 correspond to the eigenvalues of the linearization of the model for one HSC clone around the homeostatic steady state. We know from the previous section that they are negative. We now have to investigate the sign of the real parts of the roots of $\chi_{A_4}(X)$. It holds

$$\begin{aligned}
& -a_0 + a_2 a_1 \\
& = -3u_H d_{I_H} (r_H + d_{A_H}) + \underbrace{\left(b_H (I_H^* + N_{E,E_H}^*) \right)}_{=:\xi_1 > 0} + \underbrace{\left(r_H + d_{A_H} + d_{I_H} + u_H \right)}_{=:\xi_2 > 0} \times \\
& \quad \left[\underbrace{\left(b_H (r_H + d_{A_H}) (I_H^* + N_{E,E_H}^*) + b_H d_{I_H} I_H^* + b_H u_H N_{E,E_H}^* \right)}_{=:\xi_3 > 0} + \underbrace{\left(u_H (d_{A_H} + d_{I_H} + r_H) + d_{I_H} (d_{A_H} + r_H) \right)}_{=:\xi_4 > 0} \right] \\
& = (X + u_L) (X + b_L N_{E,E_H}^* + d_{I_L}) (X + r_L + d_{A_L}) - 2r_L u_L b_L N_{E,E_H}^* \quad (\text{A.48})
\end{aligned}$$

We note that

$$\det(a_4) = \chi_{A_4}(0) = u_L (b_L N_{E,E_H}^* + d_{I_L}) (r_L + d_{A_L}) - 2r_L u_L b_L N_{E,E_H}^* \quad (\text{A.49})$$

$$= u_L (b_L N_{E,E_H}^* r_L + d_{I_L} r_L + b_L N_{E,E_H}^* d_{A_L} + d_{I_L} d_{A_L}) - 2r_L u_L b_L N_{E,E_H}^* \quad (\text{A.50})$$

$$= u_L (-b_L N_{E,E_H}^* r_L + d_{I_L} r_L + b_L N_{E,E_H}^* d_{A_L} + d_{I_L} d_{A_L}) \quad (\text{A.51})$$

$$= u_L (b_L N_{E,E_H}^* (d_{A_L} - r_L) + d_{I_L} (d_{A_L} + r_L)). \quad (\text{A.52})$$

Furthermore $\det(A_4) > 0 \Leftrightarrow N_{E,E_H}^* < \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}$. Together with the expression for N_{E,E_H}^* , we obtain

$$\det(A_4) > 0 \Leftrightarrow \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} < \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}. \quad (\text{A.53})$$

The characteristic polynomial expands to

$$\chi_{A_4}(X) = X^3 + (u_L + b_L N_{E,E_H}^* + d_{I_L} + r_L + d_{A_L}) X^2 \quad (\text{A.54})$$

$$+ [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + (r_L + d_{A_L}) (b_L N_{E,E_H}^* + d_{I_L})] X \quad (\text{A.55})$$

We now check the Routh-Hurwitz Criterion. We write $\chi_{A_4}(X) := X^3 + a_2 X^2 + a_1 X + a_0$. The polynomial has only roots with negative real parts if

- (i) $a_2 > 0$ and
- (ii) $a_2 a_1 - a_0 > 0$ and
- (iii) $a_0 > 0$.

(i) is satisfied, since it is a sum of positive terms. (ii) expands to

$$[u_L (b_L N_{E,E_H}^* + d_{I_L} + r_L + d_{A_L}) + (r_L + d_{A_L}) (b_L N_{E,E_H}^* + d_{I_L})] [u_L + b_L N_{E,E_H}^* + d_{I_L} + r_L + d_{A_L}] \quad (\text{A.56})$$

$$- u_L b_L N_{E,E_H}^* d_{A_L} + u_L r_L - u_L d_{I_L} (d_{A_L} + r_L) \quad (\text{A.57})$$

$$= [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + (r_L + d_{A_L}) (b_L N_{E,E_H}^* + d_{I_L})] \quad (\text{A.58})$$

$$+ u_L [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + (r_L + d_{A_L}) (b_L N_{E,E_H}^* + d_{I_L})] \quad (\text{A.59})$$

$$- u_L b_L N_{E,E_H}^* d_{A_L} + u_L r_L - u_L d_{I_L} (d_{A_L} + r_L) \quad (\text{A.60})$$

$$= [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + (r_L + d_{A_L}) (b_L N_{E,E_H}^* + d_{I_L})] [b_L N_{E,E_H}^* + d_{I_L} + r_L + d_{A_L}] \quad (\text{A.61})$$

$$+ u_L [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + u_L (r_L + d_{A_L}) b_L N_{E,E_H}^* + u_L (r_L + d_{A_L}) d_{I_L}] \quad (\text{A.62})$$

$$- u_L b_L N_{E,E_H}^* d_{A_L} + u_L r_L - u_L d_{I_L} (d_{A_L} + r_L) \quad (\text{A.63})$$

$$= [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + (r_L + d_{A_L}) (b_L N_{E,E_H}^* + d_{I_L})] [b_L N_{E,E_H}^* + d_{I_L} + r_L + d_{A_L}] \quad (\text{A.64})$$

$$+ u_L [u_L (b_L N_{E,E_H}^* d_{A_L} + d_{I_L} + r_L + d_{A_L}) + u_L r_L b_L N_{E,E_H}^* + u_L r_L] \quad (\text{A.65})$$

which is a sum of positive terms and hence positive. As shown above (iii) holds if and only if

$$\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} < \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}. \quad (\text{A.46})$$

Taken together this implies that the homeostatic state cannot be destabilized by a second clone if

$$\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} < \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}.$$

However in the opposite case $\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} > \frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}$ the homeostatic state is destabilized. In the first case, leukemic cells introduced to the system die out whereas in the second case they grow. This implies that the relation between $\frac{d_{I_L}(r_L + d_{A_L})}{b_L(r_L - d_{A_L})}$ and $\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})}$ decides whether an additional clone introduced in the system can persist or not.

These findings provide the theoretical foundation for the following notion of *fitness*. If we define the fitness of the HSC by $F_H := \frac{b_H(r_H - d_{A_H})}{d_{I_H}(r_H + d_{A_H})}$ and the fitness of LSC by $F_L := \frac{b_L(r_L - d_{A_L})}{d_{I_L}(r_L + d_{A_L})}$, then the clone with the larger fitness can destabilize the homeostatic state of the clone with lower fitness but not vice versa.

A.8 Parameter-dependence of steady-state values

We can use the partial derivatives of the steady state cell counts to find the parameter-dependence of the steady state values. This method leads to the results discussed in section 3.1 and in table 1 of the main text.

We omit all but one of these calculations. Apart from the following example, all the cases are straightforward.

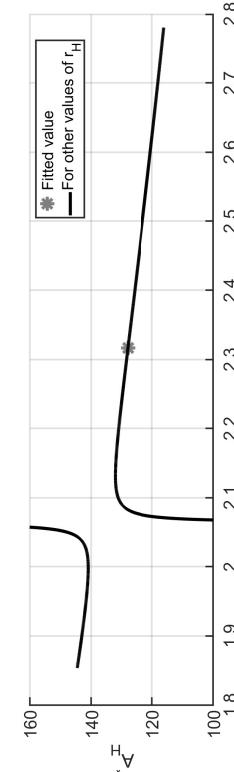
$$\begin{aligned} \frac{\partial A_H^*}{\partial r_H} &= \frac{u_H}{r_H + d_{A_H}} \left[\left(\frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})^2} - \frac{d_{I_H}}{b_H(r_H - d_{A_H})} \right) - \frac{1}{r_H + d_{A_H}} \left(K - \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} \right) \right] \quad (\text{A.66}) \\ &= \frac{u_H}{(r_H + d_{A_H})^2} \left[\left(\frac{d_{I_H}(r_H + d_{A_H})^2 - d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})^2} \right) - \left(K - \frac{d_{I_H}(r_H + d_{A_H})}{b_H(r_H - d_{A_H})} \right) \right] \quad (\text{A.67}) \\ &= \frac{u_H}{(r_H + d_{A_H})^2} \left[\frac{d_{I_H}(r_H + d_{A_H})^2}{b_H(r_H - d_{A_H})^2} - K \right] \quad (\text{A.68}) \end{aligned}$$

As such, $\frac{\partial A_H^*}{\partial r_H}$ is negative if and only if $\frac{d_{I_H}(r_H + d_{A_H})^2}{b_H(r_H - d_{A_H})^2} < K$ which is equivalent to $-\sqrt{\frac{b_H}{d_{I_H}}} K < \frac{(r_H + d_{A_H})}{(r_H - d_{A_H})} < \sqrt{\frac{b_H}{d_{I_H}}} K$. Existence of the steady state E_H^* requires that $r_H > d_{A_H}$ and $\frac{K b_H}{d_{I_H}} > \frac{r_H + d_{A_H}}{r_H - d_{A_H}}$ the latter implies $\frac{K b_H}{d_{I_H}} > 1$.

Therefore, we obtain two parameter regimes:

- $\frac{\partial A_H^*}{\partial r_H} < 0$ if $\frac{(r_H + d_{A_H})}{(r_H - d_{A_H})} < \sqrt{\frac{b_H}{d_{I_H}}} K$
- $\frac{\partial A_H^*}{\partial r_H} > 0$ if $\sqrt{\frac{b_H}{d_{I_H}}} K < \frac{(r_H + d_{A_H})}{(r_H - d_{A_H})} < \frac{b_H}{d_{I_H}} K$

Supplemental figure A7. Illustration of how A_H^* depends on the parameter r_H . Note the singularity at $r_H = d_{A_H}$.



Supplemental figure A7: Illustration of how A_H^* depends on the parameter r_H . Note the singularity at $r_H = d_{A_H}$.

A.9 Production of progenitors

The production of progenitors per unit of time is described in the model by

$$i_{D_H} = 2d_{A_H} A_H + d_{I_H} I_H \quad (\text{A.69})$$

In the steady state situation, where $\frac{dI_H}{dt} = \frac{dAH}{dt} = 0$, the production of progenitors should be in an equilibrium with the outflux of progenitors due to mature cell production. Hence determining how the steady state

production of progenitors is influenced by changes of models parameters gives insights about the effect of the respective parameter changes on steady state mature cell production. In the following we study how perturbations of the parameters impact the steady state production of progenitors. The steady state production of progenitors is $i_{D_H}^* = 2d_{A_H} A_H^* + d_{I_H} I_H^*$. Therefore, in supplementary section A.6 we showed that in general, $I_H = \frac{r_H - d_{A_H}}{r_H + d_{A_H}} A_H^*$ and $A_H^* = \frac{u_H - d_{A_H}}{r_H + d_{A_H}} N_H^*$. Therefore,

$$\begin{aligned} i_{D_H}^* &= \left(2d_{A_H} + d_{I_H} \frac{r_H - d_{A_H}}{d_{I_H}} \right) A_H^* \quad (\text{A.70}) \\ &= (r_H + d_{A_H}) A_H^* \quad (\text{A.71}) \\ &= \frac{r_H + d_{A_H}}{r_H + d_{A_H}} u_H N_H^* = u_H N_H^* \end{aligned}$$

Thus, in steady state, the production of progenitors scales with $u_H N_H^*$. Calculating the partial derivatives of $u_H N_H^*$ with respect to the parameters and assuming $N_H^* > 0$ reveals the dependence on parameter perturbations discussed in the main text: $\frac{\partial^* D_H}{\partial r_H}$ increases with K , r_H , r_H and b_H , but decreases when d_{A_H} or d_{I_H} increases. In other words, increasing differentiation of HSC decreases the production of progenitors and in turn mature blood cells.

A.10 Peripheral blood malignant cell burden can change faster or slower compared to the malignant cell burden in the stem cell niche

Simulations indicate that the malignant cell burden in the stem cell compartment and in the peripheral blood can differ significantly. Figure A8a provides an example for a scenario where the malignant cell burden in the stem cell niche increases earlier and is higher compared to the progenitor compartment, while figure A8b provides an example for a scenario where the malignant cell burden increases earlier and is higher in the progenitor compartment compared to the stem cell compartment.

A.11 Multiple stem cell doses increase chimerism in the unpreconditioned host but not in the preconditioned host

To reduce complications resulting from low circulating blood cell counts after bone marrow transplantation different approaches have been suggested. One approach involves splitting the transplant in multiple doses that are infused on subsequent days [4, 6]. In unpreconditioned mice the latter results in a higher chimerism [4]. This observation supports the concept that there exist empty niches under steady state conditions that are filled by subsequent transplantation of small cell doses [4]. Our calibrated model is able to reproduce this observation as discussed in supplementary section A.3.

It is a clinically relevant question, whether the engraftment of transplanted cells in humans can be improved if the transplant is split into multiple doses. This question not only arises in the context of bone marrow transplantation in the case of hematologic malignancy, it is also relevant when genetically engineered HSC are used to cure metabolic diseases [16, 2, 9]. Furthermore, a transplant splitting approach could be useful if it is combined with a reduced intensity conditioning (RIC) regime. This could allow to further reduce the complications of stem cell transplants in the elderly.

To obtain insights into the question how preconditioning affects the advantageous impact of splitting a transplant into multiple doses, we simulate splitting the transplant in multiple doses that are infused at homocytosis, one third of the niche-bound HSC were replaced by a leukemic clone with competitive advantage. This scenario corresponds to the setting of a newly diagnosed AML patient. Preconditioning was modeled as pre-transplant ablation of a prescribed fraction of healthy and leukemic cells. We then simulated transplantation of healthy HSC either as single dose or as multiple subsequent doses adding up to the same cell-count. To evaluate the outcome, we use the residual disease 100 days after transplantation.

The number of transplanted healthy HSC were chosen as $3.75 \cdot 10^4$ for the first dose and $2 \cdot 10^4$ for the subsequent doses to mirror the methodology of Landau et al. [14]. Note that this is two order of magnitude lower than the cell-counts described by Landau et al. [14], since we do not assume all transplanted cells be to HSC.

The results are shown in Figure A9. Our simulations indicate that the benefit of splitting a transplant into multiple doses is greater for the unpreconditioned host. In the unpreconditioned host a reduction in the residual disease at day 100 of up to 8% was achieved when the transplant was split in multiple smaller doses. However, in the preconditioned host a reduction in the residual disease was only 2%.

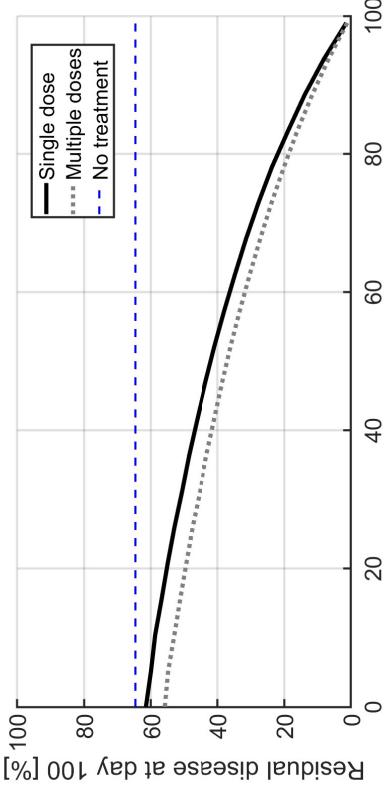


Figure A9: Residual disease after bone marrow transplantation with and without preconditioning. Before preconditioning e.g. 33% of the stem cell niche are occupied by a leukemic clone with a competitive advantage. Without treatment, the leukemic cell mass increases to approx. 65% within 100 days, shown as a dashed blue line. Before stem cells are transplanted a variable fraction of leukemic and healthy stem cells is ablated due to preconditioning. Preconditioning is assumed to be equally effective for healthy and malignant stem cells. To model the transplant either one dose of free healthy stem cells or multiple doses at subsequent days were added to the system. The total amount of transplanted cells is equal in both cases. The cell counts were based on the experimental work of Landau et al. [14] as described in the text. The effect of splitting the transplant in multiple doses decreases in case of preconditioning.

with preconditioning this effect becomes negligible. This finding is in line with results from a clinical trial [14].
The model suggests that only in scenarios without relevant preconditioning subsequent transplantation of multiple cell doses could be beneficial compared to a single dose transplant.

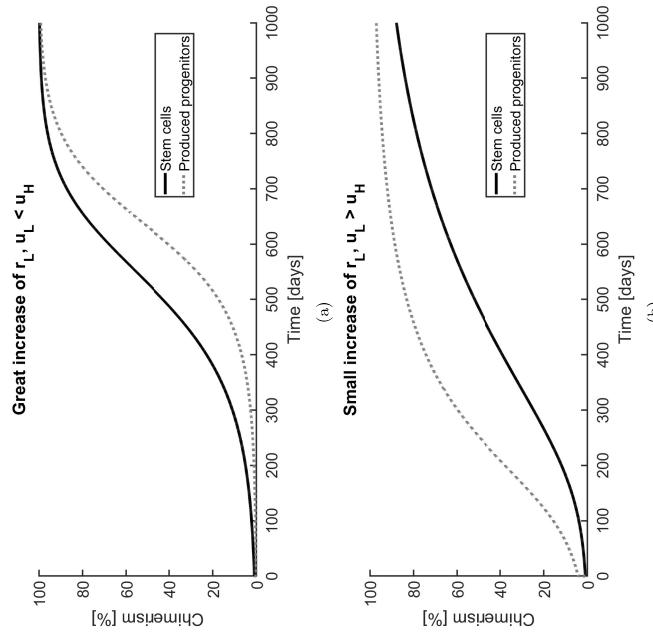


Figure A8: Time evolution of chimerism in the stem cell and in the progenitor compartment. In both panels the niche is invaded by a clone with a fitness advantage due to a higher proliferation rate. In panel b the advantage is smaller compared to panel a. In addition in the scenario depicted in panel a the invading clone has a decreased detachment rate compared to the wild-type clone, while in panel b the detachment rate of the invading clone is increased compared to the wild-type cells. Hence, although the stem cell chimerism develop similarly in both cases, the chimerism of the produced progenitors differs significantly.

A.12 Mobilizing agents could decrease the residual disease in the unpreconditioned host, but lead to an increase in the preconditioned host.

As shown in the previous section, our simulations indicate that the chimerism obtained after stem cell transplantation to an unpreconditioned host crucially depends on the amount of free niche spaces. This finding led us to question how the chimerism could be affected by stem cell mobilizing agents.

Mobilization agents force stem cells to detach from the niche. This could improve the efficiency of a bone marrow transplant following reduced or no preconditioning. One question arising in this context is whether no or reduced intensity conditioning (RIC) accompanied by mobilization agents and stem cell transplantation could be a treatment option for hematological cancers in the elderly not eligible for conventional transplantation regimens that require high dose chemotherapy. On the one hand, the use of mobilization agents could increase the number of donor cells homing to the niche and thus reduce the malignant cell burden. On the other hand detached stem cells are more prone to proliferate [21, 19, 20, 29]. For this reason stem cell mobilization could trigger the expansion of the malignant cell population. To better understand the relative contribution of these two processes to disease evolution we simulated a scenario where a malignant stem cell clone with a competitive advantage initially occupies one third of the bone marrow niche. First we consider a scenario without preconditioning. A transplant of healthy stem cells is introduced at day 0 and the detachment rate of healthy and malignant stem cells is increased at the day of transplantation due to mobilization agents. Later the detachment rate returns to its original value. The model was simulated using the parameters given in section A.3 of this supplement.

The simulations suggest that stem cell mobilization for short periods of time leads to an increased engraftment of healthy cells compared to the setting without mobilization. However, the number of malignant cells increases if the duration of mobilization treatment is longer than a critical time t_c . The stronger the mobilization effect, the earlier the malignant cell expansion dominates and the smaller the value of t_c . Figure A10 shows after which duration of mobilization treatment the malignant cell burden starts to increase and how this depends on the strength of the mobilization effect.

To consider the effect of mobilizing agents combined with preconditioning, we simulated the same scenario as above, with a malignant stem cell subpopulation occupying one third of the stem cell niche. On day 0, preconditioning was simulated by removing a certain percentage of all stem cells. We assume that preconditioning acts equally on healthy and malignant stem cells. After preconditioning, a transplant of free healthy cells was added, accompanied by mobilizing agents that were effective for 0.5 days.

In figure A11 we show the changes to the residual disease after 20 days in comparison to the same scenario without mobilization. The simulations imply that following efficient preconditioning (i.e. preconditioning that ablates large numbers of stem cells), increased mobilization leads to an increase in residual disease. However, if preconditioning is performed with reduced intensity, the increased mobilization can lead to a decrease in residual disease.

The effect of mobilizing agents may also be of relevance when genetically engineered cells have to be introduced to the stem cell niche of an unpreconditioned host, as in the case of gene therapy. Next we addressed the question whether the number of genetically engineered cells homing to the bone marrow could be increased by using mobilizing agents prior to transplantation. Fig. A12 shows that in case of no preconditioning the number of genetically engineered stem cells homing to the niche increases if mobilizing agents are administered before transplantation, however this effect becomes negligible in the context of preconditioning.

Taken together our simulations imply that mobilizing agents could improve the homing of transplanted cells to the niche only in hosts with reduced or no preconditioning.

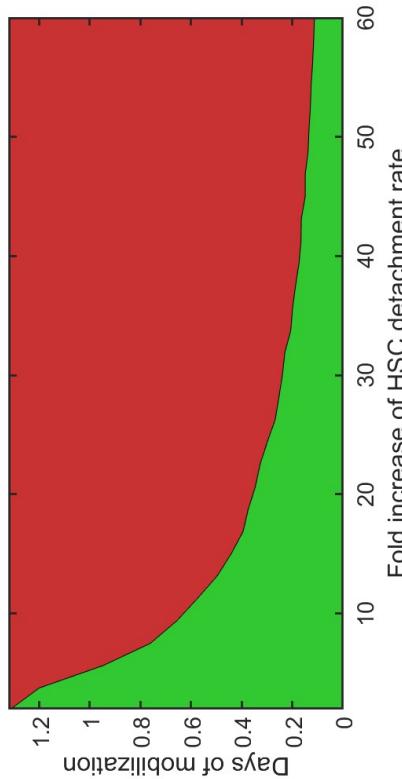


Figure A10: Impact of mobilizing agents on the non-preconditioned host. In an unpreconditioned host stem cell mobilizing agents can increase the number of graft-related cells homing to the niche. This allows for a reduction of the malignant cell burden in absence of preconditioning. However, if the effect of the mobilizing agents lasts longer than a critical time t_c , malignant cells detached from the niche expand and increase the disease burden. To assess this effect, we compare the following two hypothetical therapy approaches:
(i) bone marrow transplantation without preconditioning, (ii) bone marrow transplantation without preconditioning accompanied by mobilizing agents for at most 1.5 days. We compare the malignant cell burden after 100 days. As a beneficial impact of mobilization we define a reduction of disease burden after strategy (ii) compared to strategy (i). This is indicated by the green area. In the simulations, a leukemic clone initially occupies one third of the niche. The horizontal axis quantifies by how much the mobilizing agent increases the rate of stem cell detachment from the niche relative to wild type HSC, the vertical axis quantifies how long the mobilizing effect remains in the system.

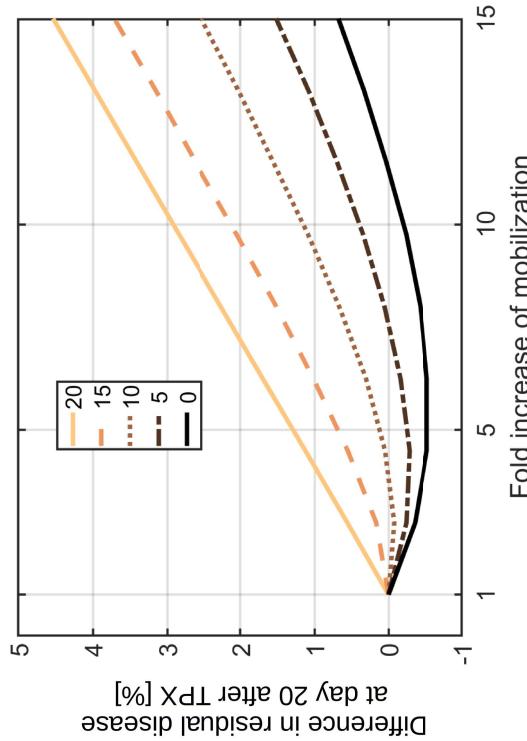


Figure A11: Impact of mobilizing agents in the preconditioned host. Before transplantation a variable amount of stem cells was ablated by preconditioning ("0" corresponds to the case without preconditioning, "20" corresponds to the case where preconditioning ablated 20% of healthy and malignant stem cells from the niche.) Post-transplantation stem cell detachment rates from the niche were increased by mobilizing agents for 0.5 days. The vertical axis shows the increase of malignant cells due to the mobilizing agent (residual disease after therapy consisting of preconditioning, bone marrow transplantation, plus mobilization minus residual disease after therapy consisting of preconditioning and bone marrow transplantation). Negative values indicate that the mobilizing agent has a beneficial effect. Mobilizing agents are not beneficial when preconditioning removes more than 10% of cells. When preconditioning removes more than 20%, the result is similar to the scenario where 20% of cells are removed.

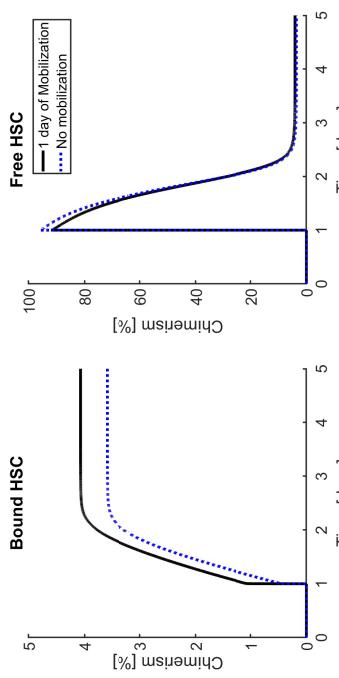


Figure A12: Impact of mobilizing agents on the homing of genetically engineered HSC in unpreconditioned hosts. In the non-preconditioned host the number of genetically engineered stem cells homing to the bone marrow can be increased by the use of mobilizing agents prior to transplantation. In the simulation depicted in black HSC mobilization has been increased by a factor of 2 during the day before transplantation. On the time of transplantation, the drug-induced mobilization was assumed to have stopped. For comparison, the blue dotted lines depicts a scenario with no increase in HSC mobilization.

References

- [1] Abkowitz, J. L., Catlin, S. N., McCallie, M. T. and Guttman, P. [2002], 'Evidence that the number of hematopoietic stem cells per animal is conserved in mammals', *Blood* **100**(7), 2665–2667.
[URL:](http://www.bloodjournal.org/cgi/doi/10.1182/blood-2002-03-0822) <http://www.bloodjournal.org/cgi/doi/10.1182/blood-2002-03-0822>
- [2] Ajati, A., Slavin, S., Aker, M., Ficara, F., Doola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabuchi, A., Carlucci, F., Marinello, E., Cattaneo, F., Vai, S., Sorrida, P., Miniero, R., Roncarolo, M. G. and Bordignon, C. [2002], 'Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning', *Science (New York, N.Y.)* **296**(5577), 2410–2413.
- [3] Ashcroft, P., Manz, M. G. and Bonhoeffer, S. [2017], 'Clonal dominance and transplantation dynamics in hematopoietic stem cell compartments', *PLOS Computational Biology* **13**(10), e1005803.
[URL:](https://doi.plos.org/10.1371/journal.pcbi.1005803) <https://doi.plos.org/10.1371/journal.pcbi.1005803>
- [4] Bhattacharya, D., Czechowicz, A., Ooi, A. L., Rossi, D. J., Bryder, D. and Weissman, I. L. [2009], 'Niche recycling through division-independent egress of hematopoietic stem cells', *The Journal of Experimental Medicine* **206**(12), 2837–2850.
[URL:](http://www.jem.org/lookup/doi/10.1084/jem.20090778) <http://www.jem.org/lookup/doi/10.1084/jem.20090778>
- [5] Boyd, A. L., Campbell, C. J. V., Hopkins, C. I., Fielding-Conyn, A., Russell, J., Ulemeck, J., Foley, R., Leber, B., Xenocostas, A., Collins, T. J. and Bhatia, M. [2014], 'Niche displacement of human leukemic stem cells uniquely allows their competitive replacement with healthy HSPCs', *The Journal of experimental medicine* **211**(10), 1925–1935.
- [6] Bucciarelli, F., Maurillo, L., Schanhuis, G., Del Principe, M., Di Veroli, A., Gurnari, C. and Venditti, A. [2019], 'The emerging role of measurable residual disease detection in AML in morphologic remission', *Seminars in Hematology* **56**(2), 125–130.
[URL:](https://linkinghub.eisnerweb.com/retrieve/pii/S003719631830101X) <https://linkinghub.eisnerweb.com/retrieve/pii/S003719631830101X>
- [7] Cashman, J., Clark-Lewis, I., Eaves, A. and Eaves, C. [2002], 'Stromal-derived factor 1 inhibits the cycling of very primitive human hematopoietic cells in vitro and in NOD/SCID mice', *Blood* **99**(3), 792–799.
[URL:](https://ashpublications.org/blood/article/99/3/792/53466/Stromal-derived-factor-1-inhibits-the-cycling-of) <https://ashpublications.org/blood/article/99/3/792/53466/Stromal-derived-factor-1-inhibits-the-cycling-of>
- [8] Catlin, S. N., Busque, L., Gale, R. E., Guttman, P. and Abkowitz, J. L. [2011], 'The replication rate of human hematopoietic stem cells in vivo', *Blood* **117**(17), 4460–4466.
[URL:](https://bloodjournal.hematologylibrary.org/content/117/17/4460.short) <https://bloodjournal.hematologylibrary.org/content/117/17/4460.short>
- [9] Fernández-García, M., Luisa Lamana, M., Hernando-Rodríguez, M., Sánchez-Domínguez, R., Bueren, J. and Yaícez, R. [2018], 'Improved Hematopoietic Gene Therapy in a Mouse Model of Fanconi Anemia Mediated by Mesenchymal Stromal Cells', *Human Gene Therapy* **29**(3), 327–336.
- [10] Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tominaga, H., Saito, N., Fukata, M., Miyamoto, T., Lyons, B., Oishima, K., Uchida, N., Taniuchi, S., Ohara, O., Akashi, K., Harada, M. and Shultz, L. D. [2007], 'Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region', *Nature Biotechnology* **25**(11), 1315.
- [11] Karpova, D. and Boing, H. [2015], 'Concise Review: CXCR4/CXCL12 Signaling in Immature Hematopoiesis—Lessons From Pharmacological and Genetic Models', *STEM CELLS* **33**(8), 2391–2399.
[URL:](http://doi.wiley.com/10.1002/stem.2054) [https://doi.wiley.com/10.1002/stem.2054](http://doi.wiley.com/10.1002/stem.2054)
- [12] Kumar, S. and Geiger, H. [2017], 'HSC Niche Biology and HSC Expansion Ex Vivo', *Trends in Molecular Medicine* **23**(9), 799–819.
- [13] Kunisaki, Y., Bruns, I., Schiermann, C., Ahmed, J., Pinho, S., Zhang, D., Mizoguchi, T., Wei, Q., Lucas, D., Ito, K., Mar, J. C., Bergman, A. and Frenette, P. S. [2013], 'Arteriolar niches maintain haematopoietic stem cell quiescence', *Nature* **502**(7473), 637–643.
[URL:](http://www.nature.com/articles/nature12612) [https://www.nature.com/articles/nature12612](http://www.nature.com/articles/nature12612)
- [14] Landau, H., Wood, K., Chung, D. J., Koehne, G., Lendvai, N., Hassoun, H., Lesokhin, A., Hoover, E., Zheng, J., Devlin, S. M. and Giralt, S. [2016], 'Fractionated stem cell infusions for patients with plasma cell myeloma undergoing autologous hematopoietic cell transplantation', *Leukemia & Lymphoma* **57**(8), 1781–1785.
[URL:](https://www.tandfonline.com/doi/full/10.3109/10428194.2015.1121256) <https://www.tandfonline.com/doi/full/10.3109/10428194.2015.1121256>
- [15] Mead, A. J. and Mullally, A. [2017], 'Myeloproliferative neoplasms stem cells', *Blood* **129**(12), 1607–1616.
[URL:](http://www.bloodjournal.org/lookup/doi/10.1182/blood-2016-10-696055) <http://www.bloodjournal.org/lookup/doi/10.1182/blood-2016-10-696055>
- [16] Morgan, R. A., Gray, D., Lomova, A. and Kohn, D. B. [2017], 'Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned', *Cell Stem Cell* **21**(5), 574–590.
[URL:](https://rupress.org/jem/article/205/4/777/47084/CXCR4-is-required-for-the-quiescence-of-the-primitive-hematopoietic-cells) <https://rupress.org/jem/article/205/4/777/47084/CXCR4-is-required-for-the-quiescence-of-the-primitive-hematopoietic-cells>
- [17] Nie, Y., Han, Y.-C. and Zou, Y.-R. [2008], 'CXCR4 is required for the quiescence of primitive hematopoietic cells', *Journal of Experimental Medicine* **205**(4), 777–783.
[URL:](https://aspublications.org/blood/article/106/4/1232/1239/Osteopontin-a-key-component-of-the-primitive-hematopoietic) <https://aspublications.org/blood/article/106/4/1232/1239/Osteopontin-a-key-component-of-the-primitive-hematopoietic>
- [18] Nilsson, S. K., Johnston, H. M., Whitty, G. A., Williams, B., Webb, R. J., Denhardt, D. T., Bertoncello, I., Bendall, L. J., Simmons, P. J. and Haylock, D. N. [2005], 'Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells', *Blood* **106**(4), 1232–1239.
[URL:](https://https://linkinghub.eisnerweb.com/retrieve/pii/S1079979615002377) <https://https://linkinghub.eisnerweb.com/retrieve/pii/S1079979615002377>
- [19] Schroeder, M. A. and DiPersio, J. F. [2012], 'Mobilization of hematopoietic stem and leukemia cells', *Journal of Leukocyte Biology* **91**(1), 47–57.
[URL:](https://doi.univly.com/10.1189/jlb.02100085) <https://doi.univly.com/10.1189/jlb.02100085>
- [20] Shen, Z.-H., Zeng, D.-F., Kong, P.-Y., Ma, Y.-Y. and Zhang, X. [2016], 'AMD3100 and G-CSF disrupt the cross-talk between leukemia cells and the endosteal niche and enhance their sensitivity to chemotherapeutic drugs in biomimetic polystyrene scaffolds', *Blood Cells, Molecules, and Diseases* **59**, 16–24.
[URL:](https://https://linkinghub.eisnerweb.com/retrieve/pii/S1079979616300237) <https://https://linkinghub.eisnerweb.com/retrieve/pii/S1079979616300237>
- [21] Shen, Z.-H., Zeng, D.-F., or AMD3100 in chemotherapy of haematological malignants?', *Medical Oncology* **32**(12), 262.
[URL:](https://link.springer.com/10.1007/s12032-015-0705-9) <https://link.springer.com/10.1007/s12032-015-0705-9>
- [22] Spivak, J. L. [2017], 'Myeloproliferative Neoplasms', *New England Journal of Medicine* **376**(22), 2168–2181.
[URL:](http://www.nejm.org/doi/10.1056/NEJMra1406186) <http://www.nejm.org/doi/10.1056/NEJMra1406186>
- [23] Stiehl, T., Ho, A. D. and Marciniak-Czochra, A. [2014], 'The impact of CD34+ cell dose on graftment after SCTs: Personalized estimates based on mathematical modeling', *Bone Marrow Transplantation* **49**(1), 30–37.
[URL:](https://linkinghub.eisnerweb.com/retrieve/pii/S10799796130005152) <https://linkinghub.eisnerweb.com/retrieve/pii/S10799796130005152>
- [24] Stiehl, T., Wang, W., Lutz, C. and Marciniak-Czochra, A. [2020], 'Mathematical modeling provides evidence for niche competition in human AML and serves as a tool to improve risk stratification', *Cancer Res.*
[URL:](https://linkinghub.eisnerweb.com/retrieve/pii/S0008547319352446) <https://linkinghub.eisnerweb.com/retrieve/pii/S0008547319352446>
- [25] Sugiyama, T., Kohara, H., Noda, M. and Nagasawa, T. [2006], 'Maintenance of the Hematopoietic Stem Cell Pool by CXCL12–CXCR4 Chemokine Signaling in Bone Marrow Stromal Cell Niches', *Immunity* **25**(6), 977–988.
[URL:](https://https://linkinghub.eisnerweb.com/retrieve/pii/S1079979613006051) <https://https://linkinghub.eisnerweb.com/retrieve/pii/S1079979613006051>
- [26] Vaidya, A. and Kale, V. [2015], 'Hematopoietic Stem Cells, Their Niche, and the Concept of Co-Culture Systems: A Critical Review', *Journal of stem cells* **10**(1), 13–31.
[URL:](https://linkinghub.eisnerweb.com/retrieve/pii/S1079979613000515) <https://linkinghub.eisnerweb.com/retrieve/pii/S1079979613000515>
- [27] Vandekerckhove, J. [2008], 'General simulated annealing algorithm'. MATLAB Central File Exchange, Last accessed: 2020-05-15.
[URL:](https://www.mathworks.com/matlabcentral/fileexchange/10548-general-simulated-annealing-algorithm) <https://www.mathworks.com/matlabcentral/fileexchange/10548-general-simulated-annealing-algorithm>
- [28] Wang, W., Stiehl, T., Raffel, S., Hoang, V. T., Hoffmann, I., Poisa-Boiro, L., Saeed, B. R., Blume, R., Manta, L., Eckstein, V., Buchbinder, T., Wucherer, P., Essers, M., Jauch, A., Trumpp, A., Marciniak-Czochra, A., Ho, A. D. and Lutz, C. [2017], 'Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia', *Haematologica* **102**(9), 1567–1577.
[URL:](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5584203/) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5584203/>
- [29] Welschinger, R., Liedtke, F., Basnet, J., Dela Pena, A., Juarez, J. G., Bradstock, K. F. and Bendall, L. J. [2013], 'Plerixafor (AMD3100) induces prolonged mobilization of acute lymphoblastic leukemia cells and increases the proportion of cycling cells in the blood in mice', *Experimental Hematology* **41**(3), 293–302.e1.
[URL:](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3617003/) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3617003/>

[30] Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.-G., Ross, J., Haug, J., Johnson, T., Feng, J., Q., Harris, S., Wiedemann, L. M., Mishina, Y. and Li, L. [2003], Identification of the haematopoietic stem cell niche and control of the niche size', *Nature* **425**(6960), 836–841.
URL: <http://www.nature.com/articles/nature02041>

[31] Zhang, Y. and Gao, Y. [2016], 'Novel chemical attempts at ex vivo hematopoietic stem cell expansion', *International Journal of Hematology* **103**(5), 519–529.

Mathematical modelling of the hematopoietic stem cell-niche system: Clonal dominance based on stem cell fitness.

Rasmus Kristoffer Pedersen^{1,3}, Morten Andersen¹, Thomas Stiehl², and Johnny T. Ottesen¹

¹Roskilde University, Denmark

²Heidelberg University, Germany

³Region Sjælland, Denmark

Abstract

Human blood cell production is maintained by hematopoietic stem cells (HSC) which give rise to all types of mature blood cells. Experimental observation of HSC in their physiologic bone-marrow microenvironment, the so-called stem cell niche, is challenging. Therefore, the details of HSC dynamics and the cellular interactions in the stem cell niche remain elusive. Mutations that give a competitive advantage to the mutated cells are the cause of clinical challenges when treating HSC-derived malignancies such as acute myeloid leukemia (AML) or the myeloproliferative neoplasms (MPNs). To investigate the significance of the interaction between the HSC and the stem cell niche in these malignancies, we propose and analyse a mechanism-based mathematical model of HSC dynamics within the bone-marrow microenvironment.

The model is based on the central hypothesis that HSC self-renewal depends on the niche. In the model, the interaction of HSC with specific niches located in the bone marrow are key to the indefinite renewal necessary for long-term maintenance of blood-production. We formulate a general model of n distinct clones that differ with respect to cell properties. We identify an attractive trapping region and compute and classify all steady states. A concept of HSC fitness naturally arises from the model analysis. HSC fitness is found to determine the asymptotic behaviour of the model, as the only steady state that is locally stable is related to the HSC-clone with the highest fitness.

Based on biological assumptions about HSC, we propose two reduced models of different complexity. A thorough mathematical analysis reveals that both reduced models have the same asymptotic behaviour as the full model.

We compare the simpler of the two models, a logistic equation of the disease burden, to clinical data of MPN-patients. The reduced model is found to agree well with data and suggests a simple interpretation and possible prediction of patient prognosis.

The proposed mathematical model and the reduced forms have the potential to provide insights into the regulation of HSC dynamics and blood cell formation, and ultimately for future advances in treatment of hematologic malignancies.

1 Introduction

Human blood consists of a vast amount of blood cells. As these mature blood cells (such as e.g. red blood cells, thrombocytes or leukocytes) die by programmed cell death or other mechanisms, a large number of new cells must be produced each day. The process of blood cell production, referred to as hematopoiesis, originates in the pool of hematopoietic stem cells (HSC). Through many steps of cell-division, self-renewal and differentiation, the diverse types of blood cells are produced, without exhaustion of the HSC population. While much research has been done in an effort to uncover the cell intrinsic regulatory mechanisms of HSC, open questions about their properties and functions still remain.

One particular area where an understanding of HSC-dynamics has a significant impact is the research of hematopoietic malignancies, such as acute myeloid leukemia (AML), chronic myeloid leukemia (CML) or the myeloproliferative neoplasms (MPNs). These diseases are believed to arise from malignant stem cells (Reya et al., 2001) that give rise to malignant cells which trigger disease-specific symptoms, such as e.g. an excess of thrombocytes. Understanding how malignant stem cells differ from wild-type HSC is crucial for efficient treatment (Dingli and Michor, 2006) and hence patient prognosis. HSC-specific niches have been in the focus of some research as they have been found to induce cellular quiescence, which can explain why some malignant stem cells respond poorly to treatment (Ishikawa et al., 2007).

Somatic mutations in the HSC genome can result in a subpopulation of HSC, referred to as a HSC clone, that differ from the normal HSC. As mutations accumulate through ageing, clonal hematopoiesis in which a substantial part of the production of blood cells is due to a single clone is expected to arise in as many as 10% of the population of people over 65 years of age. (Genovese et al., 2014). While clonal hematopoiesis by itself does not necessarily lead to symptoms of disease, it gives rise to the clinical concept of clonal hematopoiesis of indeterminate potential (CHIP) which can be a precursor or early stage of hematopoietic malignancy (Jaiswal and Ebert, 2019). Determining the significance of mutations and CHIP can be highly relevant in regards to the clinical question of understanding the risk of hematopoietic malignancies, particularly in an ageing population.

Despite years of comprehensive research on the behaviour and dynamics of HSC, both mathematically and biologically, a complete quantitative description still remains unclear. This leaves some open questions about HSC yet to be answered, such as:

- Which role do the HSC-specific niches play in maintaining HSC and blood cell counts?
- Which factors determine the outcome of competition between multiple distinct clones within the bone marrow microenvironment?
- How do mutational changes to HSC properties lead to the wide range of disease dynamics observed in patients with hematologic malignancies?
- Why are some clinical interventions highly successful for specific HSC-derived malignancies, yet less successful for other similar malignancies?
- Can insights about one malignancy guide the clinical management of other, potentially similar, malignancies?

In an effort to explore these questions, as well as many others, mathematical models have been developed throughout the literature. These include models of ordinary differential equations (Stiehl et al., 2015; Wang et al., 2017; Stiehl et al., 2020; Dingli and Michor, 2006; Andersen et al., 2017; Komarova and Wodarz, 2007; Ashcroft et al., 2017), delay differential equations (Colijn and Mackey, 2005; Park et al., 2019) as well as stochastic models (Roeder and Loeffler, 2002; Catlin et al., 2011). We refrain from giving a complete review of such work, but direct the reader to the work of Clapp and Levy (2015) or to a more general review on cancer modelling due to Altrock et al. (2015).

The mathematical models of hematologic malignancies and HSC behaviour found in the literature differ greatly in complexity, and hence the applicability of the models also varies. While complex models provide a detailed picture of a system, simpler models allow more directly for comparison with clinical data. Hence a dilemma arises where a model must have enough detail to be biologically accurate, but still simple enough to be useful in a clinical setting. Constructing complex models based on biological ideas and hypotheses allows for investigations about the nature of those ideas and hypotheses. By reducing and simplifying a complex model in mathematically or biologically reasonable ways, a simpler model arises. Such simpler models allow for comparison with data which would have been inconclusive when based on the complex model. In this way simplification helps to identify the data necessary to distinguish between different complex models. Hence, model reduction, simplification and subsequent comparison with experimental data both clarify the applicability of the model and pave the way for future experiments that can determine the validity of the hypotheses that the model was based on.

In this paper we present a model of HSC-behaviour, which captures the most significant parts of the biology, as well as the dynamical behaviour of blood cell production observed in both healthy and diseased individuals. The model is based on the dynamics of HSC within the bone-marrow microenvironment and builds on a central hypothesis about limited HSC renewal following detachment from HSC-specific niches. The processes impairing self-renewal when HSC are removed from their microenvironment are not well understood. Different from previous models of HSC dynamics (Stiehl et al., 2020; Ashcroft et al., 2017; Becker et al., 2019), we here propose that HSC can perform a limited number of divisions after detaching from the niche and then either differentiate or reattach to the niche. Reattachment to the niche is required to maintain stemness and to become reactivated at a later time.

A simplified version of the model has previously been presented by the authors (Pedersen et al., nd), and was used to investigate medical interventions such as HSC mobilization and transplantation. In particular, we found evidence that when suggesting novel treatment strategies of HSC-derived malignancies, careful attention must be paid to the treatment-induced changes in healthy and malignant stem cell dynamics to avoid potential harm.

Through mathematical analysis of the model, we determine the dynamics for non-negative initial conditions of a general model of n distinct HSC clones. By assuming that most HSC remain quiescent or inactive for extended periods of time, we propose a reduced model. The reduced form maintains the same general structure and dynamics as the full model, but allows for a more direct comparison with experimental data from literature.

Following still further simplifications, the simplest form of the model is compared to clinical data. In the most reduced form, the model is in agreement with data, suggesting that the proposed model is useful, albeit overparametrized in the full form with the current data available.

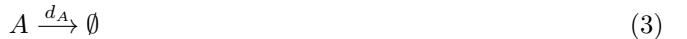
2 Model presentation

We propose a model of hematopoietic stem cells (HSC) and in particular their interaction with the bone marrow. Such a model has previously been developed by us (Pedersen et al., nd), and the present work is a generalization of that model. The biological foundation and assumptions are described here:

- HSC-specific niches are located in the bone marrow microenvironment. The specific cellular composition of the niches is not considered, and rather we consider an abstract notion of niches, assumed to be well-mixed throughout the bone marrow. It has been suggested that niche cells dynamically respond to the presence of HSC, and fluctuate in numbers (Becker et al., 2019). We hypothesize that the effect is minor and assume constant counts of niche-cells.
- HSC bind to the niches and remain quiescent for extended periods of time. The nature of how quiescence of HSC is maintained by the niches is a complex picture involving many different types of signalling (Nie et al., 2008; Nilsson et al., 2005), and is assumed to be necessary for maintaining stem cell function (Kumar and Geiger, 2017; Vaidya and Kale, 2015; Zhang and Gao, 2016). We denote niche-bound HSC as N , and denote the remaining free niches N_E .
- Niche-bound cells are assumed to detach from the niches with rate u , either spontaneously or through activation. The specifics of how this activation occurs are assumed not to impact on the behaviour of the resulting free HSC. We denote the activated free HSC as A . We assume that activated HSC can reattach to the empty niches N_E with rate b_A . These two processes can be written as:



- Upon detaching from the niche, the free HSC lose their quiescence and enter the cell-cycle (Kumar and Geiger, 2017). The detached cells are primed to undergo either self-renewal or differentiation, possibly due to signalling from the niche prior to release (Wilson and Trumpp, 2006). We assume that when HSC differentiate into more mature cells, they cannot reacquire stemness. We focus on the remaining stem cells, and differentiation is simply considered a loss of cells. If differentiation occurs with rate d_A this can be written as:



Note that this may include death of the HSC. However, under normal circumstances HSC death is assumed to occur for a negligible fraction of HSC.

If a free HSC self-renews, it divides into two HSC. Assuming the rate of division-related mutations is negligible, the two resulting cells are identical. We hypothesize that the resulting HSC are limited in their ability to further self-renew if they do not reattach to the niche. HSC following division that have to reattach to the niche to restore their full self-renewing potential are referred to as inhibited or exhausted HSC. For the the most general formulation, we consider one free HSC to self-renew with rate r and that the process results in 2γ inhibited HSC where $\gamma \geq 1$. In the supplementary material A we argue for the appropriateness of including the factor $\gamma \in \mathbb{R}$. The process can be written as:



where I denotes the inhibited HSC.

- The inhibited HSC cannot self-renew, but are still assumed to differentiate into more mature cells such as progenitors. This occurs with rate d_I , and as above we do not keep track of the resulting cells:



Inhibited HSC are hypothesized to attach to free niches, thereby gaining the ability to self-renew if they detach again. The attachment occurs with rate b_I and thus we write the process as:



The above processes are modelled as a system of differential equations, describing the rate of change of free niches, N_E , and the different states of HSC, N , A and I .

In the bone-marrow microenvironment, HSC with different cell-specific properties can co-exist for extended periods of time. We refer to HSC subpopulations with persistent and inheritable differences in cell properties as different clones. To define our model in the most general form, we model n different clones. The rates involved in the processes described above can all differ from one clone to the other, and hence we write each clone-specific rate with a subscript. Note that different clones interact only through the niches. No direct interaction between cells of different clones are considered in the model. A schematic representation of the model for $n = 2$ is shown in figure 1. The complete n -clone system of differential equations has $3n + 1$ dimensions, and is given by:

$$\dot{N}_E = \sum_{i=1}^n u_i N_i - N_E \sum_{i=1}^n (b_{I_i} I_i + b_{A_i} A_i) \quad (7a)$$

$$\dot{N}_j = b_{I_j} N_E I_j + b_{A_j} N_E A_j - u_j N_j \quad (7b)$$

$$\dot{I}_j = 2\gamma r_j A_j - b_{I_j} N_E I_j - d_{I_j} I_j \quad (7c)$$

$$\dot{A}_j = u_j N_j - b_{A_j} N_E A_j - r_j A_j - d_{A_j} A_j \quad (7d)$$

where j refers to the j^{th} clone and \cdot represents the derivative with respect to time t , $\frac{d}{dt}$. N_E denotes the abundance of empty niches, N_j are niches occupied by HSC of the j^{th} clone while A_j and I_j are free HSC of the j^{th} clone, in the active and inactive state, respectively. Based on their biological meaning, all parameters are non-negative: $u_j, b_{I_j}, b_{A_j}, r_j, d_{I_j}, d_{A_j} \geq 0$ for all j . A unique solution to system (7) exists and the solutions $N_E(t), N_j(t), I_j(t), A_j(t)$ are C^1 , since the right-hand sides of system (7) are continuous and fulfil a Lipschitz condition in the variables. As negative cell-counts are biologically meaningless, we define *feasibility* of solutions.

Definition 2.1: Feasibility

A set of solutions N_E, N_j, I_j, A_j is *feasible* if all variables N_E, N_j, I_j, A_j are non-negative for all j .

For non-negativity (and hence feasibility) of solutions, the time-derivatives for each variable must be non-negative when the given variable is zero. For all variables non-negative we find:

$$\dot{N}_{E|N_E=0} = \sum_{i=1}^n u_i N_i \geq 0 \quad (8a)$$

$$\dot{N}_{j|N_j=0} = N_E (b_{A_j} A_j + b_{I_j} I_j) \geq 0 \quad (8b)$$

$$\dot{I}_{j|I_j=0} = 2\gamma r_j A_j \geq 0 \quad (8c)$$

$$\dot{A}_{j|A_j=0} = u_j N_j \geq 0 \quad (8d)$$

for all j . Thus, due to the existence and uniqueness of solutions described above, the solutions of the system will remain non-negative for all $t > 0$, given non-negative initial conditions at $t = 0$.

From equations (7a) and (7b), $\dot{N}_E + \sum_{i=1}^n \dot{N}_i = 0$, and we define $K = N_E + \sum_{i=1}^n N_i$. As all variables are non-negative, $K \geq 0$, and in general we assume $K > 1$. Eliminating N_E , the system of equations reduces to a $3n$ dimensional system:

$$\dot{N}_j = b_{I_j} \left(K - \sum_{i=1}^n N_i \right) I_j + b_{A_j} \left(K - \sum_{i=1}^n N_i \right) A_j - u_j N_j \quad (9a)$$

$$\dot{I}_j = 2\gamma r_j A_j - b_{I_j} \left(K - \sum_{i=1}^n N_i \right) I_j - d_{I_j} I_j \quad (9b)$$

$$\dot{A}_j = u_j N_j - b_{A_j} \left(K - \sum_{i=1}^n N_i \right) A_j - r_j A_j - d_{A_j} A_j \quad (9c)$$

for all j . In addition, $N_E = K - \sum_{i=1}^n N_i$. Assuming feasible initial conditions, $N_E \geq 0$ restricts the range of $N_j(t)$ to $[0, K]$ for all j . In fact, $N_E \geq 0$ implies $\sum_{i=1}^n N_i(t) \leq K$. For feasible initial conditions, the range of $I_j(t)$ and $A_j(t)$ is $[0, \infty)$ for all j . This reformulation is an equivalent system, so results and conclusions based on one of the two formulations also holds for the other formulation. For this reason, parts of the analysis use the formulation in equations (7), while other parts make use of equations (9) to ease analysis.

Default parameters, determined in our previous work (Pedersen et al., nd), are given in table 1.

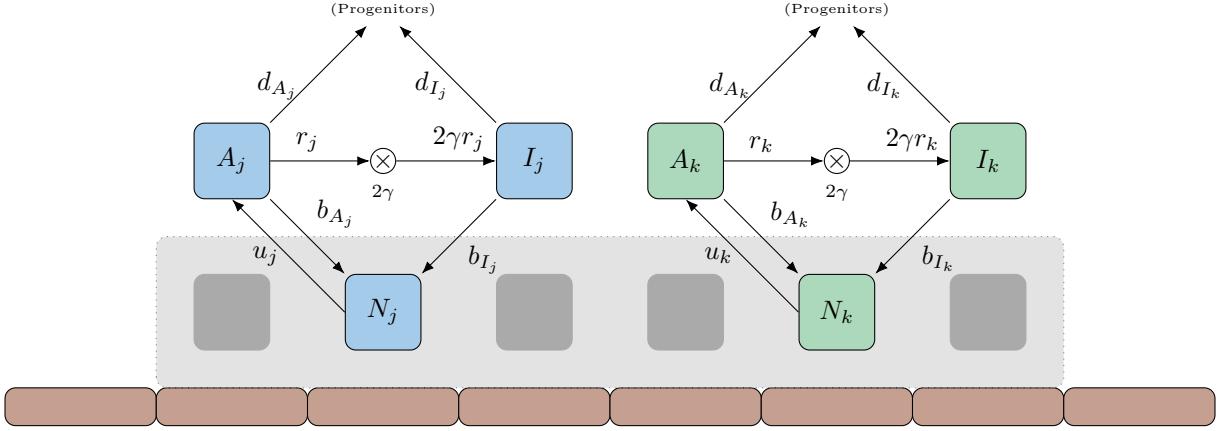


Figure 1: **Compartment diagram of the model.** The diagram depicts two distinct clones and the (indirect) competition through the niche. The two clones are denoted by subscript j and k respectively. Each box illustrates a variable while arrows illustrate the relationship between variables. The circle with the cross, \otimes , signifies a multiplication with 2γ . Note that I and A cannot change to N when there are no free niches, illustrated by the darkened boxes. The red rectangles at the bottom illustrate the bone-marrow cells that are part of the HSC niches.

K	15000 cells	u_j	0.04 day^{-1}
r_j	2.32 day^{-1}	b_{Ij}	0.96 day^{-1}
d_{A_j}	2.06 day^{-1}	d_{I_j}	3.77 day^{-1}

Table 1: **Default parameters for the proposed model.** Parameter-values were determined in our previous work (Pedersen et al., nd). Unless otherwise noted, parameters are equal for all clones j . While $b_{A_j} = 0$ was previously assumed, we here set $b_{A_j} = 0.01 \text{ day}^{-1}$.

3 Model analysis

3.1 Identifying an attractive trapping region

We showed above that the solutions of equations (9) remain non-negative for non-negative initial condition, and that $N_j(t)$ is restricted to the domain $[0, K]$ for all j . In this section we show that an attractive trapping region exists in which solutions with initial conditions inside the region stay within the region.

An upper limit for the sum of niche-bound cells is $\sum_{i=1}^n N_i = K$, and thus for any given clone $N_j = K - \sum_{k \neq j} N_k \geq 0$. Hence, N_j is bound from above, since due to equation (9a),

$$\dot{N}_j|_{\sum_{i=1}^n N_i = K} = -u_j N_j \leq 0 \quad (10)$$

Note that $\dot{N}_j = 0$ exactly when $N_j = 0$.

Extending this, note that:

$$\sum_{i=1}^n \dot{N}_i|_{\sum_{i=1}^n N_i = K} = - \sum_{i=1}^n u_i N_i < 0 \quad (11)$$

since not all $u_i N_i$ are zero when $\sum_{i=1}^n N_i = K$.

We define $\hat{A}_{j,\epsilon} = \frac{u_j K}{r_j + d_{A_j}} + \epsilon$ for any $\epsilon \geq 0$. Thus:

$$\dot{A}_j|_{A_j=\hat{A}_{j,\epsilon}} = -u_j(K - N_j) - (r_j + d_{A_j})\epsilon - b_{A_j}(K - \sum_{i=1}^n N_i) \left(\frac{u_j K}{r_j + d_{A_j}} + \epsilon \right) \quad (12)$$

$$\leq -(r_j + d_{A_j})\epsilon \leq 0 \quad (13)$$

since $\sum_{i=1}^n N_i \leq K$ and $N_j \leq K$. Hence $\dot{A}_j \leq 0$ for $A_j = \hat{A}_{j,0}$ and $\dot{A}_j < 0$ for $A_j \geq \hat{A}_{j,\epsilon}$ for any $\epsilon > 0$.

Consider an interval $\mathbb{A} = [\hat{A}_{j,\epsilon}, \hat{A}_{j,\epsilon} + \Delta]$ with $\Delta > 0$ on which $\dot{A}_j < 0$. Due to the extreme value theorem, there exists a finite μ such that $-\dot{A}_j \geq \mu, \forall A_j \in \mathbb{A}$, since \mathbb{A} is a compact subset. In particular

$$\inf_{\mathbb{A}} (-\dot{A}_j) = \min_{\mathbb{A}} (-\dot{A}_j) = \mu > 0 \quad (14)$$

and hence $\frac{1}{-\dot{A}_j} \leq \frac{1}{\mu}$ holds for all $A_j \in \mathbb{A}$.

To determine the time t_A it takes for the flow from $A_j(0) = \hat{A}_{j,\epsilon} + \Delta$ to reach $A_j(t_A) = \hat{A}_{j,\epsilon}$ we integrate over the interval:

$$t_A = \int_0^{t_A} dt = \int_{A_j(0)}^{A_j(t_A)} \frac{1}{\dot{A}_j} dA_j = \int_{\hat{A}_{j,\epsilon}}^{\hat{A}_{j,\epsilon} + \Delta} \frac{1}{-\dot{A}_j} dA_j \leq \frac{\Delta}{\mu} \quad (15)$$

and hence t_A is finite. Thus for any $\Delta > 0$, any point with $A_j(0) = \hat{A}_{j,\epsilon} + \Delta$ will move along a trajectory that enters the region $A_j \leq \hat{A}_{j,\epsilon}$ in finite time for any $\epsilon > 0$.

For I_j , we define $\hat{I}_{j,\epsilon} = \frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}} \epsilon + \epsilon$ for any $\epsilon \geq 0$. Thus, for both $A_j \leq \hat{A}_{j,\epsilon}$ and $I_j \geq \hat{I}_{j,\epsilon}$:

$$\dot{I}_j|_{I_j=\hat{I}_{j,TR,\epsilon}} \leq 2\gamma r_j \left(\frac{u_j K}{r_j + d_{A_j}} + \epsilon \right) - \left(b_{I_j} \left(K - \sum_{i=1}^n N_i \right) + d_{I_j} \right) \left(\frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}} \epsilon + \epsilon \right) \quad (16)$$

$$\leq -d_{I_j} \epsilon \leq 0 \quad (17)$$

since $\sum_{i=1}^n N_i \leq K$. Hence for $A_j \leq \hat{A}_{j,\epsilon}$ and $I_j \geq \hat{I}_{j,\epsilon}$, $\dot{I}_j \leq 0$, with $\dot{I}_j < 0$ for all $\epsilon > 0$.

Through an argument similar to the one above, it can be shown that trajectories with $I_j(0) > \hat{I}_{j,\epsilon}$ will follow a trajectory that enters the region $I_j \leq \hat{I}_{j,\epsilon}$ in finite time for any $\epsilon > 0$.

Theorem 3.1: Trapping region TR_ϵ

Defining $\hat{A}_{j,\epsilon} = \frac{u_j K}{r_j + d_{A_j}} + \epsilon$ and $\hat{I}_{j,\epsilon} = \frac{2\gamma r_j u_j K}{d_{I_j}(r_j + d_{A_j})} + \frac{2\gamma r_j}{d_{I_j}} \epsilon + \epsilon$ for any $\epsilon \geq 0$, a trapping region for equations (9) is given by:

$$TR_\epsilon = \left\{ \prod_{j=1}^n (N_j, A_j, I_j) \in \mathbb{R}^{3n} : N_j \geq 0, A_j \geq 0, I_j \geq 0, \sum_{j=1}^n N_j \leq K, A_j \leq \hat{A}_{j,\epsilon}, I_j \leq \hat{I}_{j,\epsilon}, j = 1, \dots, n \right\} \quad (18)$$

for any $\epsilon \geq 0$. For $\epsilon > 0$, TR_ϵ is an attracting trapping region.

3.2 Steady states

To understand the asymptotic behaviour of the trajectories of the model, we look for steady states in TR_0 . All points in TR_0 have feasible values of the variables, and hence existence of such steady states also implies feasibility of the steady state. We denote the steady state values with an asterisk (*).

For brevity, we use N_E^* as the empty niches in a given steady state, i.e. $N_E^* = K - \sum_{i=1}^n N_i^*$. From equations (9) it is seen trivially that:

$$A_j^* = \frac{u_j N_j^*}{r_j + d_{A_j} + b_{A_j} N_E^*} \quad (19)$$

and similarly:

$$I_j^* = \frac{2\gamma r_j u_j N_j^*}{(r_j + d_{A_j} + b_{A_j} N_E^*) (d_{I_j} + b_{I_j} N_E^*)} \quad (20)$$

Considering $\dot{N}_j = 0$ we find:

$$N_j^* = \frac{1}{u_j} N_E^* (b_{I_j} I_j^* + b_{A_j} A_j^*) \quad (21)$$

$N_j^* = 0$ for all j implies both $A_j^* = 0$ and $I_j^* = 0$, and hence a steady state in which a cell-type j has zero niche-bound cells but non-zero free cells cannot exist. However, equation (21) implies existence of a trivial steady state.

Theorem 3.2: Trivial steady state S_0^*

A trivial steady state, denoted S_0^* , always exists. In the trivial steady state $N_j^* = A_j^* = I_j^* = 0$, $\forall j \in \{1, \dots, n\}$.

The number of empty niches N_E^* in the trivial steady state is $N_E^* = K$ since $N_E = K - \sum_{i=1}^n N_i$. We proceed to investigate the case $N_j^* > 0$. From equations (19), (20) and (21) we obtain:

$$N_j^* = N_E^* \left(b_{I_j} \frac{2\gamma r_j}{(r_j + d_{A_j} + b_{A_j} N_E^*) (d_{I_j} + b_{I_j} N_E^*)} + b_{A_j} \frac{1}{r_j + d_{A_j} + b_{A_j} N_E^*} \right) N_j^* \quad (22)$$

which for $N_j^* > 0$ simplifies to

$$r_j + d_{A_j} + b_{A_j} N_E^* = b_{I_j} N_E^* \frac{2\gamma r_j}{d_{I_j} + b_{I_j} N_E^*} + b_{A_j} N_E^* \quad (23)$$

since we assume $N_E^* \geq 0$. Hence:

$$N_E^* = K - \sum_{i=1}^n N_i^* = \frac{d_{I_j} (r_j + d_{A_j})}{b_{I_j} ((2\gamma - 1)r_j - d_{A_j})}, \quad (24)$$

assuming $(2\gamma - 1)r_j \neq d_{A_j}$. Thus, requiring $N_j^* > 0$ implies a unique value of $N_E^* = K - \sum_{i=1}^n N_i^*$. For later purposes we define the *fitness* of the j^{th} clone:

Definition 3.1: Fitness

The fitness of the j^{th} clone is defined as:

$$F_j = \frac{b_{I_j} ((2\gamma - 1)r_j - d_{A_j})}{d_{I_j} (r_j + d_{A_j})}. \quad (25)$$

Hence, equation (24) can be written $K - \sum_{i=1}^n N_i^* = F_j^{-1}$. Since in general $0 \leq \sum_{i=1}^n N_i \leq K$, we find that $0 \leq F_j^{-1} \leq K$ must hold. Hence, for feasibility of a steady state in which $N_j^* > 0$ we have in particular $(2\gamma - 1)r_j > d_{A_j}$ and $F_j^{-1} < K$. In the case $F_j^{-1} = K$, we would have $\sum_{i=1}^n N_i = 0$ and hence $N_j^* > 0$ is violated and only S_0^* exists.

A_j^* and I_j^* are unique and given in terms of N_j^* due to equations (19) and (20). In addition, A_j^* and I_j^* are non-zero only if N_j^* is non-zero.

Lemma 3.1: Steady state values for free cells

For any non-trivial steady state, with $N_j^* > 0$ and $K - \sum_{i=1}^n N_i^* = F_j^{-1}$, A_j^* and I_j^* are unique and given by equation (19) and (20) respectively.

Lemma 3.1 allows us to restrict our focus to just N_j^* .

A steady state where only the j^{th} clone is non-zero exists, with $N_j^* = K - F_j^{-1}$.

Theorem 3.3: Single-clone steady state S_j^*

Given $(2\gamma - 1)r_j > d_{A_j}$ and $F_j^{-1} < K$, a single-clone steady state exists for which $N_j^* > 0$, $A_j^* > 0$, $I_j^* > 0$ and $N_E^* > 0$ while $\forall k \neq j : N_k^* = A_k^* = I_k^* = 0$. In particular, $N_j^* = K - F_j^{-1}$ and $N_E^* = F_j^{-1}$.

We denote the unique j^{th} single-clone steady state as S_j^* .

Equation (24) holds for any clone j . This allows for existence of steady states where the multiple clones co-exist, if F_j is equal for the clones.

As a particular example, we consider a scenario with two clones, 1 and 2, with equal fitness F (i.e. $F = F_1 = F_2$). The clones have equal fitness only when

$$\frac{b_{I_1} ((2\gamma - 1)r_1 - d_{A_1})}{d_{I_1} (r_1 + d_{A_1})} = \frac{b_{I_2} ((2\gamma - 1)r_2 - d_{A_2})}{d_{I_2} (r_2 + d_{A_2})} \quad (26)$$

Note that the conditions $(2\gamma - 1)r_1 > d_{A_1}$ and $(2\gamma - 1)r_2 > d_{A_2}$ as well as $F < K$ must hold for existence of a non-trivial steady state. In a co-existence steady state, the empty niches are uniquely determined as $N_E^* = F^{-1}$. In addition, it must hold that $F^{-1} + N_1^* + N_2^* = K$ which constrains the possible values of N_1 and N_2 (and in turn A_1^* , I_1^* , A_2^* and I_2^* , due to lemma 3.1) In particular, for $N_1^* = \beta$, $N_2^* = K - F^{-1} - \beta$. Since F^{-1} is uniquely determined, this is a line parametrized by β , connecting the single-clone steady states S_1^* (when $\beta = K - F^{-1}$) and S_2^* (when $\beta = 0$).

Similar relations may be found for 3 distinct clones (Resulting in a triangle of co-existence steady states) as well as higher numbers of clones.

Theorem 3.4: Co-existence steady states

Co-existence steady states in which multiple clones assume positive concentrations may exist.

A necessary and sufficient condition for co-existence is that F_j is equal for all co-existing clones.

The number of empty niches, N_E^* , in the co-existence steady states is given uniquely by $N_E^* = F_j^{-1}$.

Additionally, the bound cells must fulfil the condition $\sum_{i \in \mathcal{C}} N_i = K - F_j^{-1}$ where \mathcal{C} is the set of all co-existing clones.

This implies that co-existence steady states exists on a simplex where the dimension of the simplex is the number of co-existing clones, $\dim(\mathcal{C})$.

When there is no ambiguity, we denote the simplex of steady states S_C^* .

We note that multiple co-existence steady states can exist. As an example, consider a scenario with four clones. In such a scenario, we could have $F_1 = F_2 \neq F_3 = F_4$, which implies two separate lines of co-existence, one connecting S_1^* and S_2^* and another connecting S_3^* and S_4^* . Similarly, many different combinations of co-existence steady states can exist when two or more clones have the same fitness. However, such situations of equal fitness are unlikely to persist in a noisy biological setting. Because of this, we do not discuss the vast multitude of possible combinations of co-existence, and instead restrict our focus to situations without multiple simplexes of co-existence steady states. In some part we do however briefly comment on the case with co-existence of just two distinct clones, as it is exemplary for situations with multiple cases of equal fitness.

We sum up the feasible steady states. When the fitness of all clones are different, there are $n + 1$ steady states: A trivial steady state, S_0^* , (theorem 3.2) and n single-clone steady states, S_j^* (theorem 3.3). In case of k clones with equal fitness, a k -dimensional simplex of steady states, as described in theorem 3.4.

As a particular example, the two-dimensional case features three steady states and, if $F_1 = F_2$, a line of steady states for which $N_1^* + N_2^* = K - F_1^{-1} = K - F_2^{-1}$.

Omitting A_1^* , I_1^* , A_2^* and I_2^* due to lemma 3.1, but including N_E^* for clarity, we order a subset of variables as $(N_E^*, N_1^*, N_2^*)^T$, and the steady states are written as:

$$S_0^* = \begin{pmatrix} K \\ 0 \\ 0 \end{pmatrix}, \quad S_1^* = \begin{pmatrix} F_1^{-1} \\ K - F_1^{-1} \\ 0 \end{pmatrix}, \quad S_2^* = \begin{pmatrix} F_2^{-1} \\ 0 \\ K - F_2^{-1} \end{pmatrix} \quad (27)$$

When $F_1 = F_2$, a line of steady states also exists:

$$S_C^* = \begin{pmatrix} F_1^{-1} \\ \beta \\ K - F_1^{-1} - \beta \end{pmatrix} \quad (28)$$

where β is a number between 0 and $K - F_1^{-1}$. We refer to this line as the co-existence steady states for cell-types 1 and 2.

In figure 2 all possible combinations of steady states for one, two and three clones are illustrated. Since the unique values of A_j^* and I_j^* follow directly from N_j^* , the figures depict only the values of N_j^* .

As we describe below, the stability of steady states can be determined. In all panels of figure 2 the white circle signifies an unstable steady state, black circles signify steady states with no positive eigenvalues while the grey circle signifies a steady state which is stable only on the single-cell subspace.

3.3 Local stability of steady states

In this section, we determine the local stability of the steady states discussed above.

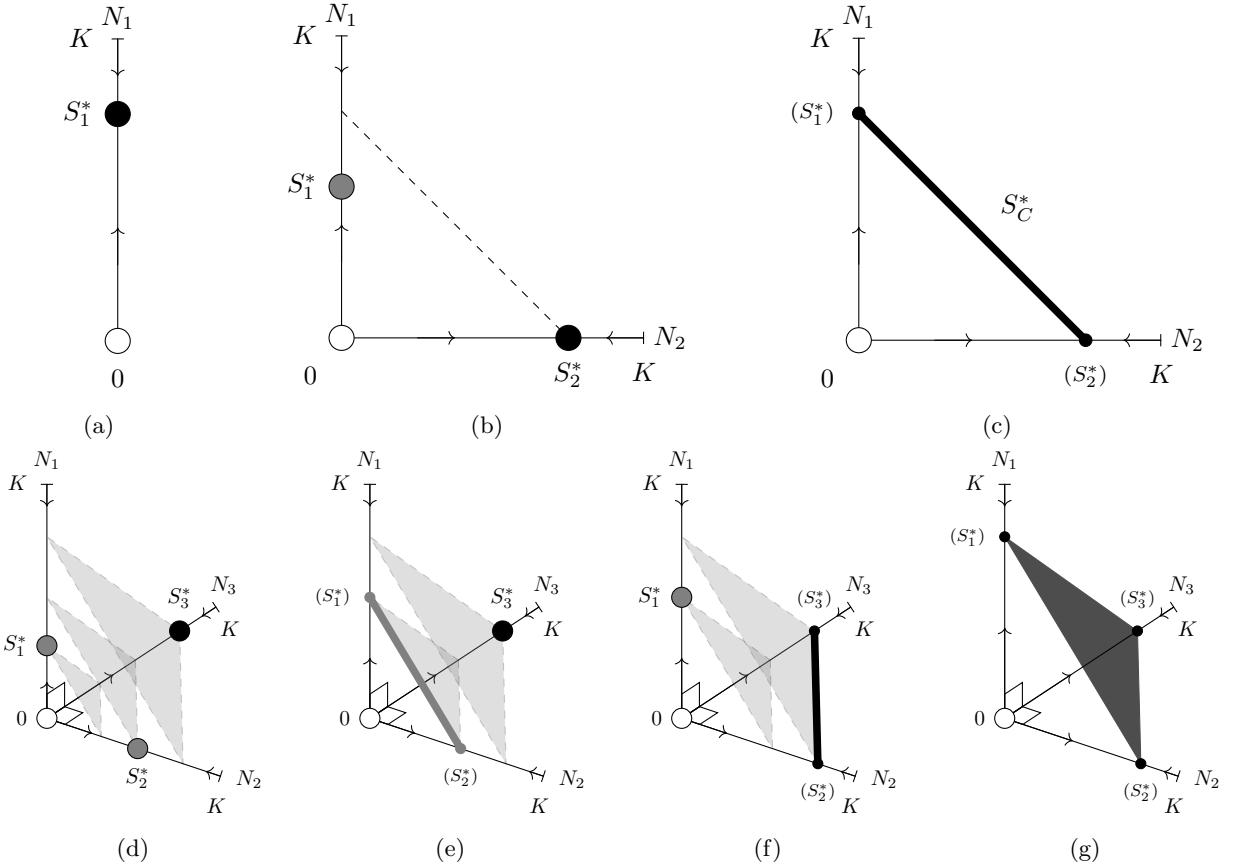


Figure 2: Illustrations of all possible combinations of steady states for one, two and three feasible clones. For clones that permit feasible single-clone steady states, all possible combinations are shown here. Hence this figure is a complete visual description of one-, two-, and three-clone scenarios up to permutations. As A_j^* and I_j^* follow directly from N_j^* , the figures depict the N_j -axes. In all figures, the white circle depicts the trivial steady state S_0^* in which all cell-counts are zero. Grey circles are steady states that are not locally stable, but are attracting within a single-clone subspace. Black circles and thick black lines depict locally stable steady states. Panel (a) shows a single-clone scenario. When $(2\gamma - 1)r_1 > d_{A_1}$, the S_1^* steady state is feasible and stable. Panel (b) depicts a scenario with two distinct clones. In particular, $F_1 < F_2$, which implies that the S_2^* steady state is stable and both the trivial and the S_1^* steady state is unstable. If no cells from clone 2 are present, the S_1^* steady state is attracting. The dashed line is only for illustrative purposes. In panel (c), a two-clone scenario is also shown, but here $F_1 = F_2$, resulting in a line of steady states connecting S_1^* and S_2^* . Numerical investigations show that this line is attracting, however, as shown in supplementary material B, the steady states have a zero eigenvalue and hence the steady states are non-hyperbolic. Panels (d) through (g) depict the full range of possible scenarios with three clones. In (d) all three clones have different fitness. In (e) $F_1 = F_2 < F_3$, while panel (f) has $F_1 < F_2 = F_3$. Finally, panel (g) depicts the scenario where $F_1 = F_2 = F_3$. The greyed out triangles in panel (d-f) are guides for the eye, while the black triangle in panel (f) is a triangle of co-existence steady states S_C^* .

Using the formulation of equations (9), we order the $3n$ variables as $(N_1, I_1, A_1, \dots, N_n, I_n, A_n)$. This allows us to write the Jacobian as a block-matrix of the form:

$$Jac = \begin{pmatrix} D_1 & G_1 & G_1 & \dots & G_1 \\ G_2 & D_2 & G_2 & \dots & G_2 \\ G_3 & G_3 & D_3 & \dots & G_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_n & G_n & G_n & \dots & D_n \end{pmatrix} \quad (29)$$

where

$$D_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j - u_j & b_{I_j} (K - \sum_{i=1}^n N_i) & b_{A_j} (K - \sum_{i=1}^n N_i) \\ b_{I_j} I_j & -b_{I_j} (K - \sum_{i=1}^n N_i) - d_{I_j} & 2\gamma r_j \\ u_j + b_{A_j} A_j & 0 & -b_{A_j} (K - \sum_{i=1}^n N_i) - r_j - d_{A_j} \end{pmatrix} \quad (30)$$

and

$$G_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j & 0 & 0 \\ b_{I_j} I_j & 0 & 0 \\ b_{A_j} A_j & 0 & 0 \end{pmatrix} \quad (31)$$

To determine the stability of the steady states described above, we investigate the eigenvalues of the Jacobian evaluated at the steady states. We note that the matrix D_j is the Jacobian for the subspace considering only the j 'th clone, i.e. the subspace with all other N_k , I_k and A_k equal to zero, for $k \neq j$.

At the trivial steady state, S_0^* , the matrices G_j are zero, and hence the eigenvalues of the full Jacobian is given by the eigenvalues of the D_j matrices, since the Jacobian is a diagonal block-matrix:

$$Jac|_{S_0^*} = \begin{pmatrix} D_1|_{S_0^*} & 0 & 0 & \dots & 0 \\ 0 & D_2|_{S_0^*} & 0 & \dots & 0 \\ 0 & 0 & D_3|_{S_0^*} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & D_n|_{S_0^*} \end{pmatrix} \quad (32)$$

where $|_{S_0^*}$ denotes evaluation in S_0^* .

In the k 'th single-clone steady state, S_k^* , any G_j where $j \neq k$ is zero. Thus the eigenvalues of the Jacobian are again given by the eigenvalues of the D_j matrices. For illustration, the Jacobian in the single-clone steady state with clone 1, $Jac|_{S_1^*}$, is given as:

$$Jac|_{S_1^*} = \begin{pmatrix} D_1|_{S_1^*} & G_1|_{S_1^*} & G_1|_{S_1^*} & \dots & G_1|_{S_1^*} \\ 0 & D_2|_{S_1^*} & 0 & \dots & 0 \\ 0 & 0 & D_3|_{S_1^*} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & D_n|_{S_1^*} \end{pmatrix} \quad (33)$$

Lemma 3.2: Stability of S_0^* and S_j^* is determined by D_j

In the trivial steady state S_0^* and in the single-clone steady states S_j^* , the eigenvalues of the Jacobian of the system of equations (9) evaluated in the given steady state values, are equal to the eigenvalues of the submatrices D_j , defined in equation (30).

Hence, the stability of the steady states can be determined by considering the submatrices D_j .

Before considering the co-existence steady states, S_C^* , we first discuss the sign of the eigenvalues of D_j .

In appendix B, we go through the process of checking the Routh-Hurwitz stability criterion for D_j . The results are summarized in the lemma below:

Lemma 3.3: Routh-Hurwitz stability criterion for D_j

Assume that S_j^* is feasible and hence $(2\gamma - 1)r_j > d_{A_j}$ as well as $K - F_j^{-1} > 0$ holds.

The matrix D_j of equation (30) evaluated in the trivial steady state, S_0^* , i.e. $D_j|_{S_0^*}$, has at least one eigenvalue with positive real part.

Evaluated in S_j^* , the j^{th} single-clone steady state, all eigenvalues of the matrix $D_j|_{S_j^*}$ have negative real part.

The eigenvalues of $D_j|_{S_k^*}$ where $k \neq j$ (i.e. the matrix with parameters of the j^{th} clone, evaluated in the single-clone steady state of the k^{th} clone) depend on the fitnesses of both the j^{th} clone and the k^{th} clone. In particular, if $F_k > F_j$, all eigenvalues have negative real part, and when $F_k < F_j$ at least one eigenvalue has positive real part.

Hence, for scenarios where no clones have equal fitness, we have determined the signs of the real parts of the eigenvalues of D_j when evaluated in S_0^* and S_j^* . This in turn determines the local stability of S_0^* and S_j^* .

Theorem 3.5: Stability of S_0^* and S_j^* for all j

If the system has a feasible single-clone steady state S_j^* , the trivial steady state S_0^* is locally unstable. Otherwise, it is locally stable.

Assuming no clones have equal fitness, a particular single-clone steady state S_k^* is stable if and only if $F_k > F_j$ holds for all $j \neq k$. If it does not hold, the steady state S_k^* is unstable.

In the special case where only one clone is considered, S_1^* is stable.

Thus, when no fitnesses are equal, the system has exactly one stable steady state and n unstable steady states. In particular, the stable steady state is the single-clone steady state of the clone with the highest fitness. Our use of *fitness* for F_j is thus justified.

Stability of the co-existence steady state

As discussed above, a simplex of solutions exists when multiple cell-types have equal fitness. As such situations are unlikely to arise in a noisy biological setting, we do not discuss them in great detail. In supplementary material B we show that in a scenario with two clones with equal fitness, the line of co-existence steady states has an eigenvalue of zero with a corresponding eigenvector along the line. Numerical investigations have shown that there is a zero eigenvalue along the simplex in general, see supplementary material B. When multiple clones have the same fitness, the Jacobian evaluated in the simplex of steady states has a zero eigenvalue. Furthermore, the numerical investigations also show that the eigenvalues depend on the fitness in the expected way: If the clones with equal fitness have a higher fitness than all other clones, the remaining eigenvalues are negative, and hence solutions of the system will approach the simplex of steady states. In the scenario where another clone has a higher fitness, the single-clone steady state of that type is stable, as expected. These relations are summarized in table 2.

In conclusion, the considerations about an equal fitness scenario shown in table 2 together with theorem 3.5 provides a complete description of hyperbolic steady states of the system and the criteria for non-hyperbolic steady states to arise.

	$F_1 > F_2 > F_3 > \dots$	$F_1 = F_2 > F_3 > \dots$	$F_1 = F_2 = F_3 > \dots$	$F_1 > F_2 = F_3 > \dots$
S_1^*	Stable	Zero eigenvalue	Zero eigenvalue	Stable
S_2^*	Unstable	Zero eigenvalue	Zero eigenvalue	Zero eigenvalue
S_3^*	Unstable	Unstable	Zero eigenvalue	Zero eigenvalue
\vdots	\vdots	\vdots	\vdots	\vdots
S_0^*	Unstable	Unstable	Unstable	Unstable
Co-existence?	No	Line (S_1^* and S_2^*)	Triangle (S_1^* , S_2^* and S_3^*)	Line (S_2^* and S_3^*)

Table 2: **Overview of the stability of steady states for examples of combinations of fitness.** All steady states are assumed to be feasible. The row labeled co-existence describes the given simplex of co-existence steady states and the points it connects. All steady states located on the co-existence simplex lead to the existence of a zero eigenvalue of the related Jacobian.

4 Model reduction

We present a model reduction. Scaling of variables together with two assumptions reduce the $3n$ dimensional system of equations (9) to an n dimensional system. The reduced system maintains most of the features and dynamics of the full system. While our considerations above yielded a complete description of the steady states of the model, a reduced model offers greater versatility as the number of parameters is reduced.

We make two general assumptions based on biological considerations:

Assumption 1 Only few niches remain empty at any given time, as found in the work of Bhattacharya et al. (2009). Hence N_E should be low in numbers compared to $N_j + I_j + A_j$ in steady state.

Assumption 2 Under most circumstances, the majority of HSC are niche-bound. Hence, in steady state, N_j are numerous compared to A_j and I_j .

Assumption 2 is in agreement with HSC mobilization studies (Yang et al., 2009) and the low division frequencies of HSC (Lee-Six et al., 2018).

Before proposing a reduced model, we first introduce a simplification by setting $b_{A_j} = 0$. Biologically, b_{A_j} is the re-attachment rate of HSC that have not yet divided after detaching from the niche. The simplification is justified, since such cells that immediately rebind can be considered included in the count of bound cells, N_j . Note that b_{A_j} does not contribute to either the steady state values N_j^* and N_E^* , or to the fitness, equation (25). The local stability of the steady states described above has been determined to be unchanged when $b_{A_j} = 0$. In addition, preliminary numerical investigations showed that the parameter should be small for the model to agree with data from literature. The model presented separately (Pedersen et al., nd) is this special case, with the additional assumption of $\gamma = 1$.

We introduce a scaling of the variables. Denoting unscaled variables with a superscript o and denoting the new scaled variables with the original notation without superscript, we define $N_j^o = \tilde{N}_j N_j$, $I_j^o = \tilde{I}_j I_j$ and $A_j^o = \tilde{A}_j A_j$. Our particular choices of scalings are $\tilde{N}_j = K$, $\tilde{I}_j = \frac{u_j}{b_{I_j}}$ and $\tilde{A}_j = \frac{u_j}{r_j} K$. Defining the smallest u_j as U , i.e. the u_j such that $u_j \leq u_k$ holds for all k , we introduce a scaled time-variable: $\tau = Ut$. The scaled version of equations (9) with $b_{A_j} = 0$ is thus:

$$\frac{dN_j}{d\tau} = \frac{u_j}{U} \left(\left(1 - \sum_{i=1}^n N_i \right) I_j - N_j \right) \quad (34a)$$

$$U \frac{dI_j}{d\tau} = 2\gamma b_{I_j} K A_j - b_{I_j} K \left(1 - \sum_{i=1}^n N_i \right) I_j - d_{I_j} I_j \quad (34b)$$

$$U \frac{dA_j}{d\tau} = r_j N_j - (r_j + d_{A_j}) A_j \quad (34c)$$

where N_j , I_j and A_j now denote the scaled variables. The scaling of N_j also scales the domain of functions that are solutions to \dot{N}_j , and hence $N_j(t) \in [0, 1]$ while still $A_j(t), I_j(t) \in [0, \infty)$. The empty niches are also scaled, such that we now have $N_E = 1 - \sum_{i=1}^n N_i \in [0, 1]$.

From the steady state considerations discussed in the previous section, we note that the steady state values of I_j^o and A_j^o scale linearly with u_j (Equations (20) and (19)), while the steady state values of N_j^o and N_E^o are independent of u_j , see theorem 3.3. In addition, the steady state value of N_E^o scales with $b_{I_j}^{-1}$. Hence, Assumption 1, stating that the steady state value of N_E^o is small compared to those of N_j^o , I_j^o and A_j^o , can be mathematically interpreted as $\frac{u_j}{b_{I_j}} \ll 1$.

We divide by $b_{I_j} K$ in equation (34b). When we defined K in the text immediately before equations (9), we assumed $K > 1$ to hold in general. Hence $\frac{U}{b_{I_j} K} \leq \frac{u_j}{b_{I_j} K} < \frac{u_j}{b_{I_j}} \ll 1$. Letting $\frac{U}{b_{I_j} K} \rightarrow 0$, we obtain a quasi-steady-state approximation for I_j for all j .

Lemma 4.1: Quasi-steady-state approximation, I_j

The expression

$$I_{j,red} = \frac{\frac{2\gamma}{d_{I_j}} A_j}{\frac{d_{I_j}}{b_{I_j} K} + 1 - \sum_{i=1}^n N_i} \quad (35)$$

is a quasi-steady-state approximation of equation (34b), which is valid under Assumption 1.

Note that $I_{j,red}$ scales with A_j .

Dividing equation (34c) by $r_j + d_{A_j}$, the expression $\frac{U}{r_j + d_{A_j}}$ appears on the left-hand-side. When $b_{A_j} = 0$, the steady state value of A_j^o , equation (19), is simply $\frac{u_j}{r_j + d_{A_j}} N_j^{o*}$. Under Assumption 2, N_j are numerous compared to I_j and A_j , or equivalently, $\frac{u_j}{r_j + d_{A_j}} \ll 1$. Hence $\frac{U}{r_j + d_{A_j}} \leq \frac{u_j}{r_j + d_{A_j}} \ll 1$ suggests the limit $\frac{U}{r_j + d_{A_j}} \rightarrow 0$, and a quasi-steady-state approximation for A_j arises.

Lemma 4.2: Quasi-steady-state approximation, A_j

The expression

$$A_{j,red} = \frac{r_j}{r_j + d_{A_j}} N_j \quad (36)$$

is a quasi-steady-state approximation of equation (34c), which is valid under Assumption 2.

The two quasi-steady-state approximations together give rise to a reduced model, which is valid under Assumption 1 and Assumption 2. We first define the domain of the reduced model.

Definition 4.1: Reduced domain, \mathbb{T}

We define the *reduced domain* as:

$$\mathbb{T} = \{(N_1, \dots, N_n) : N_j \geq 0, j = 1, \dots, n \wedge \sum_{j=1}^n N_j \leq 1\} \quad (37)$$

By substituting A_j in equation (35) by $A_{j,red}$ and subsequently substituting I_j in equation (34a) by the resulting expression for $I_{j,red}$, the reduced model is obtained.

Definition 4.2: Reduced Model

On the domain \mathbb{T} , the *Reduced Model*, is given by:

$$\dot{N}_j = u_j \left(\frac{2\gamma\rho_j(1 - \sum_{i=1}^n N_i)}{\alpha_j + 1 - \sum_{i=1}^n N_i} - 1 \right) N_j \quad (38)$$

where $\rho_j = \frac{r_j}{r_j + d_{A_j}}$ and $\alpha_j = \frac{d_{I_j}}{b_{I_j} K}$. The variables I_j and A_j are given by equations (35) and (36) respectively.

Note that equation (38) is given in unscaled time. We refer to the parameters u_j , ρ_j and α_j as the reduced parameters. While ρ_j and α_j are dimension-free, u_j has dimension $[time]^{-1}$.

4.1 Steady states of the reduced model

The reduced form of the model, equation (38), was obtained through quasi-steady-state approximations of \dot{I}_j and \dot{A}_j , and hence the steady states of the reduced form and the full system are the same. We here derive the steady states for illustration, and show that the stability of the steady states of the reduced form mirrors the full system. In the reduced model, I_j and A_j are obtained from N_j through equations (35) and (36) respectively, so we only consider N_j^* .

As discussed above, the niche-bound cells are scaled with K . Hence $0 \leq \sum_{i=1}^n N_i \leq 1$ must hold generally, and the empty niches are given as $N_E = 1 - \sum_{i=1}^n N_i$.

In steady state, we find that either $N_j^* = 0$ or

$$1 - \sum_{i=1}^n N_i^* = \frac{\alpha_j}{2\gamma\rho_j - 1} \quad (39)$$

assuming $2\gamma\rho_j \neq 1$. Note that the condition $(2\gamma - 1)r_j > d_{A_j}$ described in section 3.2, is equivalent to $2\gamma\rho_j > 1$.

As with the scaled variables, we introduce a scaling of the fitness. The scaled fitness is $f_j = KF_j$, where F_j is the fitness of the full model as defined in definition 3.1. Hence the fitness in terms of the reduced parameters ρ_j and α_j is:

$$f_j = \frac{2\gamma\rho_j - 1}{\alpha_j} \quad (40)$$

and hence equation (39) can be written as $1 - \sum_{i=1}^n N_i^* = f_j^{-1}$.

This leads to the reduced forms of the steady states.

Theorem 4.1: Trivial steady state S_0^* , reduced form

A trivial steady state, denoted S_0^* , always exists. In the trivial steady state $N_j^* = 0, \forall j \in \{1, \dots, n\}$.

When $0 < f_j^{-1} < 1$, a single-clone steady state exists and is feasible.

Theorem 4.2: Single-clone steady state S_j^* , reduced form

Given $2\gamma\rho_j > 1$ and $f_j^{-1} < 1$, a single-clone steady state exists for which $N_j^* > 0$ while $\forall k \neq j : N_k^* = 0$. In particular, $N_j^* = 1 - f_j^{-1}$ and $N_E^* = f_j^{-1}$.

We denote the j^{th} single-clone steady state as S_j^* .

As for the full model, for clones with equal fitness, co-existence steady states also exist, with the condition $\sum_{i \in \mathcal{C}} N_i = 1 - f^{-1}$ where \mathcal{C} is the set of all co-existing clones and f is the fitness for the clones.

Two-clone example

For illustration, the steady states of the two-clone reduced form, ordered as $(N_1^*, N_2^*)^T$, are:

$$S_0^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad S_1^* = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix}, \quad S_2^* = \begin{pmatrix} 0 \\ 1 - f_2^{-1} \end{pmatrix} \quad (41)$$

If $f_1 = f_2$, a line of steady states exists, given as:

$$S_C^* = \begin{pmatrix} \beta \\ 1 - f^{-1} - \beta \end{pmatrix} \quad (42)$$

where $f = f_1 = f_2$ and $\beta \in [0, 1 - f^{-1}]$. Note that both S_1^* and S_2^* are on the line, with $\beta = 1 - f^{-1}$ and $\beta = 0$ respectively.

Local stability of the steady states of the reduced model

To investigate the stability of the steady states of the reduced form, we first define:

$$g_j(N_1, \dots, N_n) = u_j \alpha_j f_j \left(\frac{1 - \sum_{i=1}^n N_i - f_j^{-1}}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right) \quad (43)$$

and

$$h_j(N_1, \dots, N_n) = -2\gamma\rho_j \alpha_j u_j \left(\frac{1}{\alpha_j + 1 - \sum_{i=1}^n N_i} \right)^2 N_j \quad (44)$$

As the functional dependence on N_1, \dots, N_n is clear, we simply write g_j and h_j for the rest of this section.

Note that equation (38) can be written as $\dot{N}_j = g_j N_j$ and that $\frac{\partial g_j}{\partial N_j} = h_j N_j^{-1}$

This allow us to write the Jacobian for the reduced system as:

$$Jac_{redu} = \begin{pmatrix} h_1 + g_1 & h_1 & \dots & h_1 \\ h_2 & h_2 + g_2 & \dots & h_2 \\ \vdots & \vdots & \ddots & \vdots \\ h_n & h_n & \dots & h_n + g_n \end{pmatrix} \quad (45)$$

In the trivial steady state, S_0^* , we find

$$g_j|_{S_0^*} = \frac{\alpha_j u_j}{\alpha_j + 1} (f_j - 1) \quad (46)$$

for all j . The feasibility criteria for the single-clone steady state S_j^* , implies that $f_j > 1$ when S_j^* is feasible. Thus the sign of g_j in the trivial steady state depends on the feasibility of the single-cell steady state.

Evaluating in the j 'th single-clone steady state S_j^* we find $g_j|_{S_j^*} = 0$. For any $k \neq j$, we find that:

$$g_k|_{S_j^*} = \frac{u_k}{\alpha_k + \frac{1}{f_j}} \left(\frac{f_k}{f_j} - 1 \right) \quad (47)$$

Hence the sign of g_k is determined by the relative fitness of the two clones. When clone k is more fit than clone j , $g_k|_{S_j^*} > 0$, and when they have the same fitness, it is zero.

We note that $N_j = 0 \Rightarrow h_j = 0$. Thus $h_j|_{S_0^*} = 0$, as well as $h_k|_{S_0^*} = 0$ for any $k \neq j$. In addition, when $N_j^* > 0$ we find that $h_j < 0$, and hence $h_j|_{S_j^*} < 0$ as well as $h_j|_{S_C^*} < 0$.

Evaluating the Jacobian in the trivial steady state we find:

$$Jac_{redu}|_{S_0^*} = \begin{pmatrix} g_1|_{S_0^*} & 0 & \dots & 0 \\ 0 & g_2|_{S_0^*} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_n|_{S_0^*} \end{pmatrix} \quad (48)$$

Since the matrix is diagonal, the eigenvalues are simply $g_j|_{S_0^*}$, and hence the stability of S_0^* is unstable if any $f_j > 1$, and stable only if for all j , $f_j < 1$, in which case S_0^* is the only feasible equilibrium.

The stability of the single-clone steady state S_j^* follows that of the full model. For illustration, we evaluate the Jacobian in the steady state for the 1st cell-type:

$$Jac_{redu}|_{S_1^*} = \begin{pmatrix} h_1|_{S_1^*} & h_1|_{S_1^*} & \dots & h_1|_{S_1^*} \\ 0 & g_2|_{S_1^*} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_n|_{S_1^*} \end{pmatrix} \quad (49)$$

The matrix is upper triangular and hence the eigenvalues are again easily found. For any $j \neq 1$, the sign of $g_j|_{S_1^*}$ is determined by $\frac{f_j}{f_1} - 1$. When $g_j|_{S_1^*} < 0$, $f_1 > f_j$. Since $h_1|_{S_1^*} < 0$ always, we note that S_1^* is stable when $f_1 > f_j$ holds for all j , and unstable if any $f_j > f_1$. This result holds for any single-clone steady state.

Finally, in a co-existence steady state, S_C^* , where m clones have the same fitness and p clones have lower fitness, the Jacobian has m rows of $h_j|_{S_C^*} < 0$ and p rows of zeros. As an example, a three dimensional case where $f_1 = f_2 > f_3$ we have:

$$Jac_{redu}|_{S_C^*} = \begin{pmatrix} h|_{S_C^*} & h|_{S_C^*} & h|_{S_C^*} \\ h|_{S_C^*} & h|_{S_C^*} & h|_{S_C^*} \\ 0 & 0 & 0 \end{pmatrix} \quad (50)$$

where $h|_{S_C^*} = h_1|_{S_C^*} = h_2|_{S_C^*}$, and hence the eigenvalues are $2h|_{S_C^*} < 0$ and 0 (with algebraic multiplicity 2).

In conclusion, the stability of the reduced model mirrors that of the full model: Fitness of clones determines the stability in such a way that the single-clone steady state for the clone with the highest fitness is stable.

4.2 Change of variables and further simplifications

Biologically, the relative frequency of a malignant clone can be considered a measure for the molecular residual disease in the stem cell compartment.

On the set $\mathbb{T} \setminus S_0^*$ in \mathbb{R}^n , we define the sum of total niche-bound cells:

$$T = \sum_{i=1}^n N_i \quad (51)$$

and the relative frequency of a clone for $j \in \{2, 3, \dots, n\}$ as:

$$C_j = \frac{N_j}{\sum_{i=1}^n N_i} \quad (52)$$

Note that we can calculate C_1 as $C_1 = 1 - \sum_{i=2}^n C_i$. As the trivial steady state $S_0^* = \{(0, \dots, 0)\}$ has been omitted, $T > 0$ holds and C_j is well-defined.

The n differential equations for the reduced model in equation (38) can be given in terms of T and C_j .

Definition 4.3: Transformed Model

On the domain $(0, 1] \times [0, 1]^{n-1}$ the *transformed model* is given as:

$$\dot{T} = T \sum_{i=1}^n g_i(T) C_i \quad (53a)$$

$$\dot{C}_j = \left(g_j(T) \sum_{i=1}^n C_i - \sum_{i=1}^n g_i(T) C_i \right) C_j \text{ for } j = 2, \dots, n \quad (53b)$$

$$C_1 = 1 - \sum_{i=2}^n C_i \quad (53c)$$

where $g_j(T)$ is given as

$$g_j(T) = u_j \alpha_j f_j \left(\frac{1 - T - f_j^{-1}}{\alpha_j + 1 - T} \right) \quad (54)$$

Lemma 4.3: Equivalence of the reduced model and the transformed model

When excluding the trivial steady state S_0^* from the domain of the reduced model, the transformed model is equivalent to the reduced model from definition 4.2 in the sense that solutions of one model can be transformed to equivalent solutions of the other model.

The function

$$\mathcal{F} : (N_1, \dots, N_n) \in \mathbb{T} \setminus S_0^* \subset \mathbb{R}^n \mapsto (T, C_2, \dots, C_n) = \left(\sum_{i=1}^n N_i, \frac{N_2}{\sum_{i=1}^n N_i}, \dots, \frac{N_n}{\sum_{i=1}^n N_i} \right) \in (0, 1] \times [0, 1]^{n-1} \subset \mathbb{R}^n \quad (55)$$

transforms points in the domain of the reduced model to the domain of the transformed model. The inverse function is trivially found as:

$$\mathcal{F}^{-1} : (T, C_2, \dots, C_n) \in (0, 1] \times [0, 1]^{n-1} \subset \mathbb{R}^n \mapsto (N_1, \dots, N_n) = \left(T \left(1 - \sum_{i=2}^n C_i \right), T C_2, \dots, T C_n \right) \in \mathbb{T} \setminus S_0^* \subset \mathbb{R}^n \quad (56)$$

The transformation \mathcal{F} and its inverse are bijective, and hence lemma 4.3 is proven.

The function $g_j(T)$ in equation (54) determines the sign of \dot{T} and \dot{C}_j . $g_j(T) = 0$ only when $T = T_1^* = 1 - f_j^{-1}$, i.e. the steady-state value of N_j in the single-clone steady state, S_1^* . For $T > 1 - f_j^{-1}$, then $g_j(T) < 0$ and

conversely $g_j(T) > 0$ when $T < 1 - f_j^{-1}$. Hence, for $C_j = 1$ and $C_i = 0$ for all $i \neq j$, we find that T will approach the j 'th single-clone steady state:

$$T \xrightarrow[t \rightarrow \infty]{} 1 - f_j^{-1} \quad (57)$$

In the transformed model, the steady states previously discussed correspond to:

$$\bar{S}_1 = \begin{pmatrix} 1 - f_1^* \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \bar{S}_2 = \begin{pmatrix} 1 - f_2^* \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \bar{S}_3 = \begin{pmatrix} 1 - f_3^* \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \quad \bar{S}_n = \begin{pmatrix} 1 - f_n^* \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \quad (58)$$

where \bar{S}_j describes $(T, C_2, C_3, \dots, C_n)$.

Co-existence steady states also exists under the same conditions as before the transformation. We restrict our attention to a two-clone scenario where the dynamics and the co-existence steady states are clearly illustrated.

4.2.1 Total cell-count and relative HSC frequency in a two-clone scenario

While multiple populations of distinct malignant clones can arise during some hematologic malignancies, we now consider malignancies with just one population of malignant stem cells.

Such malignancies can be described as a two-clone scenario, where the one clone models the healthy HSC population and the second clone models the population of malignant stem cells. In a two-clone subsystem, we write C_2 as C to be concise (and hence $C_1 = 1 - C$). Equations (53a) and (53b) can be written as:

$$\dot{T} = [g_1(T)(1 - C) + g_2(T)C] T \quad (59a)$$

$$\dot{C} = (g_2(T) - g_1(T))(1 - C)C \quad (59b)$$

We refer to this model as the *2-clone transformed model*. In section C of the supplementary material, a detailed analysis of the model is presented. We here summarize the most important results and refer the reader to the supplementary material for details.

For equal fitness, $f = f_1 = f_2$, the functions g_1 and g_2 are identical only if all parameters are equal (i.e. $u_1 = u_2$, $\alpha_1 = \alpha_2$ and $f_1 = f_2$). In this particular case, the system reduces to $\dot{C} = 0$ and $\dot{T} = g(T)T$ where $g = g_1 = g_2$.

For $f = f_1 = f_2$ but $\alpha_2 \neq \alpha_1$, the system has a line of steady state for which $T = 1 - f^{-1}$ and $C \in [0, 1]$.

Theorem 4.3: Transformed two-clone system, equal fitness

For $f = f_1 = f_2$, a line of steady states $\bar{S}_C = \{(T, C) = (1 - f^{-1}, \eta)\}$ for all $\eta \in [0, 1]$ exists. The line of steady states is attracting along the T -axis for all C . $\dot{C} = 0$ for all T only if $\alpha_1 = \alpha_2$ and $u_1 = u_2$. Otherwise $T_e^* = 1 + \alpha_1 \alpha_2 \frac{u_2 - u_1}{u_2 \alpha_2 - u_1 \alpha_1}$ exists and is a nullcline for \dot{C} , if $0 < T_e^* < 1$.

Considering now only $f_1 \neq f_2$. The steady states \bar{S}_1 and \bar{S}_2 exist, corresponding to S_1^* and S_2^* respectively. In order (T, C) they are:

$$\bar{S}_1 = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix}, \quad \bar{S}_2 = \begin{pmatrix} 1 - f_2^{-1} \\ 1 \end{pmatrix} \quad (60)$$

These two steady states are the only steady states of the system when $f_1 \neq f_2$, and the global dynamics can be determined.

Theorem 4.4: Global dynamics of the transformed two-clone system

For $f_1 \neq f_2$, the system of equations (59) has just two steady states, \bar{S}_1 and \bar{S}_2 , as defined in equations (60). There are no periodic solutions.

For $f_1 > f_2$, all solutions with $C(0) < 1$ are attracted toward \bar{S}_1 .

For $f_1 < f_2$, all solutions with $C(0) > 0$ are attracted toward \bar{S}_2 .

Details and proof of theorem 4.4 are given in section C of the supplementary material.
Nullclines for \dot{T} and \dot{C} can be determined.

Lemma 4.4: Nullclines for the 2D transformed model

\dot{C} has 2 to 4 nullclines within the domain. These are $C = 0$, $C = 1$, and the two values $T = T_{\pm}$ given in supplementary material C.

T_{\pm} depend on parameters and are not considered nullclines when they are outside the domain.

\dot{T} has a nullcline described by a value T between $1 - f_1^{-1}$ and $1 - f_2^{-1}$, and a corresponding value of $C_{null}(T)$ given as:

$$C_{null}(T) = \frac{g_1(T)}{g_1(T) - g_2(T)} \quad (61)$$

The \dot{C} -nullclines and the \dot{T} -nullcline only cross in \bar{S}_1 and \bar{S}_2 .

In figure 3 some exemplary plots of the (T, C) phase-plane are shown. The nullcline(s) for \dot{C} are shown in red, while the nullcline for \dot{T} is shown in green. Further examples are shown in supplementary C. The figures highlight changes in the relative frequency of the HSC belonging to the two clones which could be difficult to notice when considering solutions of the reduced model for two clones. In particular, one of the simulated scenarios shown in figure 3c starts from initial conditions where $T(0) = 0.3$ and $C(0) = 0.5$. In this scenario, cell counts of both clones, N_1 and N_2 , increase initially, with N_1 increasing faster. By modelling the relative frequency directly, we see that a decrease in relative frequency is followed by an increase when T approaches ≈ 0.65 . Hence even if N_1 is increasing faster than N_2 , the 1st clone is eventually outcompeted by the 2nd clone. If the 2nd clone is a malignant clone, such scenarios could have clinical significance as they could arise following treatment.

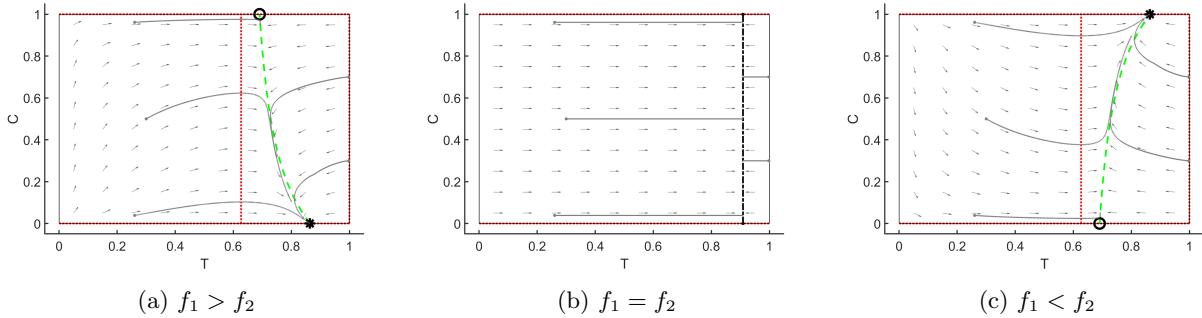


Figure 3: Phase-planes displaying exemplary scenarios of the 2-clone transformed model. The open circle denotes unstable steady states while the asterisk denotes the stable steady states. Grey lines are numerical solutions, with the same initial conditions in the three panels. The green dashed lines are the nullclines for T , while the red dotted lines are the nullclines for C . The arrows show the directional field of flows. As the flow and the values of the steady states depend on the specific parameter values, the figures are examples of typical behaviours, but differences in parameter will influence the specific flows. In panel (b) the fitness is equal, and the line of co-existence steady states, \bar{S}_C is shown as a black dash-dotted line

4.2.2 Additional simplification of a two-clone scenario

We introduce a final simplification of the model, valid close to \bar{S}_1 . Evaluating in \bar{S}_1 , i.e. $T = T_1^*$, we see from equation (54), $g_1(T_1^*) = 0$ and

$$g_2(T_1^*) = u_2 \frac{\alpha_2}{\alpha_2 + f_1^{-1}} \left(\frac{f_2}{f_1} - 1 \right) \quad (62)$$

Assuming $T_1^* \approx T_2^*$, we approximate $\dot{T} = 0$. For $T(0) = T_1^*$, a simplified expression for the relative frequency of the HSC belonging to the two clones arises, in which the expression for $g_2(T_1^*)$ above determines the model behaviour.

Definition 4.4: Logistic Approximation describing the relative frequency of the mutated clone

The *Logistic Approximation Model* is given as

$$\dot{C} = \phi(1 - C)C \quad (63)$$

where $\phi = g_2(T_1^*) = u_2 \frac{\alpha_2}{\alpha_2 + f_1^{-1}} \left(\frac{f_2}{f_1} - 1 \right)$. The sum of HSC is constant with $T = 1 - f_1^{-1}$.

Note that ϕ is not a dimension-free parameter and has units of $[time]^{-1}$, like u_j .

The logistic approximation model is valid close to \bar{S}_1 , and when $T_1^* \approx T_2^*$, i.e. when the steady-state cell-count of second clone is similar to the cell-count of the first clone in steady state. This is the case when $f_1 \approx f_2$. Hence, we assume that the initial stages of a malignancy which only has a minor fitness advantage are described accurately by the logistic approximation model.

Equation (63) is a logistic equation, with steady states $C = 0$ and $C = 1$, corresponding to \bar{S}_1 and \bar{S}_2 respectively. ϕ determines the direction of flow and hence the disease progression. When $\phi > 0$, the steady state $C = 1$ is stable, while $C = 0$ is stable for $\phi < 0$. The sign of ϕ depends solely on $\frac{f_2}{f_1}$. When $f_2 > f_1$, $\phi > 0$, while $f_2 < f_1$ implies $\phi < 0$. Thus the asymptotic behaviour of the logistic approximation is in agreement with previous analysis of the model.

ϕ does not depend on u_1 . This is an artefact of the approximation. When u_1 and u_2 are much different, only little agreement is found with numerical solutions of equations (59).

The logistic approximation model represents the simplest formulation that maintains the asymptotic behaviour of the model. Further simplification, e.g. to exponential growth during initial disease stages, fails to capture both \bar{S}_1 and \bar{S}_2 and the dependence on fitness as determinant of steady-state stability.

5 Model application to clinical data

Consider a specific hematopoietic malignancy; the group of Philadelphia-negative myeloproliferative neoplasms (MPNs), characterized by the $JAK2^{V617F}$ mutation. The allele burden of cells with the $JAK2^{V617F}$ mutation provides an estimate of the residual disease and can be interpreted as the relative frequency of the malignant clone in our model.

We introduce examples of patient-data, previously presented (Pedersen et al., 2020). The patient-data consists of measurements of the $JAK2^{V617F}$ allele burden identified in blood-cells during treatment-free periods and periods with mono-therapy with pegylated Interferon- α . We consider the measured $JAK2^{V617F}$ allele burden indicative of the frequency of mutated HSC among all HSC.

A general rule for applying models to clinical practice is to keep the models as simple as possible without losing the agreement with data to avoid overparametrization. From visual inspection, the data for both the treatment-free disease progression and the development during therapy is approximately logistic. We fit the logistic approximation from equation (63) to data, by minimizing the mean square error, using the MATLAB R2018b function `fminsearch`. In figure 4a depicts data from three untreated MPN-patients along with fits of the logistic equation. In addition, we also fit the initial value of C . Note that the first 200 days are excluded from the fit. This is done since all three patients previously received treatment and a continued decline immediately after treatment-stop is expected (Pedersen et al., 2020).

Figure 4b displays five patients with a range of different initial $JAK2^{V617F}$ allele burdens, receiving treatment with pegylated interferon alfa-2a (*Pegasys*) or pegylated interferon alfa-2b (*PegIntron*). The figure also shows a fit of the logistic approximation model to the data. Baseline-measurement have been time-shifted to coincide with a growth of the $JAK2^{V617F}$ allele burden of $\phi = 0.6$, similar to two of the patients shown in figure 4a.

The data and fits presented exemplify the range of values of ϕ for disease progression and treatment for MPN-diagnosed patients.

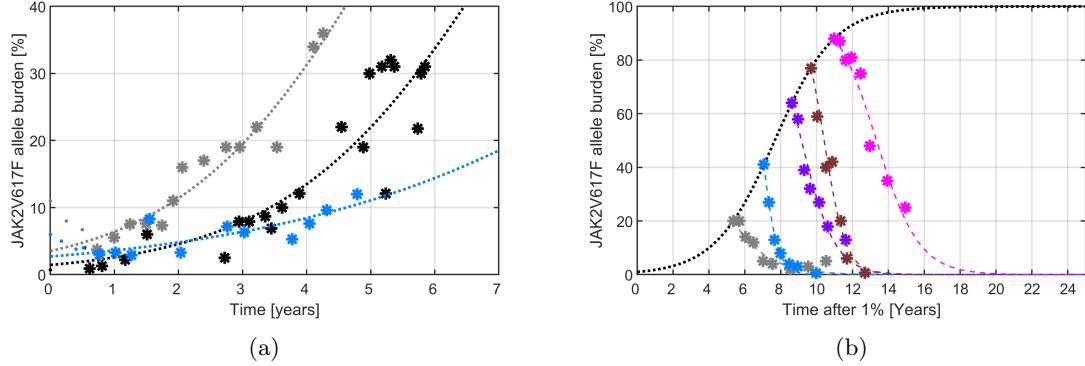


Figure 4: Fit of the logistic approximation to data from MPN-diagnosed patients. Panel (a) depicts data from three patients during a treatment-free period. Note that data during the initial 200 days were excluded from the fit, as delayed responses from previous treatment are expected to influence this period. For the fits, both the initial value of C and the value of ϕ was fitted to the data. Panel (b) shows a typical disease progression in the model in dotted black, along with data from five patients receiving treatment with interferon- α , time-shifted to coincide with the growth. At the time of treatment initialization, the value of ϕ is changed. The new value of ϕ was fitted such that it minimized the least-square error between the model and the given patient. In units of years^{-1} , the three patient-specific values of ϕ in panel (a) during untreated development of MPNs are 0.59, 0.63 and 0.30. For the development during treatment, the five patient-specific values of ϕ in panel (b) are $-0.66, -0.85, -1.1, -1.7$ and -2.3 .

6 Discussion and conclusion

Despite the vital role of HSC in the human body, many questions about their dynamics remain unanswered. The elusive processes within the bone-marrow microenvironment have yet to be fully identified and determined quantitatively. We propose a mechanism-based mathematical model of HSC behaviour based on the state-of-the-art biological knowledge about HSC behaviour. Through mathematical analysis and numerical simulations, the model can help to clarify the general understanding of HSC dynamics, and in future work be used to generate testable hypotheses that can elucidate the open questions surrounding HSC, both in health and in disease. Through a series of considerations and assumptions based on biological knowledge, multiple model simplifications arise. This leads to a *reduced* model (Definition 4.2), a *transformed* model (Definition 4.3) and a *logistic approximation* model (Definition 4.4), in addition to the *full* model.

Mathematical analysis of a proposed model of HSC dynamics

The full form of the proposed model consists of a system of ordinary differential equations, describing HSC-specific niches in the bone-marrow and n distinct clones of HSC. In the model, HSC can be in three distinguishable states; niche-bound, active and inactive. The n clones model separates subpopulations of HSC where cells pertaining to a given clone are assumed to be identical. Considering multiple clones in the model hence corresponds to scenarios where cellular heterogeneity is present in the HSC population, such as in the case of hematologic malignancies as e.g. acute myeloid leukemia or MPN. HSC are modelled to bind reversibly to the niches. Following detachment from the niches, the HSC are in the active state. The self-renewal potential of this active state decreases during divisions outside the niche. After a certain number of divisions, cells enter an inactive state and have to rebind to the niche or differentiate. During rebinding to the niche, the HSC restore their self-renewal potential and can become reactivated. The switching between the niche-bound HSC state, the active state (outside the niche, prone to divide) and inactive state (outside the niche, prone to differentiate or reattach) is the novelty of the proposed model.

A special case of the model has previously been investigated with focus on certain clinical consequences of the modelled mechanisms (Pedersen et al., nd). In contrast, the present work is a thorough mathematical analysis of a more general model.

The proposed model bears similarities to models of HSC found in the literature. Ashcroft et al. (2017) proposed a similar model of HSC dynamics and interaction with the niches, in which detachment from the niche

was also explicitly considered. The detachment was related to a subsequent release into the peripheral blood, related to an enhanced probability of stem cell death, and not related to stem cell division in the way we consider in our model. Instead, Ashcroft et al. (2017) models HSC division such that one cell maintains its position within the niche and the other enters either an empty niche or the peripheral blood. By modelling detachment from the niche as a prerequisite for HSC division, we were able to investigate which effect such detachment could have on clonal competition. Interestingly, our model implies that clonal dominance was independent of detachment rate, which suggests that detachment in the way we modelled it only plays a minor role in competition between clones. We emphasize that our suggestion of detachment as a prerequisite for division is a novel hypothesis, and requires confirmation by experimental evidence. In the work of Stiehl et al. (2020), a model of the HSC-niche system is also considered. Similarly, cell-division occurs in the niche and one of the resulting cells maintains its position in the niche. The model does not consider explicit detachment from the niche, instead free stem cells can dislodge bound stem cells from the niche leading to a cell-cell competition for specific niches. HSC differentiate after a certain number of unsuccessful attempts to occupy a niche. Hence, the model proposed by Stiehl et al. (2020) provides a biological explanation for why HSC differentiate which is lacking in our proposed model. However, a model extension with multiple stages of free cells, similar to the extension discussed in supplementary material A, could allow for a similar interpretation in our model, by modelling an increasing probability for differentiation following each division.

Becker et al. (2019) considers a model of HSC interaction with the niche. Similar to our model, niche-bound cells are considered quiescent and free cells are proliferating. In the model described by Becker et al. (2019), HSC feedback on niche cells plays a significant role, modelling a dynamic relation between HSC and niche cells. In contrast, we implicitly assume that the number of niches are constant in our model, and instead focus on the dynamics of HSC under this condition. Modelling changes in the number of niches could be considered by a more elaborate expression for the dynamics of empty niches, the processes considered by Becker et al. (2019) could be included on our model. However, due to the uncertainty about the exact cells that make up the niche, considering such specifics as reproduction and death rates of the abstract niches we consider would be difficult. Combining the findings of our model with those of Becker et al. (2019), suggest that the biological HSC-niche system is robust, and can return to homeostasis after perturbations of both cell- and niche-count.

Analysis of our model identified an attractive trapping region. Solutions to the system of differential equations with initial conditions outside this region were found to enter the region in finite time. This implies that the model is biologically reasonable and appropriate even when the initial state is perturbed.

Mathematical analysis revealed the equilibria of the system, as well as their local stability. The existence, non-negativity and stability of the steady states were found to depend on an expression consisting of clone-specific parameters, which we denote as the HSC fitness, F_j for the j^{th} clone. A trivial steady state, S_0^* , with no HSC and only empty niche-space always exists. For each clone considered, a single-clone steady state, S_j^* , exists in which the given clone has non-zero cell-counts and all other clones are vanishing. The j^{th} clone has positive cell-counts only when the respective fitness F_j is positive. When any F_j is positive, the trivial steady state is unstable. The single-clone steady state of the clone with the higher fitness is locally stable. Determining the fitness of clones thus reveals which clone is dominant in a competitive setting. If multiple clones have the same fitness, a simplex of steady states exists in which the clones co-exists. Although such situations of co-existence could be mathematically possible, it is biologically unlikely that equal fitness is attainable for extended lengths of time due to the noisy biological environment and other systemic factors. For this reason, the co-existence of steady states can be considered a mathematical artefact rather than a biological relevant feature of the system. The complete local stability considerations of the full $3n + 1$ dimensional model are described in section 3.3.

The stability of the steady states reveals that the fitness can be used to predict the outcome of competition between HSC clones. A given clone outcompetes a less fit clone, leading to extinction of the less fit clone. Comparing this dynamic to dynamics known from ecology sheds light on the nature of competing stem cells as not simply cells that interact freely with the bone-marrow niches, but instead as populations competing for a shared resource, similar to ecological systems. The fitness expression depends solely on the properties of a given clone and its interaction with the bone-marrow niches and thus it is independent of the properties of other clones. For this reason, the fitness of a clone can be determined from single-clone experiments and the result of competition can be predicted without experiments in which multiple clones are present simultaneously. The fitness is independent of the number of niches, K . This implies that the presence of many or few niches has no qualitative effect on the asymptotic behaviour; the clone with the higher fitness will out-compete other clones regardless of niche-counts. It is possible that this independence of the number of niches is a consequence of the assumption that K is constant in the proposed model. Changes to the niche-count K could arise due to e.g. age-related growth of a young individual, a specific HSC-niche interaction as suggested by Becker et al. (2019), or feedback from the immune system in response to hematologic malignancy. Such changes could affect the entire bone-marrow microenvironment, either indirectly by changing niche-counts, or indirectly through other means. This could in turn affect the proposed model and the fitness expression. By considering HSC in isolation from external factors, the proposed model provides an estimate of HSC dynamics based on the process

parameters that we considered most significant.

Biological assumptions imply a model reduction

Based on the biological assumptions that only few niches are empty in homeostasis and that the majority of HSC are bound to the niche, a reduced model is proposed. These assumptions are interpreted mathematically and imply that quasi-steady-state approximations of non-niche-bound HSC are appropriate. For n clones, the full model has $3n$ dynamic variables, whereas the reduced model is given by an n -dimensional system of ODEs, and reduce the number of parameters considered from $6n + 2$ to $3n + 1$. The steady states of the reduced model correspond to the steady state of the full model, and stability considerations reveal that the criteria for stability of the steady states depend on HSC fitness in the same way as in the full model. In the reduced model, parameters are grouped together in aggregated parameters. Thus, it is possible that specific changes to the parameters are not captured by the reduced model, e.g. if the attachment and differentiation of inactive HSC increase equally, the reduced model remains unchanged while minor changes appear in the full model. For comparing the model with experimental or clinical data however, this implies that there can be types of data where the reduced model is identifiable but the full model is not.

Transformation of the model highlights clinical relevance

We presented a transformation of the reduced model to the total HSC cell-count and a measure of the fraction of cells pertaining to a given clone. Under the transformation the trivial steady state is omitted, however the existence and stability of the remaining steady states are identical to those in the full model. The transformed variables can be related to clinical measurements. In particular, many hematologic malignancies arise from a single malignant clone in a setting where the remaining HSC are similar enough in properties to be considered identical. As such, considering the transformation for a two-clone scenario has significant clinical relevance.

The transformed model offers a different way to understand and illustrate HSC dynamics compared to the reduced model. One particular concept with clinical relevance was illustrated in figure 3c. In a scenario where a malignant clone had a higher fitness compared to healthy HSC, we observed that for low initial counts of HSC, it was possible for the relative frequency of the malignant clone to decrease initially before approaching the attracting steady state corresponding to full disease. Clinically, this is interesting as low HSC counts can arise due to treatment such as chemotherapy. Hence, the model suggests that after chemotherapy, a patient could experience a temporary decrease in disease burden, followed by an increasing disease burden. Immediately following the treatment, it could hence wrongly be interpreted that the treatment leads to long-lasting decrease of disease burden, and give false hopes for eradication of the malignancy. Even for scenarios with equal fitness, chemotherapy leading to an equal decrease for all clones can cause changes in relative frequency, as illustrated in supplementary figure A2. It is unclear how these observations apply to clinical practice, but our findings suggest evaluating the response to treatment can be nonlinear and complex.

Further simplification allows for simple comparison to clinical data

A further simplification of the transformed model for two clones may be illustrative as well as applicable in the clinical setting. Under the assumption that HSC counts in steady state do not vary significantly between the healthy state and the full-blown disease state, the transformed model simplifies to a logistic equation for the relative frequency of the malignant clone. A single aggregated parameter, ϕ , determines the dynamics of the relative frequency, revealing how independent parameters influence the general behaviour. The steady states of the logistic approximation were the same as in the full model. Local stability of the steady states was similarly determined by HSC fitness, and the criteria for clonal dominance in the logistic approximation model agree with the full model. The simplification offers an explanation for the success and frequent use of logistic or exponential models to describe hematologic malignancies. Our approach provides a theoretical foundation for otherwise empirical models of disease burden growth and decay.

We identified the parameter ϕ of the logistic equation for the relative frequency for an exemplary subset of a published MPN-patient cohort. Both treatment-free disease progression and periods of interferon- α monotherapy were found to agree well with the logistic approximation. Hence, the model provides an estimate for the development of the disease burden for the patients investigated. The clinical data consisted of measurements of the JAK2^{V617F} allele burden in blood cells. Although the disease burden of blood cells and HSC can be different in principle, evidence suggests that specifically for the JAK2^{V617F} mutation, the disease burdens in mature and immature cell compartments are strongly correlated (Takahashi et al., 2013). For this reason, we assumed the relative frequency of malignant cells to be equal for blood cells and HSC.

The single-parameter logistic expression agreed well with data of MPN patients during and following interferon- α treatment and set a lower bound for how well a more complex model could fit to data. When considering the scenario with two distinct HSC clones, the reduced model has 7 parameters and the full model has a total of 14 parameters. Hence, a fit of either model to the sparse data presented could result in significant

overparametrization. Thus, even if the logistic approximation appeared simplistic, it provided a reasonable estimate for HSC dynamics. Further investigation of how the different forms of the model compares to data may suggest novel *in vitro* or mice experiments to understand and directly quantify the HSC-processes that the model was based on. This could include various HSC labelling assays which could identify properties such as e.g. how many times an active HSC divides before exhaustion (γ) or if labelled cells always divide before attaching to the niche *in vivo*, implying $b_{A_j} = 0$.

Concluding remarks

In conclusion, mathematical analysis of a proposed model of hematopoietic stem cells allowed us to investigate the dynamics of a complex biological system which is inherently difficult to observe experimentally. Through a sequence of model reductions and simplifications based on biological knowledge, a simple logistic expression arose. Long-term persistence of a clone was determined by an expression of stem cell fitness, and criteria for clonal dominance were the same for the reduced and simplified models. By comparing the simplified model to clinical data, the model was found to be in agreement with clinical observations. Our results have clinical implications about hematopoietic stem cells when the various systems of feedback from the blood or immune system are not considered. In this work, we assumed that the effect of such feedbacks had little significance. How these feedbacks, such as e.g. inflammation of the bone-marrow, change the dynamics of the hematopoietic stem cells is an important topic to be investigated in future work.

7 Acknowledgements

This work was initiated during a visit of RKP at Heidelberg University. The authors would like to thank Anna Marciniak-Czochra (Heidelberg University, Germany) for supporting the realization of the research stay. RKP is grateful for the support of the “Produktion, Forskning og Innovation (PFI)” grant from Region Sjælland, Denmark.

Declarations of interest: none

References

- Altrock, P. M., Liu, L. L., and Michor, F. (2015). The mathematics of cancer: Integrating quantitative models. *Nature Reviews Cancer*, 15(12):730–745.
- Andersen, M., Sajid, Z., Pedersen, R. K., Gudmand-Hoeyer, J., Ellervik, C., Skov, V., Kjær, L., Pallisgaard, N., Kruse, T. A., Thomassen, M., Troelsen, J., Hasselbalch, H. C., and Ottesen, J. T. (2017). Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. *PLOS ONE*, 12(8).
- Ashcroft, P., Manz, M. G., and Bonhoeffer, S. (2017). Clonal dominance and transplantation dynamics in hematopoietic stem cell compartments. *PLOS Computational Biology*, 13(10).
- Becker, N. B., Günther, M., Li, C., Jolly, A., and Höfer, T. (2019). Stem cell homeostasis by integral feedback through the niche. *Journal of Theoretical Biology*, 481:100–109.
- Bhattacharya, D., Czechowicz, A., Ooi, A. L., Rossi, D. J., Bryder, D., and Weissman, I. L. (2009). Niche recycling through division-independent egress of hematopoietic stem cells. *The Journal of Experimental Medicine*, 206(12):2837–2850.
- Catlin, S. N., Busque, L., Gale, R. E., Guttorm, P., and Abkowitz, J. L. (2011). The replication rate of human hematopoietic stem cells *in vivo*. *Blood*, 117(17):4460–4466.
- Clapp, G. and Levy, D. (2015). A review of mathematical models for leukemia and lymphoma. *Drug Discovery Today: Disease Models*, 16(1):1–6.
- Colijn, C. and Mackey, M. C. (2005). A mathematical model of hematopoiesis—I. Periodic chronic myelogenous leukemia. *Journal of Theoretical Biology*, 237(2):117–132.
- Dingli, D. and Michor, F. (2006). Successful Therapy Must Eradicate Cancer Stem Cells. *Stem Cells*, 24(12):2603–2610.
- Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svartesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., Sklar, P., Grönberg, H., Hultman, C. M., and

McCarroll, S. A. (2014). Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*, 371(26):2477–2487.

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

Jaiswal, S. and Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science*, 366(6465).

Komarova, N. L. and Wodarz, D. (2007). Effect of cellular quiescence on the success of targeted CML therapy. *PLoS ONE*, 2(10).

Kumar, S. and Geiger, H. (2017). HSC Niche Biology and HSC Expansion Ex Vivo. *Trends in Molecular Medicine*, 23(9):799–819.

Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J. P., Martincorena, I., Anderson, E., O'Neill, L., Stratton, M. R., Laurenti, E., Green, A. R., Kent, D. G., and Campbell, P. J. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*.

Meiss, J. D. (2007). *Differential Dynamical Systems*. Society for Industrial and Applied Mathematics.

Nie, Y., Han, Y.-C., and Zou, Y.-R. (2008). CXCR4 is required for the quiescence of primitive hematopoietic cells. *Journal of Experimental Medicine*, 205(4):777–783.

Nilsson, S. K., Johnston, H. M., Whitty, G. A., Williams, B., Webb, R. J., Denhardt, D. T., Bertoncello, I., Bendall, L. J., Simmons, P. J., and Haylock, D. N. (2005). Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood*, 106(4):1232–1239.

Park, D. S., Akuffo, A. A., Muench, D. E., Grimes, H. L., Epling-Burnette, P. K., Maini, P. K., Anderson, A. R. A., and Bonsall, M. B. (2019). Clonal hematopoiesis of indeterminate potential and its impact on patient trajectories after stem cell transplantation. *PLOS Computational Biology*, 15(4).

Pedersen, R. K., Andersen, M., Knudsen, T. A., Sajid, Z., Gudmand-Hoeyer, J., Dam, M. J. B., Skov, V., Kjær, L., Ellervik, C., Larsen, T. S., Hansen, D., Pallisgaard, N., Hasselbalch, H. C., and Ottesen, J. T. (2020). Data-driven analysis of JAK2 V617F kinetics during interferon-alpha2 treatment of patients with polycythemia vera and related neoplasms. *Cancer Medicine*, 9(6):2039–2051.

Pedersen, R. K., Andersen, M., Skov, V., Kjær, L., Hasselbalch, H. C., Ottesen, J. T., and Stiehl, T. (n.d.). Combining stem cell mobilization with preconditioning: Insights about competition in the stem cell niche from mathematical modelling. *Submitted*.

Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859):105–111.

Roeder, I. and Loeffler, M. (2002). A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. *Experimental Hematology*, 30(8):853–861.

Stiehl, T., Baran, N., Ho, A. D., and Marciniak-Czochra, A. (2015). Cell division patterns in acute myeloid leukemia stem-like cells determine clinical course: A model to predict patient survival. *Cancer Research*, 75(6):940–949.

Stiehl, T., Wang, W., Lutz, C., and Marciniak-Czochra, A. (2020). Mathematical modeling provides evidence for niche competition in human AML and serves as a tool to improve risk stratification. *Cancer Research*.

Takahashi, K., Patel, K. P., Kantarjian, H., Luthra, R., Pierce, S., Cortes, J., and Verstovsek, S. (2013). JAK2 p.V617F detection and allele burden measurement in peripheral blood and bone marrow aspirates in patients with myeloproliferative neoplasms. *Blood*, 122(23):3784–3786.

Vaidya, A. and Kale, V. (2015). Hematopoietic Stem Cells, Their Niche, and the Concept of Co-Culture Systems: A Critical Review. *Journal of stem cells*, 10(1):13–31.

Wang, W., Stiehl, T., Raffel, S., Hoang, V. T., Hoffmann, I., Poisa-Beiro, L., Saeed, B. R., Blume, R., Manta, L., Eckstein, V., Bochtler, T., Wuchter, P., Essers, M., Jauch, A., Trumpp, A., Marciniak-Czochra, A., Ho, A. D., and Lutz, C. (2017). Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia. *Haematologica*, 102(9):1567–1577.

- Wilson, A. and Trumpp, A. (2006). Bone-marrow haematopoietic-stem-cell niches. *Nature Reviews Immunology*, 6(2):93–106.
- Yang, S.-H., Wang, T.-F., Tsai, H.-H., Lin, T.-Y., Wen, S.-H., and Chen, S.-H. (2009). Preharvest hematopoietic progenitor cell counts predict CD34+ cell yields in granulocyte-colony-stimulating factor-mobilized peripheral blood stem cell harvest in healthy donors. *Transfusion*, 50(5):1088–1095.
- Zhang, Y. and Gao, Y. (2016). Novel chemical attempts at ex vivo hematopoietic stem cell expansion. *International Journal of Hematology*, 103(5):519–529.

Supplementary material

A Arbitrary number of division before HSC exhaustion

In general, the number of divisions HSC perform before entering the state of exhaustion can be larger than one. In the following, we present a model where HSC become exhausted after an arbitrary finite number of J divisions. Under certain assumptions, this model is identical to the model derived in the main text. For the sake of simplicity we consider a version of the model describing only a single clone. We subdivide the compartment of active cells A to keep track of the number of divisions cells have performed. By A_i we denote the population of activated HSC that have performed $i - 1$ divisions. Hence, cells that have just detached from the niche belong to compartment A_1 , upon division each cell of compartment A_1 gives rise to two cells of compartment A_2 , et cetera.

The model with J HSC divisions before exhaustion can be written as:

$$\dot{N} = \sum_{i=1}^J (b_i(K - N)A_i) - uN \quad (\text{A.1a})$$

$$\dot{A}_1 = -b_1(K - N)A_1 - (r_1 + d_1)A_1 + uN \quad (\text{A.1b})$$

$$\dot{A}_i = -b_i(K - N)A_i + 2r_{i-1}A_{i-1} - (r_i + d_i)A_i \quad (\text{A.1c})$$

$$\dot{A}_J = -b_J(K - N)A_J + 2r_{J-1}A_{J-1} - d_J A_J \quad (\text{A.1d})$$

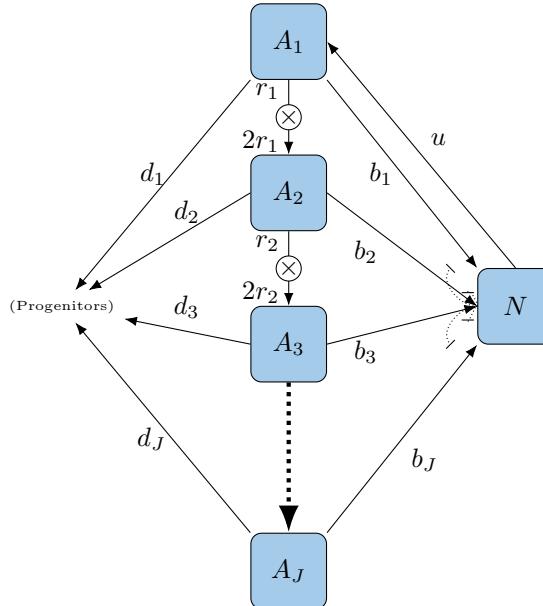


Figure A1: **Compartment diagram of the model with multiple stage of free HSC.** Each box illustrates a variable, with arrows depicting the relations between the variables. The circles with the crosses, \otimes signify a doubling.

By a quasi-steady-state approximation of A_2 type cells, we find that

$$\dot{A}_2 = 0 \Rightarrow A_2 = \frac{2r_1}{r_2 + d_2 + b_2(K - N)} A_1 \quad (\text{A.2})$$

The expression for \dot{A}_3 can then be written as:

$$\dot{A}_3 = 2r_2 \left(\frac{2r_1}{r_2 + d_2 + b_2(K - N)} \right) A_1 - d_3 A_3 - b_3(K - N)A_3 \quad (\text{A.3})$$

For the model presented in our previous work (Pedersen et al., nd), fitting to data revealed that b_{A_1} is very low. Here this corresponds to b_1 . Assuming that this also holds for the intermediate steps, we set $b_2 = 0$. Equation (A.3) then becomes

$$\dot{A}_3 = 4r_1 r_2 A_1 - d_3 A_3 - b_3(K - N)A_3 \quad (\text{A.4})$$

where $\rho_2 = \frac{r_2}{r_2 + d_2}$.

Repeating this for all division events taking place before exhaustion, i.e. assuming $b_i = 0$ and making a quasi-steady-state approximation of step i , the final step, A_J , becomes:

$$\dot{A}_J = 2r_1 \left(\prod_{i=2}^J 2\rho_i \right) A_1 - d_J A_J - b_J(K - N) A_J \quad (\text{A.5})$$

Defining $\gamma = \prod_{i=2}^J 2\rho_i$ we obtain the system described in the main text. In addition, if the probability r_i that a stem cell in compartment A_i divides is approximately equal to the probability that it differentiates, d_i , we can approximate $2\rho_i = \frac{2r_i}{r_i+d_i} \approx 1$. This implies that $\gamma = 1$ which further simplifies the system. Hence the final step, A_J , is defined as presented in the main paper.

B Details about stability of steady states

In section 3.3 of the main text, we consider the stability of the steady states for the full $3n$ system.

For the trivial steady state S_0^* and the single-clone steady states S_j^* , this depends solely on the eigenvalues of the matrix D_j defined as:

$$D_j = \begin{pmatrix} -b_{I_j} I_j - b_{A_j} A_j - u_j & b_{I_j} (K - \sum_{i=1}^n N_i) & b_{A_j} (K - \sum_{i=1}^n N_i) \\ b_{I_j} I_j & -b_{I_j} (K - \sum_{i=1}^n N_i) - d_{I_j} & 2\gamma r_j \\ u_j + b_{A_j} A_j & 0 & -b_{A_j} (K - \sum_{i=1}^n N_i) - r_j - d_{A_j} \end{pmatrix} \quad (\text{B.1})$$

The characteristic polynomial of D_j , can be written as $P_j(\lambda) = \lambda^3 + a_2\lambda^2 + a_1\lambda + a_0$ for which:

$$\begin{aligned} a_2 &= (b_{I_j} + b_{A_j})N_E + b_{I_j}I_j + b_{A_j}A_j + d_{I_j} + d_{A_j} + r_j + u_j \\ a_1 &= b_{A_j}b_{I_j}N_E^2 + b_{I_j}(b_{A_j}A_j + b_{A_j}I_j + b_{A_j}\frac{d_{I_j}}{b_{I_j}} + d_{A_j} + r_j + u_j)N_E \\ &\quad + (d_{A_j} + d_{I_j} + r_j)(b_{A_j}A_j + b_{I_j}I_j + u_j) + d_{I_j}(r_j + d_{A_j}) \\ a_0 &= b_{I_j} [b_{A_j}d_{I_j}I_j - ((2\gamma - 1)r_j - d_{A_j})b_{A_j}A_j - ((2\gamma - 1)r_j - d_{A_j})u_j] N_E \\ &\quad + d_{I_j}(r_j + d_{A_j})(b_{A_j}A_j + b_{I_j}I_j + u_j) \end{aligned}$$

$$\text{where } N_E = K - \sum_{i=1}^n N_i$$

The Routh-Hurwitz criterion can be used to determine if all roots of the polynomial P_j have negative real part. This is the case if and only if $a_2 > 0$, $a_1 > 0$, $a_0 > 0$ and $a_2a_1 > a_0$.

It can be seen from the above expressions that when N_E , A_j and I_j are non-negative, it must hold that $a_2 > 0$ as well as $a_1 > 0$. Hence when D_j is evaluated in either of the feasible steady states, it is the case that $a_2 > 0$ and $a_1 > 0$.

Whenever N_E , A_j and I_j are non-negative, we find that $a_2a_1 - a_0 > 0$. This can be seen by considering the negative parts of $-a_0$. Each negative contribution also appears in a_2a_1 and hence only positive contributions of $-a_0$ and a_2a_1 remain, and thus $a_2a_1 - a_0 > 0$

Thus, the Routh-Hurwitz stability criterion depends only on the sign of a_0 .

When evaluating in the trivial steady state, S_0^* , we find that the characteristic polynomial of $D_j|_{S_0^*}$ has $a_0|_{S_0^*} = -u_j b_{I_j} ((2\gamma - 1)r_j - d_{A_j}) (K - \frac{1}{F_j})$. The requirement for feasibility of a single-clone steady state S_j^* implies that $(2\gamma - 1)r_j - d_{A_j} > 0$ as well as $K - \frac{1}{F_j} > 0$, and hence if the j 'th single-clone steady state is feasible, $a_0|_{S_0^*} < 0$, and hence the Routh-Hurwitz stability criterion does not hold and $D_j|_{S_0^*}$ has at least one eigenvalue with positive real part.

Evaluating in the single-clone steady state, S_j^* , we find $a_0|_{S_j^*} = -a_0|_{S_0^*}$, and the Routh-Hurwitz stability criterion holds. Hence, all eigenvalues of $D_j|_{S_j^*}$ have negative real part.

Finally we consider $D_j|_{S_k^*}$ where $j \neq k$, i.e. the matrix related to the j 'th clone evaluated at the k 'th single-clone steady state. Again, assuming the j 'th clone has feasible steady states, only the sign of a_0 has to be considered.

To determine the eigenvalues of $D_j|_{S_k^*}$ for $j \neq k$, we find that the constant term of the characteristic polynomial is:

$$a_{0,D_j|_{S_k^*}} = u_j b_{I_j} ((2\gamma - 1)r_j - d_{A_j}) \frac{1}{F_k} \left(\frac{F_k}{F_j} - 1 \right) \quad (\text{B.2})$$

If $F_k > F_j$ we find $a_{0,D_j|S_k^*} > 0$, and the Routh-Hurwitz stability criterion holds, and the eigenvalues of $D_j|_{S_k^*}$ all have negative real part. Conversely, when $F_k < F_j$, the criterion does not hold, and at least one eigenvalue has positive real part.

Co-existence steady state

When two or more cell-type have equal fitness, a simplex of steady states exists, S_C^* . We here show that the system has an eigenvector along the simplex, and that the corresponding eigenvalue is zero. For simplicity, we show this for the case where only two cell-types are considered.

In this particular case, the Jacobian is given as:

$$Jac = \begin{pmatrix} D_1 & G_1 \\ G_2 & D_2 \end{pmatrix} \quad (\text{B.3})$$

When projecting the simplex of steady states onto a N_1, N_2 plane, it is a line with slope -1 . The particular non-zero eigenvector corresponding to the eigenvalue zero can be written as:

$$V_0 = \begin{pmatrix} 1 \\ \nu_{I_1} \\ \nu_{A_1} \\ -1 \\ -\nu_{I_2} \\ -\nu_{A_2} \end{pmatrix} \quad (\text{B.4})$$

where ν_{A_j} and ν_{I_j} are defined as

$$\nu_{A_j} = \frac{u_j}{b_{A_j} \frac{1}{F_j} + r_j + d_{A_j}} \quad (\text{B.5})$$

and

$$\nu_{I_j} = \frac{2\gamma r_j u_j}{\left(r_j + d_{A_j} + b_{A_j} \frac{1}{F_j}\right) \left(d_{I_j} + b_{I_j} \frac{1}{F_j}\right)} = \frac{2\gamma r_j \nu_{A_j}}{d_{I_j} + b_{I_j} \frac{1}{F_j}} \quad (\text{B.6})$$

Note that the steady state values of A_j and I_j can be written in general as:

$$A_j^* = \nu_{A_j} N_j^* \quad (\text{B.7})$$

and

$$I_j^* = \nu_{I_j} N_j^* \quad (\text{B.8})$$

The Jacobian multiplied by the eigenvector V_0 yields:

$$Jac V_0 = \begin{pmatrix} -u_1 + \nu_{I_1} b_{I_1} N_E + \nu_{A_1} b_{A_1} N_E \\ -\nu_{I_1} (b_{I_1} N_E + d_{I_1}) + \nu_{A_1} 2\gamma r_1 \\ u_1 - \nu_{A_1} (b_{A_1} N_E + r_1 + d_{A_1}) \\ -u_2 + \nu_{I_2} b_{I_2} N_E + \nu_{A_2} b_{A_2} N_E \\ -\nu_{I_2} (b_{I_2} N_E + d_{I_2}) + \nu_{A_2} 2\gamma r_2 \\ u_2 - \nu_{A_2} (b_{A_2} N_E + r_2 + d_{A_2}) \end{pmatrix} \quad (\text{B.9})$$

For existence of the co-existence steady state, we have $F = F_1 = F_2$. Evaluating at S_C^* we find:

$$Jac|_{S_C^*} V_0 = \begin{pmatrix} -u_1 + \nu_{I_1} b_{I_1} F^{-1} + \nu_{A_1} b_{A_1} F^{-1} \\ -\nu_{I_1}(b_{I_1} F^{-1} + d_{I_1}) + \nu_{A_1} 2\gamma r_1 \\ u_1 - \nu_{A_1}(b_{A_1} F^{-1} + r_1 + d_{A_1}) \\ -u_2 + \nu_{I_2} b_{I_2} F^{-1} + \nu_{A_2} b_{A_2} F^{-1} \\ -\nu_{I_2}(b_{I_2} F^{-1} + d_{I_2}) + \nu_{A_2} 2\gamma r_2 \\ u_2 - \nu_{A_2}(b_{A_2} F^{-1} + r_2 + d_{A_2}) \end{pmatrix} \quad (\text{B.10})$$

All terms of this vector are zero. This can be seen by substituting F with $F_1 = \frac{b_{I_1}((2\gamma-1)r_1-d_{A_1})}{d_{I_1}(r_1+d_{A_1})}$ in the first three components and with $F_2 = \frac{b_{I_2}((2\gamma-1)r_2-d_{A_2})}{d_{I_2}(r_2+d_{A_2})}$ in the bottom three components. Thus, we find $Jac|_{S_C^*} V_0 = 0V_0$, and the vector V_0 is indeed a non-zero eigenvector of $Jac|_{S_C^*}$, with the corresponding eigenvalue of 0.

C Details about analysis of the transformed model

In section 4.2 we present a transformation of the reduced model. When only two distinct clones are considered, the *2-clone transformed model* is given as:

$$\dot{T} = (g_1(T)(1-C) + g_2(T)C)T \quad (\text{C.1a})$$

$$\dot{C} = (g_2(T) - g_1(T))(1-C)C \quad (\text{C.1b})$$

where $g_j = u_j \alpha_j f_j \frac{1-T-f_j^{-1}}{\alpha_j+1-T}$. Only solutions to equations (C.1) that exists on the domain $\mathbb{D} = \{(T, C) \in (0, 1] \times [0, 1]\}$ are considered. In this section, we present a detailed analysis of equations (C.1).

Initially, note that solutions are indeed restricted to the domain \mathbb{D} . This can be observed by considering equations (C.1) along the boundaries of the domain. We first consider the lower boundary of T by taking $T \rightarrow 0^+$:

$$\lim_{T \rightarrow 0^+} \frac{\dot{T}}{T} = g_1(0)(1-C) + g_2(0)C \quad (\text{C.2})$$

In general, we assume that feasible steady states exists for both clones, and hence $0 < f_1^{-1} < 1$ and $0 < f_2^{-1} < 1$. This implies that $g_1(0) > 0$ and $g_2(0) > 0$, and hence $\lim_{T \rightarrow 0^+} \frac{\dot{T}}{T} > 0$, and trajectories of solutions with initial conditions within \mathbb{D} do not approach $T = 0$. The restrictions of the remaining three boundaries are determined by evaluating the differential equations at the boundaries:

$$\dot{T}|_{T=1} = g_1(1)(1-C) + g_2(1)C = -u_1(1-C) - u_2C < 0 \quad (\text{C.3})$$

$$\dot{C}|_{C=0} = 0 \quad (\text{C.4})$$

$$\dot{C}|_{C=1} = 0 \quad (\text{C.5})$$

Equilibria

While the steady states for the n -clone transformed model were presented in the main text, we briefly reiterate the steady states of the 2-clone form. $\dot{C} = 0$ implies either $C = 0$, $C = 1$ or $g_2(T) = g_1(T)$. We consider these in order. For $C = 0$, $\dot{T} = g_1(T)T$, which has just one zero, namely $T = T_1^* = 1 - f_1^{-1}$. This results in the equilibrium $\overline{S}_1 = (1 - f_1^{-1}, 0)$. Similarly, $C = 1$ implies $\dot{T} = g_2(T)T$ which has zero $T = T_2^* = 1 - f_2^{-1}$, resulting in the equilibrium $\overline{S}_2 = (1 - f_2^{-1}, 1)$.

Lastly we consider co-existence steady states, \overline{S}_C , where $0 < C_c^* < 1$. We note that \dot{T} can be re-written such that:

$$\dot{T} = (g_1(T) + (g_2(T) - g_1(T))C)T \quad (\text{C.6})$$

When $g_2(T) = g_1(T)$, we can write $\dot{T} = g_1(T)T$, which is zero only if $T = T_1^*$. Hence, an equilibrium with $0 < C_c^* < 1$ must have $T = T_1^*$. Since $g_j(T) = 0$ only holds for $g_j(1 - f_j^{-1})$, $\dot{T} = 0$ implies $g_2(T_1^*) = 0$, and hence $f_1 = f_2$. Thus, co-existence steady states \overline{S}_C only exist when $f_1 = f_2$. For $f_1 \neq f_2$, equations (C.1) has just two equilibria, \overline{S}_1 and \overline{S}_2 .

Lemma C.1: Equilibria of the 2-clone transformed model

For $f_1 \neq f_2$, equations (C.1) has just two equilibria, $\overline{S_1}$ and $\overline{S_2}$. In order (T, C) these are:

$$\overline{S_1} = \begin{pmatrix} 1 - f_1^{-1} \\ 0 \end{pmatrix} \quad \text{and} \quad \overline{S_2} = \begin{pmatrix} 1 - f_2^{-1} \\ 1 \end{pmatrix} \quad (\text{C.7})$$

When $f_1 = f_2$, a line of steady states $\overline{S_C}$ exist:

$$\overline{S_C} = \begin{pmatrix} 1 - f_1^{-1} \\ \nu \end{pmatrix} \quad (\text{C.8})$$

where $\nu \in [0, 1]$.

For $f_1 \neq f_2$, the two equilibria $\overline{S_1}$ and $\overline{S_2}$ exist on the boundary of \mathbb{D} . Hence an orbit of solutions cannot enclose either of the equilibria. For $f = f_1 = f_2$, the line of steady states $\overline{S_C}$ divides \mathbb{D} in two sub-domains, one with $T \in (0, 1 - f^{-1})$ and one with $T \in (1 - f^{-1}, 1]$. Trajectories in one sub-domain cannot enter the other as it would require leaving the domain \mathbb{D} or crossing $\overline{S_C}$, neither of which is possible. Hence, periodic orbits surrounding $\overline{S_C}$ cannot exist. As any periodic orbit in $2D$ must have at least one interior equilibrium due to index theory (Meiss, 2007), there can be no periodic solutions to equations (C.1). Thus, due to the *Poincaré-Bendixson* theorem, any solution trajectory must converge to either $\overline{S_1}$ or $\overline{S_2}$ when $f_1 \neq f_2$, or to a point on $\overline{S_C}$ for $f_1 = f_2$.

Global attraction of steady states

It can be shown trivially that when $f_1 = f_2$, $\dot{T} > 0$ for all $T \in (0, 1 - f^{-1})$ and $\dot{T} < 0$ for $T \in (1 - f^{-1}, 1]$. Hence the points on the line $\overline{S_C}$ are attractive for all initial conditions. Note however, as we show below, C is not necessarily constant when $f_1 = f_2$, see figure A2.

For $f_1 \neq f_2$, an attractive trapping region can be found. For brevity, we consider the case $f_1 < f_2$, as the case $f_1 > f_2$ follows by symmetry.

Defining

$$T_{2,\epsilon} = 1 - f_2^{-1} + \epsilon \quad (\text{C.9})$$

for any $\epsilon \geq 0$, we observe that $g_1(T_{2,\epsilon}) < 0$ and $g_2(T_{2,\epsilon}) \leq 0$, with equality only for $\epsilon = 0$. Hence

$$\dot{T}|_{T_{2,\epsilon}} = (g_1(T_{2,\epsilon})(1 - C) + g_2(T_{2,\epsilon})C) T_{2,\epsilon} \geq 0 \quad (\text{C.10})$$

where $\dot{T}|_{T_{2,\epsilon}} = 0$ only when both $C = 1$ and $\epsilon = 0$.

We present an argument analogous to our definition of the trapping region of the full model in section 3.1 of the main text.

Consider an interval $\mathbb{T} = [T_{2,\epsilon}, T_{2,\epsilon} + \Delta]$ with $\Delta > 0$ on which $\dot{T} < 0$. Due to the extreme value theorem, there exists a finite η such that $-\dot{T} \geq \eta, \forall T \in \mathbb{T}$, since \mathbb{T} is a compact subset. In particular

$$\inf_{\mathbb{T}} (-\dot{T}) = \min_{\mathbb{T}} (-\dot{T}) = \eta > 0 \quad (\text{C.11})$$

and hence $\frac{1}{-\dot{T}} \leq \frac{1}{\eta}$ holds for all $T \in \mathbb{T}$.

To determine the time t_Δ it takes for the flow from $T(0) = T_{2,\epsilon} + \Delta$ to reach $T(t_\Delta) = T_{2,\epsilon}$ we integrate over the interval:

$$t_\Delta = \int_0^{t_\Delta} dt = \int_{T(0)}^{T(t_\Delta)} \frac{1}{\dot{T}} dT = \int_{T_{2,\epsilon}}^{T_{2,\epsilon} + \Delta} \frac{1}{-\dot{T}} dT \leq \frac{\Delta}{\eta} \quad (\text{C.12})$$

and hence t_Δ is finite. Thus for any $\Delta > 0$, any point with $T(0) = T_{2,\epsilon} + \Delta$ will move along a trajectory that enters the region $T \leq T_{2,\epsilon}$ in finite time for any $\epsilon > 0$.

Defining

$$T_{1,-\epsilon} = 1 - f_1^{-1} - \epsilon \quad (\text{C.13})$$

a similar argument can be made that any point with $T(0) < T_{1,-\epsilon}$ enter the region in finite time. Hence there is a region $\mathbb{M} = \{(T, C) \in [1 - f_1^{-1}, 1 - f_2^{-1}] \times (0, 1)\} \subset \mathbb{D}$ which is an attractive trapping region for all points in $\mathbb{D} \setminus \mathbb{M}$

Observe that for all $T \in \mathbb{M}$, it holds that $g_1(T) \leq 0$ with equality only in $T = T_1^*$, and $g_2(T) \geq 0$ with equality only in $T = T_2^*$. Hence $g_2(T) - g_1(T) > 0$ since $T_1^* \neq T_2^*$. Thus, for $T \in \mathbb{M}$:

$$\lim_{C \rightarrow 0^+} \frac{\dot{C}}{C} = \lim_{C \rightarrow 0^+} (g_2(T) - g_1(T)) (1 - C) = g_2(T) - g_1(T) > 0 \quad (\text{C.14})$$

and hence $\overline{S_1}$ is unstable.

We define $\mathbb{M}_\epsilon = \{(T, C) \in [1 - f_1^{-1}, 1 - f_2^{-1}] \times (\epsilon, 1)\}$ for any $\epsilon > 0$. The region \mathbb{M}_ϵ , only contains one equilibrium, $\overline{S_2}$. As there can be no periodic solutions, we have determined that $\overline{S_2}$ is attractive for all trajectories with $C(0) > 0$. As \mathbb{M} is an attractive trapping region for all points in $\mathbb{D} \setminus \mathbb{M}$, this result extends to all points in \mathbb{D} . For the case $f_1 > f_2$, a similar construction with $C \in [0, 1 - \epsilon]$ for all $\epsilon > 0$ can be made, implying that $\overline{S_1}$ is an attractor for all initial conditions with $C(0) < 1$.

Lemma C.2: Limit sets of solutions

For $f_1 < f_2$, there are two equilibria, $\overline{S_1}$ and $\overline{S_2}$. $\overline{S_2}$ is an attractor for all points with $C(0) > 0$.

For $f_1 > f_2$, there are two equilibria, $\overline{S_1}$ and $\overline{S_2}$. $\overline{S_1}$ is an attractor for all points with $C(0) < 1$.

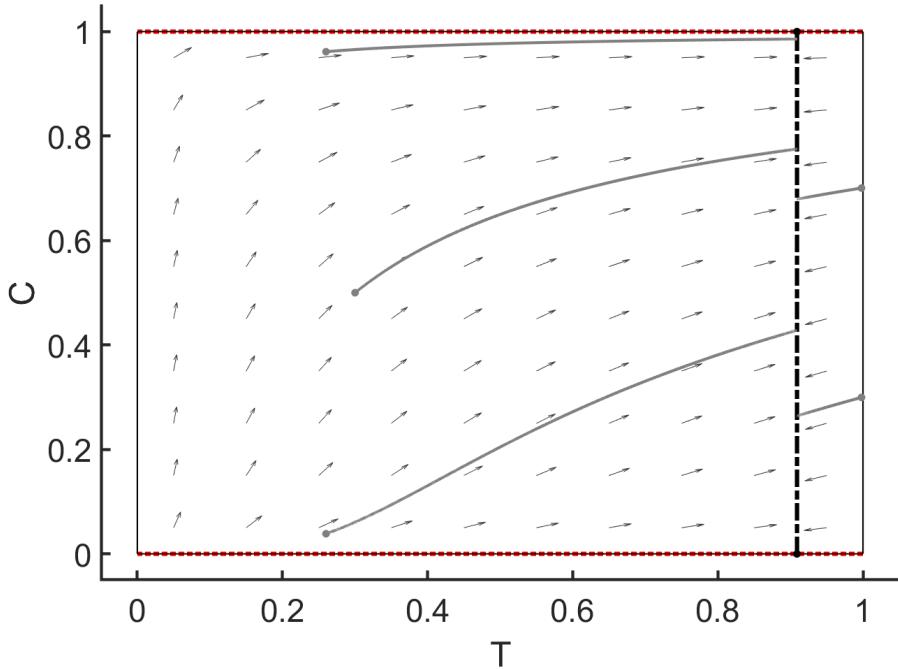


Figure A2: **Illustrative example of an equal fitness scenario.**

In the simulation, $\rho_1 = \rho_2$ and $\alpha_1 = \alpha_2$ and hence $f_1 = f_2$. However $u_2 = 5u_1$, which causes the abundance of the clone with the higher fitness to increase as solutions approach the co-existence equilibrium.

Conditions for $g_2(T) = g_1(T)$.

The expression $g_2(T) = g_1(T)$ can be written as a second order polynomial in T . Defining $\mu_j = u_j \alpha_j f_j$:

$$g_2(T) = g_1(T) \quad (\text{C.15})$$

$$u_2 \alpha_2 f_2 \frac{1 - T - f_2^{-1}}{\alpha_2 + 1 - T} = u_1 \alpha_1 f_1 \frac{1 - T - f_1^{-1}}{\alpha_1 + 1 - T} \quad (\text{C.16})$$

$$\mu_2 \frac{1 - T - f_2^{-1}}{\alpha_2 + 1 - T} = \mu_1 \frac{1 - T - f_1^{-1}}{\alpha_1 + 1 - T} \quad (\text{C.17})$$

$$aT^2 + bT + c = 0 \quad (\text{C.18})$$

where

$$a = \mu_2 - \mu_1 \quad (\text{C.19a})$$

$$b = -\mu_2 ((1 - f_2^{-1}) + (\alpha_1 + 1)) + \mu_1 ((1 - f_1^{-1}) + (\alpha_2 + 1)) \quad (\text{C.19b})$$

$$c = \mu_2(1 - f_2^{-1})(\alpha_1 + 1) - \mu_1(1 - f_1^{-1})(\alpha_2 + 1) \quad (\text{C.19c})$$

If $a = 0$, $b = 0$ and $c = 0$, then g_1 and g_2 are identical functions. Assuming $\mu_1 = \mu_2 = \mu$, $a = 0$ is fulfilled. Equation (C.19b) then becomes:

$$b = \mu (f_2^{-1} - f_1^{-1} + \alpha_2 - \alpha_1) \quad (\text{C.20})$$

For $b = 0$, it must hold that $f_1^{-1} = f_2^{-1} + \alpha_2 - \alpha_1$. This allows for a simplification of c :

$$c = \mu (f_2^{-1} + \alpha_1) (\alpha_2 - \alpha_1) \quad (\text{C.21})$$

And hence $\alpha_2 = \alpha_1$ implies $c = 0$, since all parameters are positive.

Writing up $g_2(T) = g_1(T)$ again, with $\alpha = \alpha_1 = \alpha_2$, we obtain:

$$\mu_2 \frac{1 - T - f_2^{-1}}{\alpha + 1 - T} = \mu_1 \frac{1 - T - f_1^{-1}}{\alpha + 1 - T} \quad (\text{C.22})$$

$$\mu_2(1 - T - f_2^{-1}) = \mu_1(1 - T - f_1^{-1}) \quad (\text{C.23})$$

$$\mu_2(1 - T - f_2^{-1}) - \mu_1(1 - T - f_1^{-1}) = 0 \quad (\text{C.24})$$

$$-(\mu_2 - \mu_1)T + \mu_2(1 - f_2^{-1}) - \mu_1(1 - f_1^{-2}) = 0 \quad (\text{C.25})$$

which is linear in T . If $\mu_1 = \mu_2$, the expression simplifies to $f_1 = f_2$.

Finally, when both $f_1 = f_2$ and $\alpha_1 = \alpha_2$, the functions g_1 and g_2 are identical only when $u_1 = u_2$. Thus, it is only possible for $g_2(T) = g_1(T)$ to hold for all T when all parameters are equal.

Lemma C.3: Condition for $g_2 = g_1$

The functions $g_1(T)$ and $g_2(T)$ are equal for all T if and only if $f_1 = f_2$, $\alpha_1 = \alpha_2$ and $u_1 = u_2$, in which case g_1 and g_2 are identical functions.

We emphasize that $f_1 = f_2$ is *not* sufficient for $g_2(T) = g_1(T)$ to hold for all T .

Nullclines for \dot{T} and \dot{C}

Assuming that g_1 and g_2 are not identical, and hence that $f_1 \neq f_2$, $\alpha_1 \neq \alpha_2$ and $u_1 \neq u_2$, it is possible for the second order polynomial of equation (C.18) to have one or two real roots between 0 and 1. These values, T_{\pm} , are nullclines for \dot{C} , shown in red in the figures of this section and in the main text. It is possible for both T_{\pm} to be between 0 and 1 simultaneously, see figure A3, and for neither to be between 0 and 1.

For equal fitness, $f = f_1 = f_2$, the differential equation for T can be written:

$$\dot{T} = \left(\frac{u_1 \alpha_1}{\alpha_1 + 1 - T} (1 - C) + \frac{u_2 \alpha_2}{\alpha_2 + 1 - T} C \right) (1 - T - f^{-1}) f T \quad (\text{C.26})$$

$\dot{T} = 0$ has two solutions: $T_f^* = 1 - f^{-1}$ and

$$T_o^* = 1 + \alpha_1 \alpha_2 \frac{u_1 (1 - C) + u_2 C}{u_1 \alpha_1 (1 - C) + u_2 \alpha_2 C} \quad (\text{C.27})$$

Since $C \in [0, 1]$, and all parameters are positive, $T_o^* > 1$ and hence not feasible. Thus, in a scenario with equal fitness, the sign of \dot{T} only changes at $T_f^* = 1 - f^{-1}$.

For $f_1 \neq f_2$, the sign of \dot{T} is determined by $g_1(T)(1 - C) + g_2(T)(C)$, from which a function $C_{null}(T)$ can be defined:

$$C_{null}(T) = \frac{g_1(T)}{g_1(T) - g_2(T)} \quad (\text{C.28})$$

$\dot{T} = 0$ for all $(T, C) = (T_n, C_{null}(T_n))$ and hence $C_{null}(T)$ is a nullcline for T . In the two single-clone steady states, we find $C_{null}(1 - f_1^{-1}) = 0$ and $C_{null}(1 - f_2^{-1}) = 1$.

Observe that

$$C_{null}(T) \xrightarrow[g_1(T) \rightarrow g_2(T)]{} \pm \infty \quad (\text{C.29})$$

and thus the $C_{null}(T)$ is undefined for $T = T_{\pm}$ and the nullclines do not cross, as expected from the analysis of steady states above.

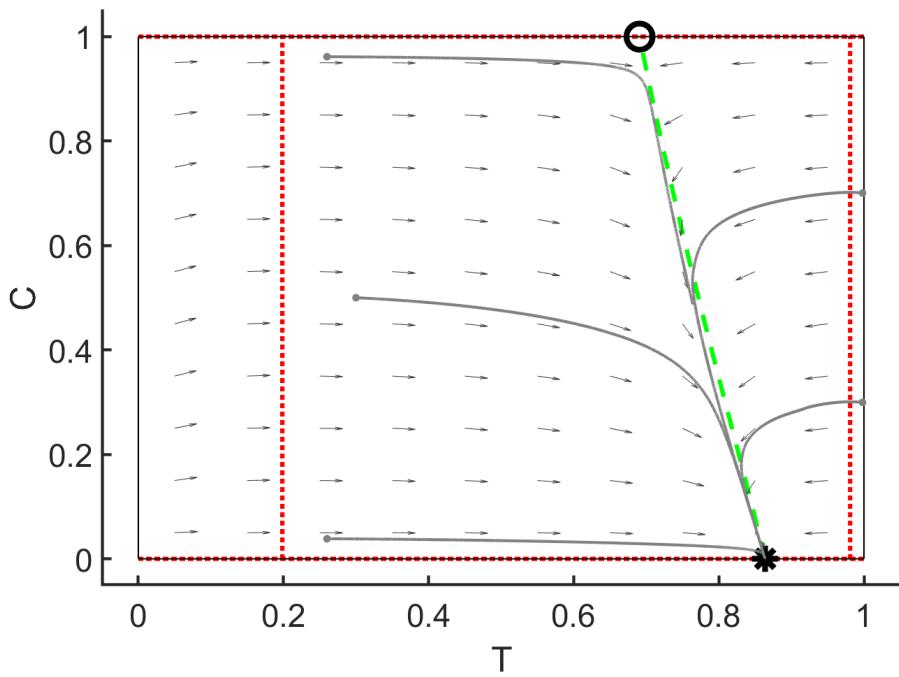


Figure A3: **Phase plane illustration of two \dot{C} -nullclines.** The figure illustrates a particular choice of parameters such that the two roots of $g_2(T) - g(1) = 0$, T_{\pm} , are both between 0 and 1. The particular choice of parameter for the figure were $\rho_1 = 0.529$, $\rho_2 = 0.627$, $\alpha_1 = 0.008$, $\alpha_2 = 0.079$, $u_1 = 0.113$ and $u_2 = 0.038$.

Appendix B

Poster presentations

Treatment of Blood Cancers and the Importance of Quiescence

Modelling of Quiescent Stem Cells

in Relation to Myeloproliferative Neoplasms

Rasmus Kristoffer Pedersen (rakrpe@ruc.dk)

Cancitis Group, dirac.ruc.dk/cancitis

IMFUFA, Department of Science and Environment, Roskilde University
Supervisors: Johnny T. Ottesen, Hans C. Hasselbalch and Morten Andersen



Background

Myeloproliferative Neoplasms (MPNs) is a family of diseases in the bone marrow, which lead to an excessive production of hematopoietic cells. Through a mechanism-based modelling approach, the behaviour of the hematopoietic system has been modelled with satisfactory results [1]. Preliminary results comparing the results of the model to clinical data show great promise for future prediction of the development of the disease. This work aims to extend the model, based on mechanisms suggested by clinical intuition, particularly, the addition of *quiescent* (or *dormant*) stem cells.

Assessment of clinical data

While the mathematical model considers different types of cells, clinical decisions often have to rely on certain indicators of the disease level. On such indicator is the *JAK2 allele burden*, shown to be well associated with the MPN diagnosis [2]. The JAK2 allele burden measures the ratio of mutated mature cells to the total number of mature cells. In the mathematical model, this would correspond to $\frac{y_1}{x_1 + y_1}$.

In figure 1, previously unpublished time-series measurements of the JAK2 allele burden are shown for patient undergoing treatment with *Interferon- α* .

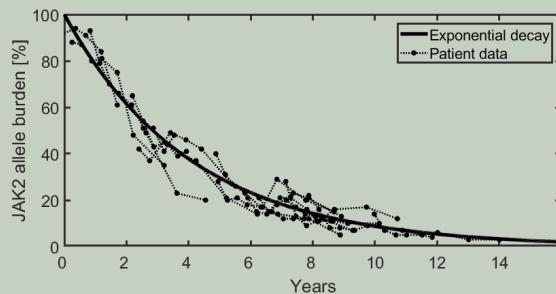


Figure 1: Simple decay rate of JAK2 allele-burden for patients treated with *Interferon- α*

The development is shown to follow that of an exponential decay with a half-life of 2.85 years, and thus, the development of the disease can easily be estimated, regardless of the specific level.

However, for a small group of patients, the behaviour of the JAK2 allele burden during treatment with *Interferon- α* displays a delayed reaction, with a short period of growth before the decay starts. These patients are shown in figure 2, with the simple treatment response curve added for comparison.

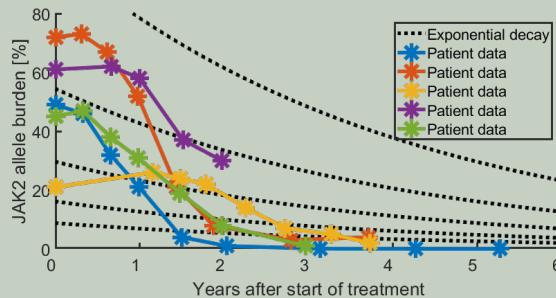


Figure 2: JAK2 allele-burden for the delayed reaction group of patients treated with *Interferon- α*

References

- [1] M. Andersen, Z. Sajid, R. K. Pedersen, J. Gudmand-Høyer, C. Ellervik, V. Skov, L. Kjær, N. Pallisgaard, T. A. Kruse, M. Thomassen, J. Troelsen, H. C. Hasselbalch, and J. T. Ottesen. Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. *PLOS ONE*, 12(8):1–18, 08 2017.
- [2] T. S. Larsen, N. Pallisgaard, M. B. Møller, and H. C. Hasselbalch. The JAK2 V617F allele burden in essential thrombocythemia, polycythemia vera and primary myelofibrosis - Impact on disease phenotype. *European Journal of Haematology*, 79(6):508–515, 2007.

Mathematical model

The Cancitis model [1] comprises a system of six coupled ordinary differential equations (ODE's). The model consists of the healthy stem cells (x_0) and differentiated healthy blood cells (x_1), while MPN-mutated counterparts are represented by y_0 and y_1 . The debris of dead cells is denoted a , while the dimensionless number s represent the load on the immune system. The proposed extension adds two additional ODE's representing quiescent healthy stem cells and MPN-mutated stem cells, denoted x_q and y_q respectively. The model is shown schematically in figure 3.

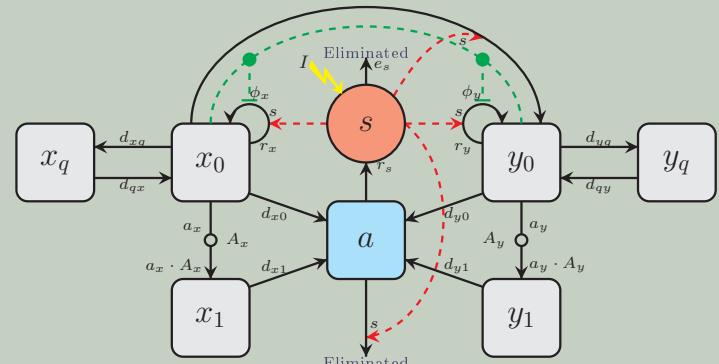


Figure 3: Compartment diagram

The model has been shown to describe the time-development of MPN's well, while demonstrating the effects of inflammation (determined by parameter I). This type of ODE model is prevalent in the literature, and comparison with state-of-the-art models are being done.

Results

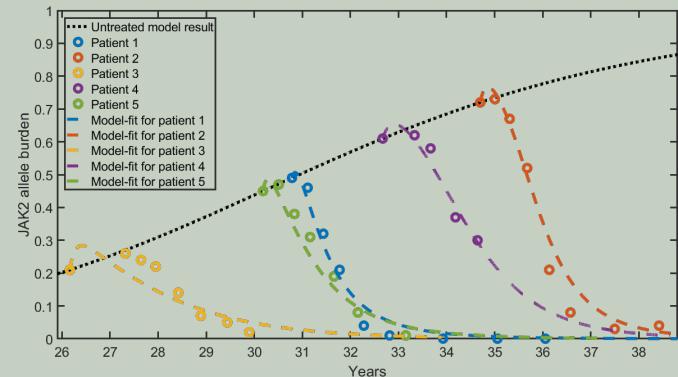


Figure 4: Model simulations fitted with the five patients after untreated growth

Through analysis of literature and communications with clinical professionals, it was determined that the parameters mostly influenced during treatment with *Interferon- α* , were the self-renewal rates r_x and r_y , as well as the activation rates of the quiescent stem cells. Modifying these rates yields in silico treatments fitting well with data, as shown in figure 4.

Conclusions and future perspectives

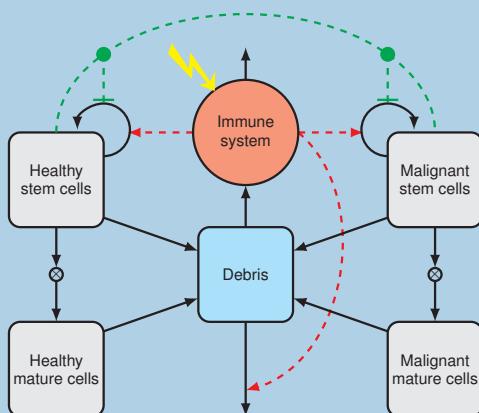
The addition of the quiescent stem-cell compartments allows for an in silico treatment more closely resembling the clinical intuition about the effects of *Interferon- α* . Preliminary fitting of the parameters of the extended model shows great promise, enabling the model to successfully describe the development of the JAK2 allele-burden for a minor group of patients which previously could not be modelled. While the parameters are currently individually estimated by hand, future work formalising an algorithm for the determination of the parameters based on medical knowledge could prove useful for clinical assessments of the response to treatment for individual patients. As such, this could ultimately lead to personalised treatment schemes, based on the initial response to treatment.



Joint work with: Thomas Stiehl^b, Hans C. Hasselbalch^c,

Morten Andersen^a and Johnny T. Ottesen^a

^a: Roskilde University, Denmark. ^b: Heidelberg University, Germany. ^c: Zealand University Hospital, Denmark



Motivation and background

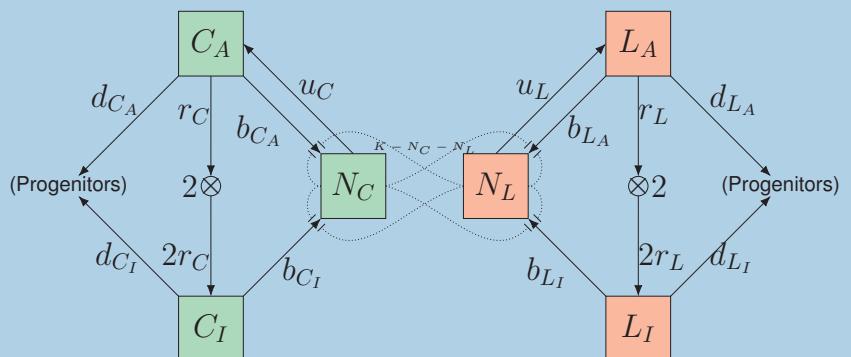
Andersen et al (2017) present a mechanism-based mathematical model describing the hematopoietic (blood-producing) system of the human body during the development of the Myeloproliferative Neoplasms (MPNs), a group of slowly developing blood cancers. The model consists of a six-dimensional system of ODEs. Through time-scale considerations, the model can be reduced to a simpler two-dimensional system describing the slow dynamics (Ottesen et al, 2019). So far, model analysis has granted some insight about the cytokines related to MPNs as well as prediction of patient-specific treatment response.

However, some changes to the model are currently being considered:

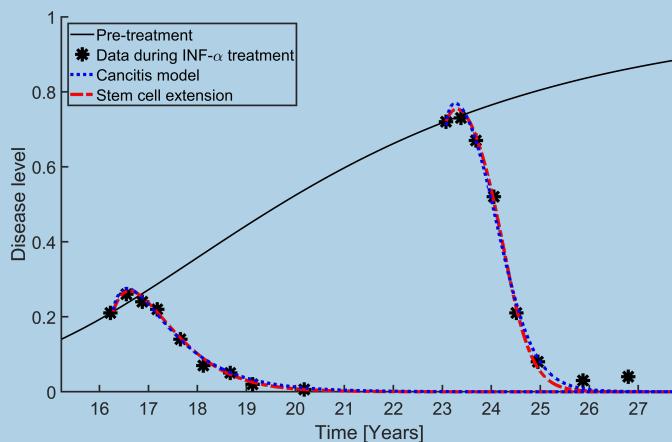
- A subset of patients display response that are not readily reproducible by the present model formulation.
- The treatment considered (INF- α) is known to affect features of the hematopoietic system not currently included in the model.

Mechanisms of stem cells

- The figure depicts a compartment diagram of a separate model considering solely the hematopoietic stem cells within the bone marrow micro-environment.
- Inspired by the work of Wang et al (2017), the model considers competition between stem cells in a finite stem cell niche of capacity K .
- Stem cells bound to the niche are considered quiescent while free stem cells can divide, differentiate or return to quiescence by rebinding with the niche.



Results



- Time-scale considerations allow for a reduction of the stem cell model to a system of two ODEs, modelling the slow time-scale.

$$\begin{aligned} \dot{N}_C &= \left(\frac{2(1 - N_C - N_L)}{\alpha_C + 1 - N_C - N_L} \left(\frac{r_C}{r_C + d_{CA}} \right) - 1 \right) N_C \\ \dot{N}_L &= \mu \left(\frac{2(1 - N_C - N_L)}{\alpha_L + 1 - N_C - N_L} \left(\frac{r_L}{r_L + d_{LA}} \right) - 1 \right) N_L \end{aligned}$$

Where $\alpha_C = d_{CI}/(b_{CI}K)$, $\alpha_L = d_{LI}/(b_{CI}K)$ and $\mu = u_L/u_C$

- Stem cell compartments in the Cancitis model can be replaced with the slow time-scale stem cell model leading to an extended model.
- The figure displays disease development, two examples of data for patients undergoing treatment as well as simulate treatment in the models.
- Both the normal and extended model can reproduce the responses to treatment, by step-wise changing parameters at treatment onset.

Preliminary conclusions

- When modelling treatment-response of MPN patients, the Cancitis model implies **increased elimination** of malignant cells as response to treatment.
- The model can be extended to include the **stem cell dynamics** more accurately, which allows for **alternative hypotheses about which parameters are affected by treatment**.
- Preliminary results suggests that **increased activation of quiescent stem cells** ($u_C \uparrow$ and $u_L \uparrow$) and increased differentiation ($d_{LI} \uparrow$) lead to the delayed responses seen in data.

References

- Andersen, M., Sajid, Z., Pedersen, R. K., Gudmand-Hoeyer, J., Ellervik, C., Skov, V., Kjaer, L., Pallisaard, N., Kruse, T. A., Thomassen, M., Troelsen, J., Hasselbalch, H. C., Ottesen, J. T. (2017). Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. PLOS ONE, 12(8), e0183620. <https://doi.org/10.1371/journal.pone.0183620>
- Ottesen, J. T., Pedersen, R. K., Sajid, Z., Gudmand-Hoeyer, J., Bangsgaard, K. O., Skov, V., Kjaer, L., Knudsen, T.A., Pallisaard, N., Hasselbalch, H. C., Andersen, M. (2019). Bridging blood cancers and inflammation: The reduced Cancitis model. Journal of Theoretical Biology, 465, 90-108. <https://doi.org/10.1016/j.jtbi.2019.01.001>
- Wang, W., Stiehl, T., Rafel, S., Hoang, V. T., Hoffmann, I., Poisa-Beiro, L., Saeed, B. R., Blume, R., Manta, L., Eckstein, V., Bochler, T., Wucher, P., Essers, M., Jauch, A., Trumpp, A., Marciniauk-Czochra, A., Ho, A. D., Lutz, C. (2017). Reduced hematopoietic stem cell frequency predicts outcome in acute myeloid leukemia. Haematologica, 102(9), 1567-1577. <https://doi.org/10.3324/haematol.2016.163584>