© 2020

Leonardo Miranda Santana

ALL RIGHTS RESERVED

MATHEMATICAL MODELING OF TUMOR PROGRESSION AND CHEMOTHERAPY REGIMENS

By

LEONARDO MIRANDA SANTANA

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

Written under the direction of

Gyan Bhanot

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2020

ABSTRACT OF THE DISSERTATION

Mathematical Modeling of Tumor Progression and Chemotherapy Regimens

By LEONARDO MIRANDA SANTANA

Dissertation Director:

Gyan Bhanot

Despite all important advances in the treatment of cancer over the last decades, preventing disease recurrence remains a challenge. Resistance of tumors to chemotherapy can be caused not only by selection of drug-resistant clones over the course of treatment, but also by the presence of tumor foci that are in a dormant state and are not targeted/killed by DNA-damaging agents. Such dormant tumor foci can eventually transition to a state of active growth, causing disease recurrence. In this thesis, we propose a stochastic model to describe recurrence of generic tumors, in which tumor foci can transition between a chemoresistant dormant state and a chemosensitive state of active growth. We develop a framework to determine the time-dependent probability that an undetectable residual tumor will become large enough to be detectable, and model the effect of chemotherapy on recurrence by switching the death rate of active tumor foci at the treatment time cap. We fit our model to data from a clinical trial for maintenance chemotherapy with the poly(ADP-ribose) polymerase (PARP) inhibitor olaparib in ovarian cancer, and use parameters from the fits to predict recurrence-free survival when chemotherapy dosage or duration are increased. In

this context, we also investigate how recurrence and cure are affected by transition rates between dormant and active states within the tumor, and predict how the effectiveness of increasing chemotherapy dosage or duration for improving long-term recurrence-free survival depends on these rates. Our results should be useful in planning optimized chemotherapy dosage and duration for cancer treatment, especially in cancer types for which no targeted therapy is available.

Acknowledgments

The work described in Chapters 2, 3 of this thesis is published in [1]: Leonardo M. Santana, Shridar Ganesan, and Gyan Bhanot, A quasi birth-and-death model for tumor recurrence, *J. Theor. Biol.*, 480: 175–191, Copyright Elsevier (2019). The work described in Chapters 4 and 5 is currently being prepared for publication.

I am immensely grateful to Gyan Bhanot for being an extremely kind and helpful advisor and for giving me the opportunity to be his Ph.D. student. Without his patience, wisdom and words of encouragement, I would have had a significantly harder time to complete my work while managing my teaching duties at the Physics and Math departments. I would like to thank him for doing such a great job in helping me transition to a research area that was completely new to me. As a scientist who has contributed to so many areas of knowledge, he has been an inspiration to me and I learned a great deal from him.

I would like to thank Dr. Shridar Ganesan, for pointing out the biological ideas that gave birth to this work, and for his input as a clinician and cancer researcher. I am also thankful to all the people in the computational cancer genomics group at CINJ, for such a great learning experience at the Systems Biology and Bioinformatics group meetings. In particular, Anshuman Panda and Amartya Singh gave me valuable advice that helped me in my transition to cancer computational biology.

In the Physics Department, I would like to thank Alexandre Morozov, for suggesting the continuum limit approach that helped me move forward with this project. I would also like to thank Emil Yuzbashyan, for his mentorship during my experience in the Condensed Matter Theory group.

I am thankful to Ronald Gilman, for being such a supportive Graduate Program Director, and to all the faculty in the Physics Department from whom I had the privilege to learn physics by attending their enlightening courses during my first years in the program.

I would like to thank Mark Croft and Abdelbaki Brahmia, for all their support during my time as a physics recitation/lab instructor, and for their organizational skills that made my life easier during my teaching experience throughout these years.

I am very grateful to my family, especially my parents, for all their emotional support, for investing in my education and for believing in my capabilities. Without them, my academic endeavors would never have been possible.

I am very thankful to my girlfriend Siana, for all her love and support, and for always being so sweet and caring. And a special thanks to Aniket and Adrian, for being great friends during all these years, and for helping me keep my sanity with their sense of humor and positive outlook on life.

Dedication

To my mother Vanessa (in memorian)

Table of Contents

Abstract ii

Acknowledgments iv

Dedication vi

1. Introduction 1

1.1. Overview 1

1.2. Tumor viruses and the discovery of oncogenes 3

1.3. Mitogenic signaling and cancer 4

1.4. Tumor suppressor genes 7

1.5. Cell cycle clock and G1 checkpoint control 9

1.6. The p53 protein 11

1.7. A few remarks on the influence of tumor cell cycling dynamics on response to chemotherapy 14

2. A Quasi-Birth-and-Death Model for Tumor Recurrence 18

2.1. Overview of the discrete-state model 19

2.2. Master equation for discrete-state transition probabilities 22

2.3. Continuum limit of the discrete-state QBD model 24

2.3.1. Forward Kolmogorov approach 25

2.3.2. Boundary conditions for the forward equation 27

2.3.3. Backward Kolmogorov approach 30

2.3.4. Boundary conditions for the main backward equation 31
2.3.5. Probability flux through the absorbing boundaries 32
2.3.6. Equations for time-dependent probabilities of cure and recurrence . 32
2.3.7. Recurrence time probability distribution and moments 34

3. Analytical solutions and simulations 36
3.1. Stationary solution at large times 36
3.2. Approximate solution of the mean recurrence time equation 39
3.3. Simulations 43
3.4. Moments of the tumor-size distribution 47
3.5. Fitting the model to recurrence-free survival data 50
3.6. Application of the model to chemotherapy simulation 54

4. Fitting the model to data from an ovarian cancer neo-adjuvant trial with Olaparib 59
4.1. Ovarian cancer clinical trial with Olaparib 59
4.2. Setting up the data from the clinical study 61
4.3. Fitting the model to the placebo KM curve 63
4.4. Fitting the model to the Olaparib KM curve 68

5. Predicting Recurrence-Free Survival as a function of Olaparib dosing and scheduling 73
5.1. Predictions for long-term recurrence-free survival of patients in the Olaparib clinical trial 73
5.2. Predicting the effect of dormancy on recurrence and cure rates of patients treated with Olaparib 79
5.3. Predicted tumor-growth dynamics for the Olaparib study and effects of dormancy on dynamics 84

6. Conclusions and outlook 89

Appendix A. Structure of the infinitesimal transition matrix 92

Appendix B. Solution of the M/M/1/N queue with absorbing boundary states 95

Appendix C. The Kaplan-Meier survivorship curve 99

Appendix D. Algorithm to reconstruct individual patient data from Kaplan-Meier curves 101

Chapter 1

Introduction

1.1 Overview

Complex multicellular organisms have evolved to control cell proliferation using sophisticated mechanisms that determine the fate of a cell, stasis, replication or death. These mechanisms are critical in maintaining order and function in complex body tissues. In cancer, some of these mechanisms fail because of abnormalities in genes that regulate them, causing the defective cells to proliferate uncontrollably, sometimes leading to dysregulated growth, which is a hallmark of cancer [2, 3].

Cell replication or death are controlled by two types of genes, oncogenes and tumor suppressor genes. Oncogenes are responsible for promoting cell proliferation, regulating blood supply, building tissue etc, whereas tumor suppressor genes act as a brake on these processes, regulating them to ensure cell and tissue homeostasis by controlling a variety of cellular signaling pathways. It is the equilibrium between these two forces, as they respond to both internal and external stimuli, that keeps cell proliferation under control, and preserves the integrity of normal cellular functions encoded in the genome. Cancer is essentially the result of the disruption of this equilibrium caused by genetic abnormalities.

In addition to internal cell-cycle control, there are a number of pathways that guard against internal or external effects that damage the DNA. These mechanisms that repair or protect the DNA are vital to preserve genomic integrity and maintain normal cellular function. Some examples of these mechanisms, established by billions of years of evolution, are cell-cycle arrest, the DNA proofreading and repair machinery, apoptosis (programmed

cell death) and senescence. These mechanisms destroy or inactivate cells that have accumulated DNA damage beyond the possibility of repair. The genes that control DNA repair sometimes lose their function due to mutations or epigenetic silencing, which leads to genomic instability and accumulation of DNA damage [4]. Tumors originate when cells with such abnormalities are permitted to survive or replicate, as a result of mutations in apoptosis pathways, DNA-damage checkpoints, DNA-repair checkpoints or mitotic checkpoint pathways [5, 6].

Carcinogenesis is a multistep process, typically requiring the acquisition of several mutations for cancer initiation, growth and metastasis [7]. There are several factors that can cause or increase the risk of developing cancer, the most common being exposure to chemicals or radiation that cause DNA damage, inherited genetic defects, and infection with certain viruses. Regardless of the primary cause of the cancer, all tumors acquire certain biological capabilities as they develop, which allow them to sustain their growth and proliferate. Hanahan and Weinberg have proposed at least six such capabilities that a tumor must acquire: sustained proliferative signaling, evasion of growth suppressors, resistance to programmed cell death, replicative immortality, induction of angiogenesis (the process by which new blood vessels are formed), and activation of invasion and metastasis pathways [3]. These evasion mechanisms in tumors are often the result of increased mutation rates, resulting in chromosomal abnormalities, which create an altered karyotype, leading to altereted pathways and tumor heterogeneity, thereby creating multiple ways in which tumors can survive immune surveillance and therapy.

A large number of genes associated with cancer development and progression have been identified, and their roles in specific pathways have been established [8]. The enormous variety of molecular types of cancer results from the large number of possible combinations of defective oncogenes and tumor suppressor genes that contribute to the neoplastic state [9]. A key tumor suppressor gene, which is the most frequently mutated (and most frequently cited) gene in human cancers, is the transcription factor p53, also known as the “Guardian

of the Genome" [10–12]. As the name suggests, it plays an important role in preventing cancer, by identifying damaged cells, pausing progression of the cell cycle from G1 phase into S phase or from G2 phase into M phase, giving the cell time to effect appropriate repair. When repair fails, p53 induces apoptosis or regulated death of the cell, and therefore acts as a tumor suppressor [13]. Most cancers find it necessary to inactivate p53 using a variety of strategies, and to eliminate the barrier imposed on cell replication by its surveillance of the cellular state, which then allows the tumor to proliferate and invade distant tissues [14].

In the next sections, a few major mechanisms in the development of cancer are reviewed, especially those pertaining to mitogenic signaling, cell-cycle control and apoptosis, and the roles of key genes are described. A brief summary of important discoveries relevant to the molecular biology of cancer is given along with appropriate references, which may serve as an introductory guide to the subject.

1.2 Tumor viruses and the discovery of oncogenes

The discovery that normal chicken embryo fibroblasts can be transformed into tumor cells by infection with the Rous sarcoma virus (RSV) paved the way for major discoveries in cancer biology. It lead to the realization that certain RNA viruses, known as retroviruses, make DNA strands from their own RNA templates (reverse transcription) and insert them into the genome of the host cell during their replication cycle. The answer to the puzzle of how RSV could transform a normal cell into a tumor cell was the presence of a gene in the viral genome (the v-*src* oncogene) that caused the transformation [15]. The sequence of the v-*src* gene was later discovered to be closely related to the DNA sequence of a gene present in normal avian DNA, c-*src*, making it the first oncogene to be identified [16].

Tumor retroviruses can sometimes transform cells even in the absence of an oncogene of cellular origin in its genome. However, cells infected with such viruses take a significantly longer time to develop a tumor than those infected with oncogene-carrying viruses such as RSV. An investigation of lymphomas induced by the avian leukosis virus (ALV) revealed

that the cellular transformation could be explained by insertion of the retroviral genome adjacent to the c-*myc* proto-oncogene, causing it to be activated through the viral promoter [17]. Although integration of the ALV retroviral genome occurs at random locations of the host cell's genome, on rare occasions, this insertion happens to be adjacent to the c-*myc* gene. Since these early studies, both types of tumor retroviruses (carrying or not carrying an oncogene in their genome) have allowed researchers to identify many oncogenes [18].

The key discovery that oncogenes carried by transforming retroviruses can also be found in cell lines transformed by DNA transfection provided further evidence to the idea that the same oncogenes can be activated by different mechanisms, such as retroviruses and the action of mutagens. Moreover, a number of experiments have revealed that oncogene activation can occur in a variety of ways, such as structural changes in the protein due to mutations, overexpression of the oncogene due to copy-number alterations or chromosomal translocations, or changes in related regulatory pathways (e.g. suppression of tumor suppressor genes that control the oncogene) [2].

1.3 Mitogenic signaling and cancer

The maintenance of tissue architecture in Metazoa relies on coordinated cellular signaling, through which cells can send mitogenic stimuli to other cells, thus promoting growth in a tightly controlled manner. Such signaling is usually achieved through the activation of cell-membrane receptors by ligands known as growth factors. Such receptors, once activated, emit signals to downstream proteins of the mitotic signaling circuitry. These signaling mechanisms are highly conserved across the metazoan genomes, and their study has allowed scientists to unveil the biochemical machinery underlying these processes in humans via the study of model organisms such as flies, worms and mice [2].

Many oncogenes found to be drivers of human cancers have been shown to induce mitotic activity by modification of growth signaling. The non-receptor tyrosine kinase *src*, which was the first proto-oncogene to be discovered in vertebrates (following the discovery of its

retroviral counterpart, v-*src*, the first viral oncogene to be discovered), provided hints to how a single oncogene can transform a cell and create a variety of cancer phenotypes. A number of experiments revealed that *src* acts as a kinase that phosphorylates tyrosine amino-acid residues in multiple downstream proteins involved in mitotic pathways, thus having a key role in mitogenic signaling [15].

Other proteins were discovered to function as signal transducers located at the cell membrane, receiving growth signals from outside the cell and communicating them to other proteins in the cytoplasm or directly relocating to the nucleus to activate specific pathways. The first such protein to be discovered was EGF-R, the receptor protein for the Epidermal Growth Factor (EGF) protein [19]. It was found that EGF-R receptors contain an extracellular domain that binds to its ligand EGF, and a cytoplasmic domain that is similar to the *src* gene product, which acts as a tyrosine kinase (for this reason, EGF-R is part of the so-called RTK class of proteins, standing for tyrosine kinase receptor).

The EGF receptors at a cell membrane are an interface through which the cell senses mitogenic stimuli, representing the input layer of the EGF network (or the *Her* network in humans). The signals received from growth factors are processed by complex signaling cascades that communicate with the cellular nucleus, where proteins at the output layer of mitogenic cascades function as transcription factors, which bind to DNA promoter sites of key genes involved in such processes as apoptosis, migration, growth, adhesion and differentiation [2].

Genomic sequence analysis of EGF-R revealed its similarity to v-*erb-B*, an oncogene originally discovered in the genome of the avian erythroblastosis virus (AEV), a retrovirus known for its ability to transform normal cells into tumor cells [20]. More specifically, it was discovered that the viral gene v-*erb-B* clearly inherited (over the course of its evolution) the genomic sequence of an EGF-R lacking the EGF-binding extracellular domain. The resulting protein product of the viral gene acts as malfunctioning receptor that is *constitutively* activated, transducing continuous signals telling the cell to proliferate even without

external stimulus.

Many cancer types are associated with overexpression or mutations of EGF-R or other proteins of the cellular *erb-B* family (also known as Her in humans). One such example is the *Her2/Neu* protein in *Her2+* breast cancer. The biochemical mechanism that favors malignancies in mutant or overexpressed *erb-B* receptors can be understood by a dimerization model [21]. Each receptor is composed of two protein subunits, each of which is free to move laterally on the two-dimensional surface of the cell membrane. When a growth factor binds to one of the subunits, the latter moves randomly over the membrane until it finds its complementary subunit, thus activating the receptor by forming a dimer. When receptor proteins are overexpressed, dimer formation becomes more likely. The dimer model also explains how mutations in the *erb-B* family of genes can cause receptors to fire even in the absence of growth-factor ligands, since dimer formation (and thus activation of cytoplasmic signaling) is favored by certain mutations of the subunit proteins. The discovery of *erb-B* receptors and their role in cancer formation has lead to the development of targeted long-term therapies (e.g. Herceptin for *Her2+* breast cancer) that suppress receptor signaling either by blocking their extracellular growth-factor binding domain (i.e., by acting as growth-factor antagonists), or by inhibiting kinase activity of their cytoplasmic RTK domain.

After the discovery of EGF, which mainly acts on certain epithelial cells, several other growth factors that act on a variety of tissues were discovered, and their roles in the pathogenesis of many cancers were characterized [22]. Likewise, a variety of transmembrane surface receptors other than RTKs were discovered to be associated with human cancers such as, for example, integrins [23], which function as sensors that cells use to control their adherence to the extracellular matrix (this is particularly important in the process of tumor metastasis).

Increased mitogenic signaling in cancers can also be caused by the response of a cell to its own overexpressed growth factors. This occurs, for example, in cells transformed by

the Simian sarcoma virus. An oncogene present in this retrovirus, v-*sis*, was discovered to be sequentially related to the Platelet-Derived Growth Factor (PDGF) found in serum. Infected cells express the v-*sis* gene at high levels, increasing firing of the cell's own PDGF receptors to induce mitogenic activity [24, 25].

A key family of proteins that lie downstream of RTK receptors in mitogenic signaling networks is *Ras*, originally discovered as an oncogene of the rat sarcoma viruses. These proteins are GTPases, acting as molecular switches that have a central role in mediating signal transduction from the cell membrane to several downstream signaling cascades ending in the nucleus [26]. When a *Ras* protein is bound to guanosine diphosphate (GDP) it is "off", meaning that it has low affinity to its downstream effectors, whereas it is turned "on" when the bound GDP gets replaced by a guanosine triphosphate (GTP), meaning that it binds to its downstream effectors (mainly Raf, Ral-GEF and PI3K) with high affinity. Mediating between RTK receptors and *Ras* is a guanine nucleotide exchange factor known as *Sos*, which binds to adapter proteins (*Shc*, *Grb2*) that attach to transphosphorylated (active) RTK cytoplasmic tails and induces the release of GDP bound to *Ras*, allowing binding of GTP. In other words, *Sos* has the ability to turn the *Ras* switch "on" when an RTK receptor at the cell membrane fires. An activated wild-type *Ras* protein shuts itself off through its own intrinsic GTPase activity, hydrolizing GTP into GDP, especially in the presence of GTPase-activating proteins (GAPs), thus forming an activation/deactivation cycle. This feature of the *Ras* protein acts as a negative feedback loop that tightly controls signal transduction to downstream cascades. Mutated *Ras* proteins can become oncogenic when they lose their ability to shut off due to aminoacid substitutions in their GTPase domain.

1.4 Tumor suppressor genes

An important group of genes in cancer formation are the tumor suppressor genes (TSGs), which act either as "gatekeepers", preventing progression into the cell cycle or inducing

apoptosis, or as "caretakers", maintaining genomic integrity [2]. Loss of functional TSGs removes barriers preventing cell proliferation and survival, which can lead to malignant transformation.

In 1971, Alfred G. Knudson proposed that the loss of function of a single TSG observed frequently in retinoblastoma patients (the *Rb1* gene) only occurs when *both* alleles of the gene are inactivated [27]. This lead to the realization that inactive TSGs are generally recessive, an idea that came to be known as the "two-hit hypothesis". If one of the TSG alleles is inactive due to germ line or somatic mutations, the second allele can be inactivated if it is hit by mutations or epigenetic silencing due to methylation of the gene promoter, or if it is lost due to a *loss of heterozygozity* (LOH) event. Exceptions to this genetic mechanism have been observed in TSGs that are haploinsufficient, i.e., when the functional loss of only one allele is advantageous for the development of the tumor [28] (a notable example is the the TP53 gene encoding the p53 protein, although tumors with one mutated TP53 allele usually end up deleting the wild-type copy to completely eliminate any residual tumor-suppressive function). The recessive nature of most TSGs helps their identification by the frequent LOH occurrences found in tumors (especially in familial cancers). A common scenario is a mutated allele of a TSG and the loss of the second allele due to a deletion in the homologous chromosome. Genomic analysis of tumors can identify candidate TSGs, which can then be "knocked out" in the germ line of mouse models to validate them as TSGs.

Because of its recessive nature, a TSG allele with an inactivating mutation can be inherited in the germ line, unlike mutant oncogenes, which are usually incompatible with normal embryonic development. This means that familial cancers are usually linked to the presence of a defective TSG allele, which increases the likelihood of cancer development due to LOH or other mechanisms that inactivate the wild-type TSG allele.

Besides coding for proteins involved in key cellular processes such as cell-cycle control, detection/repair of DNA damage, negative regulation of mitogenic signal transduction, cell

differentiation and migration, among others, some TSGs have also been found that produce micro RNAs (miRNAs) instead of coding for proteins. Such miRNAs are responsible for post-transcriptional regulation of many genes, with implications for cancer development and progression [29].

The next two sections will be mainly focused on the roles of the two most important tumor suppressor proteins in cancer: the pRb protein, which plays a central role in cell-cycle control through repression of DNA replication, and the p53 protein, a transcription factor that protects the genome by triggering DNA repair, cell-cycle arrest, apoptosis or senescence.

1.5 Cell cycle clock and G1 checkpoint control

All metazoans share essentially the same regulatory machinery that controls their cell cycle, which likely appeared in a common ancestor hundreds of millions of years ago. The cell cycle is an organized and regulated process, which proceeds through a sequence of phases, during which the cell prepares proteins necessary for the complex process of cell division, makes critical decisions to proceed to the next phase, pause to repair damaged DNA, or undergo apoptosis. During the first or G1 phase, the cell grows and prepares proteins for the S ("synthesis") phase, during which a major step in the process of division takes place, viz, the replication of DNA. The decision of whether a cell enters the S phase is controlled by the G1 checkpoint (restriction point), before the end of the G1 phase. This decision depends on both external and internal factors, such as mitogenic signaling and DNA integrity, respectively. After DNA replication (S phase), the cell enters another phase of growth known as G2, during which it encounters a second checkpoint (G2), at which the cell cycle program checks the integrity of the replicated DNA, before it decides whether to enter the M phase (mitosis) or to halt the process and repair damaged DNA. Finally, during mitosis (M phase), the cell undergoes a series of events that allow it to divide into two identical daughter cells. During the so-called metaphase of mitosis, a third checkpoint

(spindle assembly checkpoint) restrains the separation of sister chromatids until they are properly attached to the spindle apparatus, to ensure that each daughter cell gets exactly one copy of each chromosome. In addition to the G1, S, G2 and M phases, cells can also enter a quiescent (i.e., dormant) state known as G0 (classic examples are nerve and heart cells, which are permanently arrested in G0 after differentiation), or exit it and reenter G1 when sufficiently stimulated by extracellular mitogenic signals (in tumor cells, this no longer required due to self-sustained mitogenic signaling and absence of barriers against proliferation, so that G0/G1 transitions become dysregulated).

Mitogenic signals transduced from cell surface receptors to the nucleus through signaling cascades (see Section 1.3) end up with activation of transcription factors that promote transcription of key proteins of the cell cycle circuit known as D cyclins, which are part of a family of four different types of proteins (D, E, A and B cyclins) present in the checkpoint control circuit, whose function is to activate partner proteins known as cyclin-dependent kinases (CDKs). Yet another set of proteins negatively regulate CDKs, as expected for a cyclic process. Essentially, progression through the cell cycle and the correct sequence of events are determined by changes in the levels of D, E, A and B cyclins during the different phases of the cycle [2, 30].

The most critical decision during the cell cycle occurs at the G1 checkpoint, where the cell cycle clock checks if all prerequisites are met before allowing the cell to replicate its DNA [6, 31]. At the core of the G1 checkpoint control circuit is the pRb protein, encoded in the *Rb* tumor suppressor gene [11]. Essentially, this protein represses DNA replication, unless it is hyperphosphorylated.

While it is unphosphorylated or hypophosphorylated, the pRb protein binds to a transcription factor of the E2F family, preventing it from promoting expression of genes involved in DNA replication. The production of D-type cyclins in response to extracellular signals (which become autocrine in tumor cells) activates their partner CDKs (either CDK4 or CDK6) by forming D-CDK4/6 complexes, which in turn phosphorylate pRb proteins. This

phosphorylation inactivates the repressor state of pRb, which releases the previously bound E2F transcription factors, thus activating transcription of late G1 genes. This is the mechanism by which the G1 checkpoint gate is opened, allowing the cell to enter the S phase [11]. Once the cell proceeds beyond the G1 checkpoint, it follows an automatic program regulated by other cyclins (E, A and B) and CDKs (CDK2 and CDC2), which unlike the D-cyclin program, does not depend on extracellular signals [2]. Among the genes whose transcription is promoted by active E2Fs are the gene encoding cyclin E and also the E2F1 gene (which is one of the transcription factors of the E2F family itself). Cyclin E binds to CDK2, forming an E-CDK2 complex that drives hyperphosphorylation of pRb, further increasing the levels of cyclin E and activated E2F1 and creating a feed-forward loop that ensures a rapid and irreversible transition into the S phase.

Mutations in the *Rb* gene are very common in cancer, and the role of inactivated pRb proteins in the development and survival of tumors is quite obvious: a defective pRb protein never binds to the transcription factor E2F regardless of its phosphorylation state, thus removing the barrier that prevents the cell from proceeding to the S phase. Loss of function of the pRb protein can also occur by the presence of viral oncoproteins that block it from binding to E2F, by the action of oncogenes that lead to increased production of D cyclins, or by oncogenes that deregulate pRb phosphorylation [2].

1.6 The p53 protein

A hallmark of tumor cells is their ability to evade both apoptosis (programmed cell death) and permanent exit from the cell cycle, which occur in normal cells as they respond to extreme internal or external stresses that jeopardize normal cellular function [3]. In order to preserve the integrity of the tissues where they belong, the cells of complex organisms were designed by evolution to halt their progression in the cell cycle to allow repair of damaged DNA, to permanently stop proliferating or to undergo apoptosis when they face physiologic stress [32]. Tumor survival depends on bypassing the mechanisms that control

these processes, which depend on the orchestrated action of a complex regulatory network within the cell. The p53 protein is the key gatekeeper at the center of this molecular network and, not surprisingly, it is a major target of tumorigenesis [33].

As was the case of many important breakthroughs in the molecular biology of cancer, p53 was originally discovered in connection with tumor viruses. The protein was first detected in the 1970s using cell lines that had been transformed by the SV40 tumor virus and subsequently injected into mice. The immune response against the transformed cells produced antibodies that reacted with a viral protein ("large T") capable of binding to (and by doing so, inactivating) a cellular protein that came to be identified as the tumor suppressor protein p53 [34, 35].

The p53 protein is a transcription factor that receives signals from a variety of sensors and surveillance systems in the cell, which detect physiologic stresses such as DNA damage (caused by radiation or chemicals), hypoxia, dysregulated growth signaling pathways, blockage of DNA transcription or replication, etc. Such signals trigger the p53 alarm by increasing the levels of the p53 protein, which then takes action by activating transcription of a variety of genes involved in cell-cycle arrest, DNA repair, senescence or apoptosis [2].

Under healthy conditions, a normal cell keeps its levels of p53 very low, by requiring a high turnover rate of the protein. In some sense, p53 regulation is done by having it produced all the time and degraded quickly, like pressing both the accelerator and the brake of a car. This allows rapid action when necessary. If the break is released, the car immediately accelerates, and if the accelerator is released, the car immediately stops. In the case of p53, when the surveillance systems of the cell detect signs of physiologic stress, the levels of p53 rise rapidly to effect a response. This rapid increase in p53 levels is made possible by regulatory proteins such as Mdm2 and ARF [36]. The Mdm2 protein, which is actually a product of one of the genes activated by p53, works as a negative regulator by binding to the p53 tetramer and ubiquitinating it, thus tagging it for degradation by proteasomes. The ARF protein upregulates p53 by preventing Mdm2 from marking it for

degradation. Critically high levels of E2Fs caused by deregulation of the G1 checkpoint control activate ARF, leading to an increase in p53 levels, which triggers tumor-suppressive response.

Cell-cycle arrest can be triggered by p53 through the activation of a number of proteins, the most important of which is the CDK inhibitor $p21^{Cip1}$. This protein plays an important role in halting cell cycle progression by inhibiting CDK2 or other cyclins, thereby inhibiting phosphorilation of pRb (and consequently the expression of genes regulated by E2F), firing of replication origins and DNA-synthesis proteins [37]. At the same time, $p21^{Cip1}$ protects cells from apoptosis and halts DNA replication by binding to PCNA (proliferating cell nuclear antigen), thus allowing DNA repair.

Apoptosis is triggered by p53 when upstream surveillance proteins detect extensive DNA damage that cannot be repaired or substantially high mitogenic signaling within the cell. There are two distinct mechanisms to induce apoptosis: the intrinsic (mitochondria dependent) and the extrinsic (death receptor dependent) apoptotic programs, both of which eventually converge to the same downstream executioner caspases [38]. The intrinsic pathway starts with p53 activating proteins such as Bax, which induce mitochondrial outer membrane permeabilization (MOMP), releasing pro-apoptotic proteins such as Cytochrome c from the mitochondrial intermembrane space into the cytosol. This initiates an apoptotic caspase cascade, ending with the activation of executioner caspases, which cleave a range of substrates, such as downstream caspases, nuclear proteins, plasma membrane proteins and mitochondrial proteins, ultimately leading to cell death. Although p53 primarily triggers apoptosis by the intrinsic program, it also regulates expression of certain death receptors (such as Fas) that initiate the extrinsic program when their cognate ligands FasL bind to them. These activated death receptors then bind to/activate the Fas-associated proteins with death domain (FADD), forming the death-inducing signaling complex (DISC), which in turn initiates a caspase cascade ending up at the same executioner caspases as those in the intrinsic program.

Because p53 is a highly connected node in a vast network of proteins, and because of the critical role of this protein for such a variety of cellular responses against tumorigenesis, the inactivation of p53 is often a priority in early evolving tumors, which take advantage of this weakness to grow and survive. The loss of p53 function is advantageous at multiple levels, allowing tumors to evade cell-cycle arrest and apoptosis in spite of dysregulated mitogenic signaling, while also acquiring an increased tolerance to mutations, thereby activating new oncogenes and inactivating new tumor suppressor genes throughout the process [2, 3, 33].

1.7 A few remarks on the influence of tumor cell cycling dynamics on response to chemotherapy

The advent of the first chemotherapeutic drugs for leukemia and lymphoma in the 1940s was an important milestone in the history of cancer treatment and research [39]. Since then, a large number of cytotoxic drugs with anti-cancer properties have been discovered, and targeted therapies developed for specific cancers, such as hormonal therapy with tamoxifen for ER+ breast cancer. For cancers such as serous ovarian cancer and triple negative (ER-/HER2-/PR-) breast cancer, where such adjuvant (post-surgery) targeted therapies are not available, chemotherapy with DNA-damaging agents remains the sole option after surgery and radiation [40]. By targeting rapidly cycling cells, which are often already at the limit of their stability and unable to repair additional DNA damage, chemotherapy with cytotoxic drugs systemically attacks growing tumors. Side effects on other tissues, especially on cells with a high turnover rate, such as skin and the intestinal epithelium, can be moderate to severe, depending on the duration and intensity of treatment.

A large number of mathematical models have been proposed to predict optimal regimens of adjuvant chemotherapy to specify duration, dosage levels or dosing protocols, with the goal of reducing tumor recurrence or limiting metastasis. These models may be classified by their mathematical approach, which can be either stochastic or deterministic and linear or nonlinear, and to the nature of the biological assumptions underlying them (see

[41] for a review). Examples of such models include optimal-control-theory models [42], game-theoretical models [43, 44], as well as models of drug resistance and/or chemotherapy scheduling, which can be either stochastic [45–49] or deterministic [50–53]. Many of these models focus on optimization of chemotherapy dosing strategies, especially when tumors are treated with a combination of drugs, and the multiple types of clones in the tumor are resistant to different drugs in the combination.

In current clinical practice, chemotherapy is usually given at maximum tolerated dose for the minimum possible duration, which is usually 3-6 months, with the assumption that this strategy maximizes the benefit to the patients while decreasing the duration of possible side effects [54]. This strategy is based on modeling tumor cells as continuously dividing at some fixed deterministic rate [55]. Norton and Simon [56] proposed that tumor growth follows a type of sigmoid curve known in the literature as the Gompertzian function, and also proposed the tumor-regression hypothesis, that cell kill is proportional to the growth rate of the untreated tumor [57, 58]. Since the Gompertzian growth rate decreases as the tumor grows, they concluded that it becomes increasingly more difficult to kill the tumor as its size increases. This model provides the rationale for administrating the "maximum tolerated dose" and maximizing dose density, the goal being to efficiently kill the tumor when it is small and growing rapidly. However, the effectiveness of this treatment paradigm is unclear and not entirely consistent with some clinical and experimental data [59, 60] (see also Ref. [58] for a detailed literature review).

It is known that the *stage* of a cancer, which is related to the tumor size and degree of lymph node involvement, is an excellent predictor of prognosis, independently of cancer type or therapy. The larger the tumor, the more difficult it is to effect a cure. Likewise, tumor *grade*, which is a measure of tumor aggressiveness, is also a good predictor of outcome. However, it has been suggested that the fractional impact of treatment on improved survival is higher for patients with late-stage or high-grade tumors than for patients with early-stage or low-grade tumors (see, for example, the studies [61] and [62]). This observation suggests

that unlike Norton and Simon proposed in their deterministic Gompertzian growth model, tumor cells may not all be in continuous growth, and that actively growing tumors tend to be more responsive to cytotoxic drugs than those that are mostly in a dormant or resting state [63].

In [64], a stochastic alternative to Norton and Simon's deterministic Gompertzian model is proposed for breast cancer, where tumors are not in continuous growth, but can be either in a dormant state or in an active-growth state. It has been suggested that a major cause of resistance to chemotherapy is the presence of large numbers of dormant tumor foci in G0 phase, where they can repair damage induced by chemotherapy, and can then transition in a stochastic manner to a state of active growth [60, 64, 65]. In the presence of such dormant tumor foci, it is clear that the effect of chemotherapy on recurrence rates does not last for a long time after chemotherapy ends. Several clinical trials [66–68] have shown that improved recurrence rates for patients receiving chemotherapy revert to rates for the control group (who received no chemotherapy) in a relatively short time after termination of treatment, suggesting the presence of residual dormant tumor foci that survived treatment and regenerated the cancer. These data also suggest that short-duration chemotherapy only targets the tumor foci that are actively cycling during chemotherapy. Tumor foci that are mostly dormant during chemotherapy evade treatment and can cause recurrence when they transition to a state of active growth. This suggests that chemotherapy may be more effective if its duration is optimized according to the rate at which dormant tumor foci transition to active growth.

These observations suggest the following two hypotheses: 1) For a given cancer type, there is a characteristic time for dormant tumor foci to transition to active growth; 2) Dormant tumor foci are often resistant to chemotherapeutic agents that induce DNA damage, while active tumor foci are relatively more sensitive. Based on these hypotheses, we develop a mathematical model and framework to study the impact of variation in dosage and

duration of chemotherapy on recurrence rates, and the influence of transitions between dormant and active-growth states on chemotherapy effectiveness (Chapters 2 and 3). We apply our model to data from a phase 3 ovarian cancer clinical trial with the poly(ADP-ribose) polymerase (PARP) inhibitor olaparib, and make predictions for long-term progression-free survival of the patients in different scenarios (Chapters 4 and 5). If data is available, our model can be adapted to optimize dosage and duration in tumors other than serous ovarian cancer, in the hope that it may serve as a guide for the design of experiments and clinical trials that may eventually lead to better, more optimized chemotherapy regimens.

Chapter 2

A Quasi-Birth-and-Death Model for Tumor Recurrence

The tumor model described in this chapter was inspired by previous work on the effect of quiescence (i.e., the presence of dormant tumor foci) on tumor growth, such as the work of Komarova and Wodarz [46], which inspired the stochastic model described below, or to earlier deterministic versions of their model, proposed in Refs. [50], [51], or [52]. However, in contrast to these earlier studies, our focus is on predicting recurrence-free survival for a given treatment regimen, especially the long-term recurrence-free fraction and the time to recurrence. To this end, we propose special boundary conditions that will be described in Sections 2.2 and 2.3, which to the best of my knowledge, are a novel contribution to tumor modeling. Such boundary conditions are the key idea to make the connection between the model and clinical data for disease recurrence of generic tumors. For simplicity and as a first exercise, we will explore treatment regimens controlled by a single parameter μ_A, the death rate of actively-dividing tumor foci. Later on, we will add new parameters representing chemotherapy duration and dosing, which will be important in the application of the model to clinical data (Chapters 4 and 5). Also, unlike previous work, we will only model transitory chemoresistance resulting from dormancy, without discussing the important but more difficult stochastic issue of resistance from acquired mutations. The work described in this chapter and in Chapter 3 is published in [1]: Leonardo M. Santana, Shridar Ganesan, and Gyan Bhanot, A quasi birth-and-death model for tumor recurrence, *J. Theor. Biol.*, 480: 175–191, Copyright Elsevier (2019).

2.1 Overview of the discrete-state model

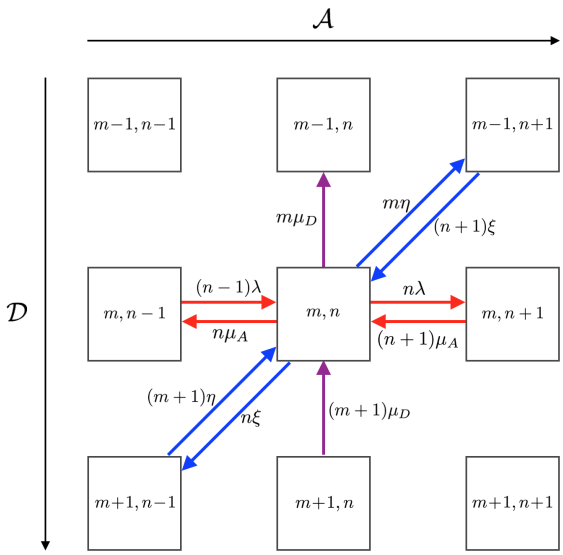
After surgery and radiation therapy, we assume that cancer patients retain a number of residual, undetectable tumor foci that may eventually grow and create a detectable tumor (recurrence). The foci can transition from a dormant (non-dividing, chemoresistant) state to an active (dividing, chemosensitive) state and vice versa with rates η and ξ respectively. Chemotherapy affects only active foci, which can either double or die, with rates λ and μ_A respectively. The dynamics of this process is modeled as a Quasi Birth-and-Death (QBD) process (for an introduction to this topic, see [69, 70]) that describes the stochastic time evolution of the number of active and dormant tumor foci in a patient, resulting in either tumor recurrence (when the number of foci is large enough to be detected) or remission (when there are no foci left).

To describe our model, it is simplest to use standard notations borrowed from Quantum Mechanics. Denoting by $|m, n\rangle$ the state with m dormant ($\mathcal{D}$) tumor foci and n active ($\mathcal{A}$) foci, the goal is to predict the time evolution of the joint probability distribution $p_{mn}(t)$, given the initial condition $p_{mn}(0)$. Without loss of generality, we can choose $p_{mn}(0) = \delta_{m,m_0}\delta_{n,n_0}$, where m_0 and n_0 are respectively initial counts of dormant and active foci. This is so because the latter initial condition defines a Green's function, from which the solution for any arbitrary initial condition can be constructed as a convolution of transition probabilities, as a result of linearity.

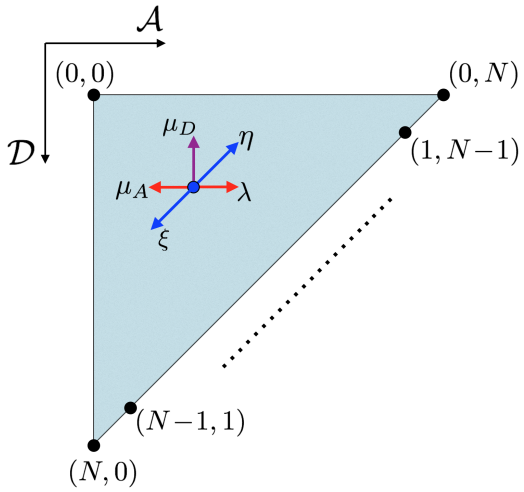
Our QBD model for tumor recurrence is a Markov process on the two-dimensional lattice of all possible Fock states $|m, n\rangle$ with the following transition rules:

- At any given time, an active ($\mathcal{A}$) tumor focus may either double or die at rates λ and μ_A, respectively. During chemotherapy (between time $t = 0$ and time $t = t_{chemo} > 0$), the death rate is $\mu_A = \mu_{chemo}$ and after treatment (beyond time $t = t_{chemo}$), it decreases to a lower baseline rate $\mu_A = \mu_0$.

- By definition, a dormant ($\mathcal{D}$) tumor focus cannot double, but it could in principle



(a) Possible transitions on the lattice of states (m, n), where m is the number of dormant ($\mathcal{D}$) tumor foci and n is the number of active ($\mathcal{A}$) foci.



(b) Boundaries of the state space $\mathcal{D} \otimes \mathcal{A}$. The boundary conditions are absorbing at the cure state $(0, 0)$ and also at the recurrence line $m + n = N$.

Figure 2.1: Structure of the Fock-like state space of the QBD model for tumor recurrence. Figures reproduced from [1].

die at some rate μ_D. For simplicity, we will eventually set $\mu_D = 0$. This is because dormant foci can repair chemotherapy damage as we argued above, i.e., they are chemoresistant.

- An active ($\mathcal{A}$) tumor focus may transition to dormancy ($\mathcal{D}$) at rate ξ and a dormant tumor focus may become active at rate η.

- Let $\bar{T}_D$ and $\bar{T}_A$ be the respective times that a tumor focus spends, on average, in the dormant and active states within its life cycle (defined in our context as the time between the birth of the tumor focus and its doubling). It then follows that the $\mathcal{D} \to \mathcal{A}$ and $\mathcal{A} \to \mathcal{D}$ hopping rates are given by $\eta = 1/\bar{T}_D$ and $\xi = 1/\bar{T}_A$, respectively, so the doubling rate is given by $\lambda = 1/(\bar{T}_A + \bar{T}_D)$. Hence, the rates η, ξ and λ are related by the equation

$$\frac{1}{\lambda} = \frac{1}{\eta} + \frac{1}{\xi}. \tag{2.1}$$

This constraint reduces the number of parameters in the model by one and allows us

to parametrize the rates η, ξ by their ratio $\nu \equiv \xi/\eta$ as follows:

$$\xi = \lambda(1+\nu)\ , \quad \eta = \lambda(1+1/\nu) \qquad (\nu > 0). \tag{2.2}$$

- All the tumor foci within the dormant ($\mathcal{D}$) or the active ($\mathcal{A}$) sub-population are equally likely to undergo a transition independently. This means that each transition probability on the two-dimensional lattice of states is proportional to either the population of dormant foci (m) or the population of active foci (n). The transition probabilities of any transitions to states beyond nearest-neighbor states are assumed to be second order in infinitesimal time, so the model is "skip free". The transition probabilities to neighboring states are then given by

$$\begin{aligned}
\Pr\{m-1, n; t+\Delta t | m, n; t\} &= \mu_D m \Delta t + \mathcal{O}(\Delta t^2) \\
\Pr\{m, n-1; t+\Delta t | m, n; t\} &= \mu_A n \Delta t + \mathcal{O}(\Delta t^2) \\
\Pr\{m, n+1; t+\Delta t | m, n; t\} &= \lambda n \Delta t + \mathcal{O}(\Delta t^2) \\
\Pr\{m-1, n+1; t+\Delta t | m, n; t\} &= \eta m \Delta t + \mathcal{O}(\Delta t^2) \\
\Pr\{m+1, n-1; t+\Delta t | m, n; t\} &= \xi n \Delta t + \mathcal{O}(\Delta t^2)
\end{aligned} \tag{2.3}$$

For transitions to states beyond nearest-neighbors, i.e., for $|m-m'| > 1$ or $|n-n'| > 1$, we assume

$$\Pr\{m', n'; t+\Delta t | m, n; t\} = \mathcal{O}(\Delta t^2). \tag{2.4}$$

A diagram of the state space showing the transitions above is given in Fig. 2.1.

- If the total number of tumor foci $m+n$ reaches a sufficiently large number N, the tumor becomes detectable and no further transitions are allowed, i.e., the disease recurrence is defined by means of absorbing boundary conditions at the recurrence boundary $m+n = N$. The absorbing boundary condition at the extinction state $(0,0)$ (remission or cure) is automatically satisfied, since the transition rates are proportional to either m or n.

2.2 Master equation for discrete-state transition probabilities

From the transition rules described in Section 2.1, it follows that the time evolution of the state probabilities $p_{mn}(t)$ is given by the master equation

$$\frac{d\,|p(t)\rangle}{dt} = \hat{Q}\,|p(t)\rangle\,, \tag{2.5}$$

where $|p(t)\rangle$ is the probability vector, whose components are the state probabilities $p_{mn}(t)$, i.e., $|p(t)\rangle = \sum_{mn} p_{mn}(t)\,|m,n\rangle$. The infinitesimal transition operator $\hat{Q} = \hat{Q}_{bulk} + \hat{Q}_{edge}$ consists of a "bulk" part $\hat{Q}_{bulk}$ and an edge correction $\hat{Q}_{edge}$ that imposes the absorbing boundary conditions at the recurrence line $m+n=N$. In second-quantized language, the bulk part is given by

$$\begin{aligned}\hat{Q}_{bulk} = {} & (\lambda\,\hat{a}_+ + \mu_A\,\hat{a}_- + \xi\,\hat{d}_+\,\hat{a}_- - \lambda - \mu_A - \xi)\,\hat{n} + \\ & + (\mu_D\,\hat{d}_- + \eta\,\hat{d}_-\,\hat{a}_+ - \mu_D - \eta)\,\hat{m}\end{aligned} \tag{2.6}$$

and the edge correction is given by

$$\begin{aligned}\hat{Q}_{edge} = \textstyle\sum_{m=0}^{N} [\; & -\mu_D m\,|m-1,N-m\rangle\langle m,N-m| \\ & -\mu_A(N-m)\,|m,N-m-1\rangle\langle m,N-m| \\ & -\xi(N-m+1)\,|m,N-m\rangle\langle m-1,N-m+1| \\ & +(\lambda+\mu_A+\xi)(N-m)\,|m,N-m\rangle\langle m,N-m| \\ & -\eta\,(m+1)\,|m,N-m\rangle\langle m+1,N-m-1| \\ & +\,(\mu_D+\eta)\,m\,|m,N-m\rangle\langle m,N-m|\;]\,.\end{aligned} \tag{2.7}$$

In Eq. (2.7), $\hat{d}_+$ and $\hat{d}_-$ are respectively creation and annihilation operators for dormant tumor foci and $\hat{a}_+$ and $\hat{a}_-$ are those for the active foci. These operators are defined as:

$$\begin{aligned}\hat{d}_\pm\,|m,n\rangle &= |m\pm1,n\rangle\,, \\ \hat{a}_\pm\,|m,n\rangle &= |m,n\pm1\rangle\,.\end{aligned} \tag{2.8}$$

The number operators $\hat{m}$ and $\hat{n}$ are defined in the usual fashion:

$$\hat{m}\,|m,n\rangle = m\,|m,n\rangle\,,$$
$$\hat{n}\,|m,n\rangle = n\,|m,n\rangle\,. \tag{2.9}$$

The structure of the matrix $\hat{Q}$ is block-tridiagonal in the linear space $\mathcal{D}\otimes\mathcal{A}$, with block indices (m,m') that run across the Fock states of $\mathcal{D}$, and with indices (n,n') within each nonzero block that run across the Fock states of $\mathcal{A}$. The $N+1$ blocks along the diagonal are themselves tridiagonal and decrease in size as the $\mathcal{D}$-space index n increases. The structure of $\hat{Q}$, as described in Appendix A, is sufficiently complicated that explicit analytical solutions are not straightforward, although analytical formulas are available for the inversion of general tridiagonal and even certain types of block-tridiagonal matrices [71, 72]. Analytical and even stable numerical methods for general level-dependent QBD processes, i.e., QBD models with a block-tridiagonal matrix structure where the blocks are not constant along the diagonals [73], are scarce in the literature. Although a matrix-analytic method has been developed for these models in [74], it still relies on the ability to solve non-trivial matrix equations. Numerical methods have been developed for finding stationary distributions in level-dependent QBD models [75], but generally applicable numerical methods for finding transient solutions (other than the expensive matrix exponentiation) have yet to be developed [76]. For a special class of level-dependent QBD models with applications in biology and epidemiology, a method based on a continuous-fraction representation of the Laplace-transformed transition probabilities has recently been developed [76]. However, it is not applicable to the model defined by Eqs. (2.5), (2.6) and (2.7), where all the transition and birth/death rates can be nonzero.

For the moment, we will disregard constraints on the parameters (Eq. (2.1)) and consider the special case $\eta = \xi = \mu_D = 0$. In this limit, the system reduces to a continuous-time birth-and-death process with transition rates λn and $\mu_A n$ (otherwise known in queuing theory as the $M/M/1/N$ queue [69]) and with absorbing boundary states. The version

of this model when $N \to \infty$ (i.e., semi-infinite Markov chain) and reflecting boundary conditions has been studied extensively and analytical solutions have been obtained for its transient analysis using several techniques [77–81]. A version of the $M/M/1/N$ queuing model that is more relevant to our analysis is the one with finite N and absorbing boundary states, which is solved analytically in [82], where the transient solution is obtained and a simple expression is given for the large-time probability π_{rec} of absorption at the state $n = N$ (corresponding to recurrence in our model), under the initial condition $p_n(0) = \delta_{n,n_0}$:

$$\pi_{rec} = \frac{1 - (\mu_A/\lambda)^{n_0}}{1 - (\mu_A/\lambda)^N} \tag{2.10}$$

The probability of absorption at the zero-particle state $n = 0$ (corresponding to remission or cure) is then $\pi_{cure} = 1 - \pi_{rec}$. In the limit $N \to \infty$, we note that this model has a phase transition at $\mu_A/\lambda = 1$:

$$\pi_{rec} = \begin{cases} 1 - (\mu_A/\lambda)^{n_0} & \text{if} \quad \mu_A/\lambda \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

More details on the solution of the $M/M/1/N$ queuing model with absorbing boundary states can be found in Appendix B. In Section 3.1, it will be shown that a similar stationary solution also occurs in general in the QBD model defined by Eqs. (2.5), (2.6) and (2.7). We will show this both analytically in the continuum limit and in simulations of the discrete-state stochastic process.

2.3 Continuum limit of the discrete-state QBD model

A simple approach that is suitable for our tumor recurrence model is to take the continuum limit of the master equation (2.5), i.e., take the large N limit. Since the reciprocal of the detectable tumor size $(1/N)$ is a natural small parameter in the model, the master equation can be expanded in powers of $1/N$ and transformed into a continuum equation. The resulting partial differential equation may then be solved using well-developed methods [83].

There are two alternative ways to represent the time evolution of the stochastic process at hand [83–87], namely the *forward* master equation (2.5) and the *backward* master equation

$$\frac{d}{dt}\langle P(m,n;t)| = \langle P(m,n;t)|\,\hat{Q}^{\dagger}, \tag{2.12}$$

where $\hat{Q}^{\dagger}$ is the adjoint of the operator $\hat{Q}$ defined in Eqs. (2.6), (2.7) and the state vector $\langle P(m,n;t)|$ is defined as the probability to end up at a particular state $\langle m,n|$ at time t, starting from *any* state $\langle m_0,n_0|$ at time $t=0$, i.e.,

$$\langle P(m,n;t)| \equiv \sum_{m_0,n_0} \langle m_0,n_0|\,\Pr\{m,n;t|m_0,n_0;0\}. \tag{2.13}$$

When either Eq. (2.5) or Eq. (2.12) is used in the large-N expansion, we get respectively the *forward* or the *backward* Kolmogorov equation in the continuum limit $1/N \ll 1$ by truncating the expansion after the second term, as discussed further below.

2.3.1 Forward Kolmogorov approach

We can define a continuum limit of the state space $\Omega = \{(m,n)|m \geq 0; n \geq 0; m+n \leq N\}$ by defining continuous variables $x = m/N$ and $y = n/N$, which represent the dormant and active foci sub-populations as a fraction of the detectable tumor size N, respectively. We define a "tumor focus" as the resolution scale of our model: for example, we can define a focus as 1/1000 of a detectable tumor, in which case $N = 1000$ is a natural definition of recurrence (detectable tumor size). When $N \gg 1$, the lattice of discrete states (m,n) becomes a continuum as the spacings $\delta x = \delta y = 1/N$ decrease to zero. The discrete probabilities $p_{mn}(t)$ can then be replaced by a smooth probability density $\rho(\vec{x},t) = N^2 p_{mn}(t)$, where $\vec{x} = (x,y)$ is an arbitrary point in the bounded region $\Omega = \{(x,y)|x \geq 0; y \geq 0; x+y \leq 1\}$.

We proceed to take the continuum limit of the master equation by first replacing the raising/lowering operators (see Eqs. (2.5), (2.6) and (2.8)) by the corresponding translation

operators in the continuum, i.e.,

$$\hat{d}_{\pm} \to e^{\mp \delta x \, \partial/\partial x} \quad ,$$
$$\hat{a}_{\pm} \to e^{\mp \delta y \, \partial/\partial y} \quad , \tag{2.14}$$

where $\delta x = \delta y = 1/N$.

To see that the operators (2.14) are indeed the correct continuum versions of $\hat{d}_{\pm}$ and $\hat{a}_{\pm}$, let us apply (for example) the operator $\hat{a}_{\pm}$ to an arbitrary state vector $|\rho\rangle$:

$$\begin{aligned} \hat{a}_{\pm} |\rho\rangle &= \int \rho(\vec{x}) \; \hat{a}_{\pm} |\vec{x}\rangle \, d^2x \\ &= \int \rho(\vec{x}) \, |x, y \pm \delta y\rangle \, d^2x \\ &= \int \rho(x', y' \mp \delta y) \, \left|\vec{x}\,'\right\rangle d^2x' \\ &= \int e^{\mp \delta y \, \partial/\partial y'} \rho(\vec{x}\,') \, \left|\vec{x}\,'\right\rangle d^2x' \end{aligned} \tag{2.15}$$

The reason for the opposite signs between $\hat{a}_{\pm}$ and the exponent of $e^{\mp \delta y \, \partial/\partial y}$ is that the creation and annihilation operators are passive transformations, whereas the continuum translation operators are defined as active transformations. A Kramers-Moyal expansion of the master equation can then be obtained by expanding the operator (2.6) in powers of $\delta x = \delta y = 1/N$. Retaining only the first and the second terms in this large-size expansion, we obtain the two-dimensional Fokker-Planck equation

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = -\vec{\nabla} \cdot \vec{J}(\vec{x}, t), \tag{2.16}$$

which is a local continuity equation with a probability current density given by

$$\begin{aligned} \vec{J}(\vec{x}, t) = & \left[-(\eta x - \xi y)(\hat{e}_x - \hat{e}_y) - \mu_D x \hat{e}_x + (\lambda - \mu_A) y \hat{e}_y \right] \rho(\vec{x}, t) - \\ & - \frac{1}{2N} \left\{ (\hat{e}_x - \hat{e}_y) \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y} \right) \left[(\eta x + \xi y) \rho(\vec{x}, t) \right] + \right. \\ & \left. + \hat{e}_x \frac{\partial}{\partial x} \left[\mu_D x \rho(\vec{x}, t) \right] + \hat{e}_y \frac{\partial}{\partial y} \left[(\lambda + \mu_A) y \rho(\vec{x}, t) \right] \right\}, \end{aligned} \tag{2.17}$$

where each drift term is a product of the respective transition rate by the probability density, along the unit vector in the direction of the transition, and the terms proportional to $1/N$

represent diffusion, with a diffusion tensor that involves off-diagonal terms (i.e., the terms that involve mixed second derivatives $\partial^2\rho/\partial x\partial y$ are nonzero) and is dependent on the state variables (x, y).

Since Eq. (2.16) gives the probability density $\rho(\vec{x}, t)$ at any state $\vec{x} = (x, y)$ at time $t \geq 0$, given the initial condition $\rho(\vec{x}, 0)$, it corresponds to the well-known *forward* Kolmogorov equation [83–87]. Here we define the initial condition to be sharply peaked at a state $\vec{x}_0 = (x_0, y_0)$, i.e., $\rho(\vec{x}, 0) = \delta(\vec{x} - \vec{x}_0)$. Note that the normalization of the probability density $\rho(\vec{x}, t)$ is *not* preserved by the time evolution, because the probability flux exits through the recurrence boundary $x + y = 1$, and also accumulates at the cure state $(0, 0)$.

2.3.2 Boundary conditions for the forward equation

The boundary conditions for Eq. (2.16) are the following. At the recurrence line $x + y = 1$, the boundary condition is absorbing, i.e., $\rho(\vec{x}, t) = 0$. In the vicinity of the cure state $(0, 0)$ we define $\rho(\vec{x}, t) = 0$ on the line $x + y = \varepsilon$, where $\varepsilon \ll 1/N$ is a small parameter. In the weak limit (in the distributional sense) $\varepsilon \to 0$, the small region $x + y \leq \varepsilon$, representing the cure state, becomes a single absorbing point where the probability density collapses to a Dirac delta function weighted by the probability of cure before time t, which will henceforth be denoted by $p_{cure}(t)$. In other words, we can separate the probability density $\rho(\vec{x}, t)$ in a regular part $\rho_{reg}(\vec{x}, t)$ and a Dirac-delta component:

$$\rho(\vec{x}, t) = \rho_{reg}(\vec{x}, t) + p_{cure}(t)\delta(\vec{x}) \tag{2.18}$$

In Section 2.3.3, an equation for the function $p_{cure}(t)$ is derived (see Eq. (2.37)) using the backward Kolmogorov approach. In Section 3.1, the large-time limit of $p_{cure}(t)$ is determined in closed form in terms of the initial condition. The single-absorbing-point boundary condition above can be defined more rigorously by setting $\rho(\vec{x}, t) = 0$ on the line $x + y = \varepsilon$, where $\varepsilon \ll 1/N$ is sufficiently small (i.e., this line can be thought of as an absorbing boundary through which the probability flux enters a small region near the

origin), and by *defining*

$$\rho(\vec{x},t) = \frac{6}{\varepsilon^2}\left(1 - \frac{x+y}{\varepsilon}\right) p_{cure}(t) \tag{2.19}$$

for any (x,y) within the small region $\{(x,y)|x \geq 0; y \geq 0; x+y \leq \varepsilon\}$. When the probability density (2.19) is integrated over this small region, we get exactly $p_{cure}(t)$, i.e., the small region near the origin approximately represents the cure state. In the weak limit (in distributional sense) $\varepsilon \to 0$, the probability density at the origin becomes a Dirac delta distribution.

At the boundaries $x = 0$, $\varepsilon \leq y \leq 1$ and $y = 0$, $\varepsilon \leq x \leq 1$, the boundary conditions are both reflecting, i.e., the normal component of the probability current density must vanish. In other words, $\vec{J}_{reg}(\vec{x},t) \cdot \vec{n} = 0$, where $\vec{n}$ is the outward normal and $\vec{J}_{reg}(\vec{x},t)$ is the regular part of the current density defined by Eq. (2.17). This means there cannot be any flux crossing the boundaries $x = 0$ or $y = 0$, except at the cure state $(x,y) = (0,0)$, at which the regular part of $\rho(\vec{x},t)$ vanishes, whereas the delta peak works as a single absorbing point. The absorbing boundary conditions should be interpreted as follows: while the single absorbing point at the origin pins any probability that it absorbs to the cure state (thus the delta peak), the probability flux through the recurrence boundary $x + y = 1$ exits to the outer region $\{(x,y)|x \geq 0; y \geq 0; x+y \geq 1\}$ and never returns.

The boundary condition near the origin, as properly defined above, ensures that the Fokker-Planck equation (2.16) gives the correct probability conservation equation in its integral form. Indeed, using the boundary-condition scheme described above and integrating Eq. (2.16) over the area $\Omega_\varepsilon = \{(x,y)|x \geq 0; y \geq 0; \varepsilon \leq x+y \leq 1\}$, we get

$$\dot{p}_{risk}(t) \equiv \frac{d}{dt}\int_{\Omega_\varepsilon} \rho(\vec{x},t) d^2x = -\int_{\Omega_\varepsilon} \vec{\nabla} \cdot \vec{J}(\vec{x},t) d^2x, \tag{2.20}$$

where $p_{risk}(t)$ is the probability to be at risk until time t (i.e., to have neither relapsed nor been cured before time t) and $\dot{p}_{risk}(t) \equiv d\,p_{risk}(t)/dt$. Defining $p_{rec}(t)$ and $p_{cure}(t)$, respectively, as the probabilities of recurrence and cure at any time $\leq t$, we can write

$$p_{risk}(t) \equiv 1 - p_{rec}(t) - p_{cure}(t). \tag{2.21}$$

Using the divergence theorem and the reflecting boundary condition $\vec{J}(\vec{x},t)\cdot\vec{n}=0$ at the boundaries $x=0$, $\varepsilon\le y\le 1$ and $y=0$, $\varepsilon\le x\le 1$, we get

$$\frac{d}{dt}\int_{\Omega_\varepsilon}\rho(\vec{x},t)d^2x=-\int\limits_{\substack{x+y=1\\(1,0)\to(0,1)}}\vec{J}(\vec{x},t)\cdot\vec{n}\;dl-\int\limits_{\substack{x+y=\varepsilon\\(0,\varepsilon)\to(\varepsilon,0)}}\vec{J}(\vec{x},t)\cdot\vec{n}\;dl, \tag{2.22}$$

where $\vec{n}$ is the outward normal. This is equivalent to the statement

$$\dot{p}_{risk}(t)=-\dot{p}_{rec}(t)-\dot{p}_{cure}(t), \tag{2.23}$$

which is consistent with Eq. (2.21) above, as it should be.

Equation (2.16) is separable with respect to time, i.e., it can be reduced to an eigenvalue problem for the *forward* Fokker-Planck operator $\mathcal{L}_f$ defined by recasting Eq. (2.16) in the form $\dot{\rho}(\vec{x},t)=\mathcal{L}_f\rho(\vec{x},t)$, subject to the boundary conditions described above. However, the equation is not separable in the coordinates (x,y) and also depends on these variables explicitly through coefficients *inside* the differential operators. Finding a basis of two-variable eigenfunctions of $\mathcal{L}_f$ satisfying the mixed boundary conditions described above is a difficult problem, but unnecessary for our main goal, which is to derive an equation for the *recurrence-free survival function* $S(t)$, defined as the probability of no-recurrence before time t, i.e.,

$$S(t)\equiv\int_{\Omega_\varepsilon}\rho(\vec{x},t)d^2x+p_{cure}(t)=1-p_{rec}(t), \tag{2.24}$$

where the domain of integration is the region $\Omega_\varepsilon=\{(x,y)|x\ge 0;y\ge 0;\varepsilon\le x+y\le 1\}$ and $p_{rec}(t)$ and $p_{cure}(t)$ are respectively the time-dependent probabilities of recurrence and cure before time t.

The function $S(t)$ defined in Eq. (2.24) establishes the connection between data (recurrence-free survival curves) and the model. The *backward* Kolmogorov approach is often more appropriate to first-passage time problems [84–86, 88] and will be used in combination with Eq. (2.17) to derive an equation for $S(t)$. The current density derived from the forward equation (Eq. (2.17)) will be used to derive formulas for the probability flux into the cure state or through the recurrence boundary. Note that unlike the forward equation, the

backward equation cannot be expressed as a local continuity equation for the conservation of probability. Furthermore, the boundary conditions for the backward equation will be derived from those of the forward equation in Section 2.3.3.

2.3.3 Backward Kolmogorov approach

Starting with Eq. (2.12) and using the adjoint of the master equation operator (2.6), we can derive the backward Kolmogorov equation by means of a Kramers-Moyal expansion similar to that leading to the forward equation (2.16). It should be noted that the differential operators in the backward equation act on functions of the initial-state variables (x_0, y_0):

$$\begin{aligned}\frac{\partial \rho(\vec{x},\vec{x}_0,t)}{\partial t} &= [-(\eta x_0 - \xi y_0)(\hat{e}_x - \hat{e}_y) - \mu_D x_0 \hat{e}_x + \\ &\quad +(\lambda - \mu_A) y_0 \hat{e}_y] \cdot \vec{\nabla}_{\vec{x}_0} \rho(\vec{x},\vec{x}_0,t) + \\ &\quad + \sum_{i=1}^{2}\sum_{j=1}^{2} D_{ij}(\vec{x}_0) \frac{\partial^2 \rho(\vec{x},\vec{x}_0,t)}{\partial x_{0i} \partial x_{0j}} \\ &\equiv \mathcal{L}_b \; \rho(\vec{x},\vec{x}_0,t), \end{aligned} \tag{2.25}$$

where $\mathcal{L}_b$ is the backward operator, $(x_{01}, x_{02}) \equiv (x_0, y_0)$, and $D_{ij}(\vec{x}_0)$ are the components of the diffusion tensor

$$\begin{aligned}\mathbb{D}(\vec{x}_0) &= \begin{pmatrix} D_{11}(\vec{x}_0) & D_{12}(\vec{x}_0) \\ D_{21}(\vec{x}_0) & D_{22}(\vec{x}_0) \end{pmatrix} \\ &= \frac{1}{2N} \begin{pmatrix} \eta x_0 + \xi y_0 + \mu_D x_0 & -(\eta x_0 + \xi y_0) \\ -(\eta x_0 + \xi y_0) & \eta x_0 + \xi y_0 + (\lambda + \mu_A) y_0 \end{pmatrix}. \end{aligned} \tag{2.26}$$

The backward equation (2.25) is somewhat different from Eq. (2.16) in that the non-constant coefficients appear *outside* the differential operators. It is subject to the *final* condition that some state $\vec{x}$ will be reached at time t, starting from *anywhere* $(\vec{x}_0)$ in the state space. This explains why the backward equation is often more useful for first-passage time problems than its forward counterpart.

2.3.4 Boundary conditions for the main backward equation

The boundary conditions for the backward equation (2.25) can be derived from those of the forward equation (2.16) as follows (see e.g. [87]). Let f and g be arbitrary square-integrable functions defined on the domain $\Omega_\varepsilon = \{(x,y)|x \geq 0; y \geq 0; \varepsilon \leq x+y \leq 1\}$, satisfying the forward and the backward equations/boundary conditions, respectively. Let us consider the L^2 inner product

$$\langle \mathcal{L}_f f, g\rangle \equiv \int_{\Omega_\varepsilon} g\ \mathcal{L}_f f\ d^2x = -\int_{\Omega_\varepsilon} g\ \vec{\nabla}\cdot\vec{J}_f\ d^2x, \tag{2.27}$$

where $\vec{J}_f$ is the current density as defined in Eq. (2.17) for the density function f. Integrating the right-hand side of Eq. (2.27) by parts, it can be shown that

$$\langle \mathcal{L}_f f, g\rangle = \langle f, \mathcal{L}_b g\rangle - \oint_{\partial\Omega_\varepsilon} \left[g\ \vec{J}_f + f(\mathbb{D}\vec{\nabla}g)\right]\cdot\vec{n}\ dl, \tag{2.28}$$

where $\mathbb{D}(\vec{x})$ is the diffusion tensor defined in Eq. (2.26).

Since the backward operator is the adjoint of the forward operator, i.e., $\mathcal{L}_b = \mathcal{L}_f^\dagger$, we must have $\langle \mathcal{L}_f f, g\rangle = \langle f, \mathcal{L}_b g\rangle$ for any square-integrable functions f, g defined on the domain Ω_ε that satisfy the forward and backward equation/boundary conditions, respectively. Therefore, it follows that the boundary term on the right-hand side of Eq. (2.28) must vanish for any such f, g. This means that given boundary conditions on any function f satisfying the forward equation, the boundary conditions for any function g satisfying the backward equation have to be chosen in such a way that the integrand on the second term of Eq. (2.28) vanishes. It then follows that $g = 0$ for absorbing boundaries and $(\mathbb{D}\vec{\nabla}g)\cdot\vec{n} = 0$ for reflecting boundaries. Hence, on both absorbing boundaries $x_0 + y_0 = \varepsilon$ and $x_0 + y_0 = 1$, the boundary conditions are Dirichlet homogeneous, i.e., $\rho(\vec{x}, \vec{x}_0, t) = 0$. Using Eqs. (2.25) and (2.26), we can also show that the boundary conditions at the reflecting boundaries are

$$\frac{\partial\rho(\vec{x},\vec{x}_0,t)}{\partial x_0}\Big|_{x_0=0} = \frac{\partial\rho(\vec{x},\vec{x}_0,t)}{\partial y_0}\Big|_{x_0=0} \tag{2.29}$$

at the boundary $x_0 = 0$, $\varepsilon \leq y_0 \leq 1$ and

$$\frac{\partial\rho(\vec{x},\vec{x}_0,t)}{\partial x_0}\Big|_{y_0=0} = \frac{\partial\rho(\vec{x},\vec{x}_0,t)}{\partial y_0}\Big|_{y_0=0} \tag{2.30}$$

at the boundary $y_0 = 0$, $\varepsilon \leq x_0 \leq 1$.

2.3.5 Probability flux through the absorbing boundaries

The probability flux through the recurrence boundary can be obtained by integrating the normal component of the current density (2.17) along the line $x + y = 1$,

$$\dot{p}_{rec}(\vec{x}_0, t) = -\dot{S}(\vec{x}_0, t) = \int\limits_{\substack{x+y=1 \\ (1,0)\to(0,1)}} \vec{J}(\vec{x}, \vec{x}_0, t) \cdot \vec{n}\, dl, \tag{2.31}$$

where $S(\vec{x}_0, t)$ is the recurrence-free survival function (see Eqs. (2.24) and (2.22)) and $\dot{p}_{rec}(\vec{x}_0, t) \equiv \frac{\partial}{\partial t} p_{rec}(\vec{x}_0, t)$. For $\mu_D = 0$, using the boundary condition that the probability density has to vanish at the recurrence line $x + y = 1$, we get

$$\dot{p}_{rec}(\vec{x}_0, t) = -\frac{1}{2N}(\lambda + \mu_A) \int_0^1 y \left.\frac{\partial \rho(x, y, \vec{x}_0, t)}{\partial y}\right|_{x=1-y} dy. \tag{2.32}$$

Similarly, using the scheme described in Section 2.3.2 for the boundary condition near the origin (see discussion below Eq. (2.18) and also Eq. (2.22)), for $\mu_D = 0$ we find that the probability flux into the cure state $(0, 0)$ is given by

$$\dot{p}_{cure}(\vec{x}_0, t) = -\frac{1}{2N}(\lambda + \mu_A) \lim_{\varepsilon \to 0+} \int_0^\varepsilon y \left.\frac{\partial \rho(x, y, \vec{x}_0, t)}{\partial y}\right|_{x=\varepsilon-y} dy. \tag{2.33}$$

2.3.6 Equations for time-dependent probabilities of cure and recurrence

Partial differential equations for $p_{rec}(t)$ and $p_{cure}(t)$ (respectively defined above as the time-dependent probabilities of recurrence and cure at any time $\leq t$) can be derived by first doing the differentiation and integration operations on the right-hand sides of Eqs. (2.32) and (2.33) on both sides of Eq. (2.25). These operations commute with the backward operator $\mathcal{L}_b$ defined in Eq. (2.25), because the latter only acts on the initial-condition variables (x_0, y_0). By doing so, we get

$$\begin{aligned} \ddot{p}_{rec}(\vec{x}_0, t) &= \mathcal{L}_b\, \dot{p}_{rec}(\vec{x}_0, t), \\ \ddot{p}_{cure}(\vec{x}_0, t) &= \mathcal{L}_b\, \dot{p}_{cure}(\vec{x}_0, t), \end{aligned} \tag{2.34}$$

where $\mathcal{L}_b$ is the backward operator defined in Eq. (2.25) and we have used the notation $p_{rec}(\vec{x}_0, t)$, $p_{cure}(\vec{x}_0, t)$ to recall the dependence on the initial condition $\vec{x}_0 = (x_0, y_0)$. Integrating from time $t = 0$ to an arbitrary time t, we get

$$\begin{aligned} \dot{p}_{rec}(\vec{x}_0, t) - \dot{p}_{rec}(\vec{x}_0, 0) &= \mathcal{L}_b \, p_{rec}(\vec{x}_0, t), \\ \dot{p}_{cure}(\vec{x}_0, t) - \dot{p}_{cure}(\vec{x}_0, 0) &= \mathcal{L}_b \, p_{cure}(\vec{x}_0, t), \end{aligned} \tag{2.35}$$

where we have used the initial conditions $p_{rec}(\vec{x}_0, 0) = p_{cure}(\vec{x}_0, 0) = 0$. Assuming a Dirac-delta peak for the initial state-space probability density, i.e., $\rho(\vec{x}, \vec{x}_0, 0) = \delta^2(\vec{x} - \vec{x}_0)$, it follows from Eqs. (2.32) and (2.33) that the initial flux into both absorbing boundaries vanishes, i.e., $\dot{p}_{rec}(\vec{x}_0, 0) = \dot{p}_{cure}(\vec{x}_0, 0) = 0$. Therefore, it follows that $p_{rec}(\vec{x}_0, t)$ and $p_{cure}(\vec{x}_0, t)$ satisfy the backward equations

$$\dot{p}_{rec}(\vec{x}_0, t) = \mathcal{L}_b \, p_{rec}(\vec{x}_0, t), \tag{2.36}$$

$$\dot{p}_{cure}(\vec{x}_0, t) = \mathcal{L}_b \, p_{cure}(\vec{x}_0, t). \tag{2.37}$$

Since the recurrence-free survival function is defined as $S(\vec{x}_0, t) = 1 - p_{rec}(\vec{x}_0, t)$, it must satisfy the PDE

$$\dot{S}(\vec{x}_0, t) = \mathcal{L}_b \, S(\vec{x}_0, t). \tag{2.38}$$

For $\mu_D = 0$, we can simplify the backward equation (2.25) by transforming to the new variables $z_0 \equiv x_0$, $w_0 \equiv x_0 + y_0$,

$$\frac{\partial \rho(\vec{z}, \vec{z}_0, t)}{\partial t} = \mathcal{L}_b \, \rho(\vec{z}, \vec{z}_0, t), \tag{2.39}$$

where $\vec{z} \equiv (z, w)$ and the transformed backward operator is given by

$$\begin{aligned} \mathcal{L}_b = &- \left[\eta z_0 - \xi(w_0 - z_0)\right] \frac{\partial}{\partial z_0} + (\lambda - \mu_A)(w_0 - z_0)\frac{\partial}{\partial w_0} + \\ &+ \frac{1}{2N}\left[\eta z_0 + \xi(w_0 - z_0)\right] \frac{\partial^2}{\partial z_0^2} + \frac{1}{2N}(\lambda + \mu_A)(w_0 - z_0)\frac{\partial^2}{\partial w_0^2} \end{aligned} \tag{2.40}$$

In the new variables (z_0, w_0), the boundary conditions for Eqs. (2.36), (2.37), (2.38) and (2.39) are summarized in Table 2.1. The Neumann boundary conditions for Eq. (2.39) at

Table 2.1: Boundary conditions for the PDEs satisfied by a few relevant functions of the initial-state variables $\vec{z}_0 = (z_0, w_0)$, where $z_0 \equiv x_0$ and $w_0 \equiv x_0 + y_0$. Table reproduced from [1].

Function	Equation	Boundary			
		$w_0 = \varepsilon$	$w_0 = 1$	$z_0 = 0$	$z_0 = w_0$
$\rho(\vec{z}, \vec{z}_0, t)$	$\dot{\rho} = \mathcal{L}_b\, \rho$	$\rho = 0$	$\rho = 0$	$\frac{\partial \rho}{\partial z_0} = 0$	$\frac{\partial \rho}{\partial z_0} = 0$
$p_{rec}(\vec{z}_0, t)$	$\dot{p}_{rec} = \mathcal{L}_b\, p_{rec}$	$p_{rec} = 0$	$p_{rec} = 1$	$\frac{\partial p_{rec}}{\partial z_0} = 0$	$\frac{\partial p_{rec}}{\partial z_0} = 0$
$p_{cure}(\vec{z}_0, t)$	$\dot{p}_{cure} = \mathcal{L}_b\, p_{cure}$	$p_{cure} = 1$	$p_{cure} = 0$	$\frac{\partial p_{cure}}{\partial z_0} = 0$	$\frac{\partial p_{cure}}{\partial z_0} = 0$
$S(\vec{z}_0, t)$	$\dot{S} = \mathcal{L}_b\, S$	$S = 1$	$S = 0$	$\frac{\partial S}{\partial z_0} = 0$	$\frac{\partial S}{\partial z_0} = 0$
$T_{rec}^{(1)}(\vec{z}_0)$	$\mathcal{L}_b \left[p_{rec}(\infty)\, T_{rec}^{(1)} \right] = -p_{rec}(\infty)$	undef.*	$T_{rec}^{(1)} = 0$	$\frac{\partial T_{rec}^{(1)}}{\partial z_0} = 0$	$\frac{\partial T_{rec}^{(1)}}{\partial z_0} = 0$

* The mean recurrence time diverges at the cure-state boundary $w_0 = \varepsilon$.

the boundaries $z_0 = 0$ and $z_0 = w_0$, given in Table 2.1, follow immediately from Eqs. (2.29) and (2.30), and those for Eqs. (2.36), (2.37) and (2.38) follow from the equations for the probability flux through the absorbing boundaries $x + y = 1$ and $x + y = \varepsilon$, given in 2.3.5.

The PDEs (2.36), (2.37) and (2.38), subject to the boundary conditions given in Table 2.1, are all separable in time, but not in the initial-state variables (z_0, w_0). However, they can be solved analytically in the large-time limit $t \to \infty$, as will be shown in Section 3.1 below.

2.3.7 Recurrence time probability distribution and moments

A key random variable in our model is the recurrence time henceforth denoted by T_{rec}, corresponding to the first passage time through the recurrence boundary. The normalized probability that T_{rec} lies between t and $t + dt$ can be obtained from a simple application of Bayes' theorem [86]:

$$\Pr\{t < T_{rec} < t + dt \,|\, \text{recurrence}\} = \frac{\dot{p}_{rec}(t)dt}{p_{rec}(\infty)} \tag{2.41}$$

In other words, the probability density function (PDF) for the recurrence time T_{rec} is $\dot{p}_{rec}(t)/p_{rec}(\infty) = -\dot{S}(t)/p_{rec}(\infty)$, where the probability $p_{rec}(\infty) = \int_0^\infty \dot{p}_{rec}(t)dt$ that recurrence takes place at any time $t < \infty$ is given in closed analytic form in Section 3.1

below (see Eq. (3.3)). The ratio $p_{rec}(t)/p_{rec}(\infty)$ then gives the cumulative distribution function (CDF) of T_{rec}.

Let

$$T_{rec}^{(n)} \equiv \int_0^\infty t^n \frac{\dot{p}_{rec}(t)}{p_{rec}(\infty)} dt \tag{2.42}$$

denote the n-th moment of the recurrence time T_{rec} ($n = 0, 1, 2, \dots$). From Eqs. (2.36) and (2.42), we get the hierarchy of equations below [84, 87], where the n-th moment of T_{rec} is related to its $(n-1)$-th moment:

$$\mathcal{L}_b \left[p_{rec}(\infty)\, T_{rec}^{(n)}(\vec{z}_0) \right] = -n\, p_{rec}(\infty)\, T_{rec}^{(n-1)}(\vec{z}_0) \tag{2.43}$$

Here we have used the notation $T_{rec}^{(n)}(\vec{z}_0)$ to recall the dependence on the initial condition $\vec{z}_0$. Since the function $p_{rec}(\infty)$ also depends on the initial condition through the variable w_0 (see Eq. (3.3) below), it cannot be taken outside the backward operator $\mathcal{L}_b$, because the latter acts on the initial condition variables (z_0, w_0).

The boundary condition for Eq. (2.43) at the absorbing boundary $w_0 = 1$ is $T_{rec}^{(n)}(\vec{z}_0)] = 0$. From Eq. (2.42) and Table 2.1, it also follows that Neumann boundary conditions $\frac{\partial}{\partial z_0}[p_{rec}(\infty)\, T_{rec}^{(n)}(\vec{z}_0)] = 0$ must be imposed at both reflecting boundaries $z_0 = 0$ and $z_0 = w_0$.

Letting $n = 1$ in Eq. (2.43), we get an equation for the Mean Recurrence Time (MRT), denoted here by $T_{rec}^{(1)}$:

$$\mathcal{L}_b \left[p_{rec}(\infty)\, T_{rec}^{(1)}(\vec{z}_0) \right] = -p_{rec}(\infty) \tag{2.44}$$

This is a key equation in our analysis, which will be used to find an approximate analytical formula for the MRT.

Chapter 3

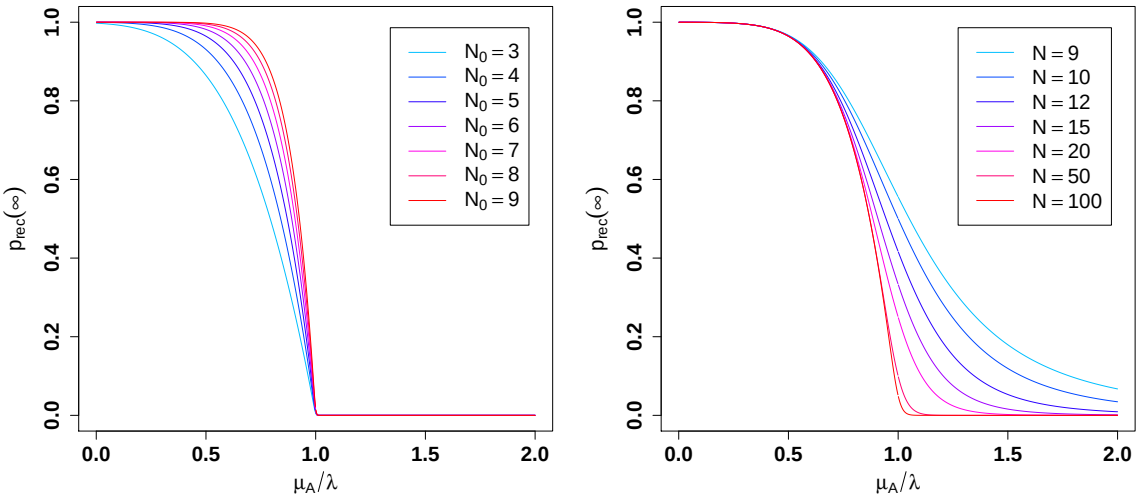
Analytical solutions and simulations

In this chapter, we discuss our main results, namely the stationary solution of Eq. (2.36) at large times in closed analytic form, as well as the "outer solution" of the mean recurrence time (MRT) equation (2.44) at leading (zeroth) order in $1/N$, which is approximately valid everywhere except inside thin boundary layers that stretch along the reflecting barriers $z_0 = 0$ and $z_0 = w_0$. For several choices of the parameters and initial conditions, these solutions are compared against simulations. We then describe a simple procedure to fit the model to recurrence-free survival data, using serous ovarian cancer data downloaded from the public database *The Cancer Genome Atlas* [89] as an example. Finally, using simulations, we show how the mean recurrence time (MRT) or the asymptotic (large-time) probability of no recurrence can be increased by extending chemotherapy in time or by increasing the tumor death-rate parameter during chemotherapy.

3.1 Stationary solution at large times

The stationary solution of Eq. (2.36), which satisfies $\dot{p}_{rec}(\vec{z}_0, t) = 0$, gives the large-time probability $p_{rec}(z_0, w_0, \infty)$ of absorption at the recurrence boundary $w = 1$, given the initial state (z_0, w_0). In the large-time limit $t \to \infty$, the probability of cure is given by $p_{cure}(\vec{z}_0, \infty) = 1 - p_{rec}(\vec{z}_0, \infty)$: recurrence or cure are the only possible fates at large times. Therefore, $p_{rec}(z_0, w_0, \infty)$ satisfies the *homogeneous* backward equation

$$\mathcal{L}_b \, p_{rec}(z_0, w_0, \infty) = 0, \tag{3.1}$$



(a) Varying N_0 for $N = 500$.

(b) Varying N for fixed $N_0 = 5$.

Figure 3.1: Large-time probability of recurrence *vs.* μ_A/λ for $\mu_D = 0$, obtained analytically using the backward Kolmogorov approach (see Eq. (3.3)). Panel 3.1a shows the effect of changing the total initial number of tumor foci (N_0), whereas panel 3.1b shows the effect of a finite detectable-tumor size N on the phase transition. Figures reproduced from [1].

where the backward operator in the variables (z_0, w_0) is given by Eq. (2.40). The boundary conditions are given in Table 2.1.

An ansatz solution to Eq. (3.1) would be a recurrence probability that only depends on the *total* initial number of tumor foci w_0 and not specifically on what fraction of this initial number are dormant foci (z_0), i.e., $\frac{\partial}{\partial z_0} p_{rec}(z_0, w_0, \infty) = 0$. This type of solution automatically satisfies the Neumann boundary conditions at the reflecting boundaries $z_0 = 0$ and $z_0 = w_0$ (see Table 2.1). Using the ansatz above and Eq. (2.40), the homogeneous PDE (3.1) becomes the following ODE in the variable w_0:

$$\frac{1}{2N}(\lambda + \mu_A)\frac{d^2 p_{rec}(\infty)}{dw_0^2} + (\lambda - \mu_A)\frac{dp_{rec}(\infty)}{dw_0} = 0, \tag{3.2}$$

where $p_{rec}(\infty) \equiv p_{rec}(z_0, w_0, \infty)$ only depends on w_0. This equation can be easily solved for the Dirichlet boundary conditions given in Table 2.1. The solution is

$$p_{rec}(\infty) = \frac{1 - e^{-2RN_0}}{1 - e^{-2RN}}, \tag{3.3}$$

where $N_0 = w_0 N$ is the initial total number of tumor foci and

$$R \equiv \frac{1 - \mu_A/\lambda}{1 + \mu_A/\lambda}. \tag{3.4}$$

In Eq. (3.3), the limit $\varepsilon \to 0$ has been taken. In the large-time limit $t \to \infty$, the probability of cure is given by

$$p_{cure}(\infty) = 1 - p_{rec}(\infty). \tag{3.5}$$

In Fig. 3.1, the function (3.3) is plotted for different values of a detectable tumor size N for a fixed initial number of foci N_0 and also for different initial conditions N_0 at a fixed N.

In the limit $N \to \infty$, this solution displays a phase transition at $\mu_A/\lambda = 1$:

$$p_{rec}(\infty) = \begin{cases} 1 - e^{-2RN_0} & \text{if} \quad \mu_A/\lambda \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Thus, in the limit $N \to \infty$, the large-time probability of cure for $\mu_A/\lambda \leq 1$ is $p_{cure}(\infty) = 1 - p_{rec}(\infty) = e^{-2RN_0}$, which is the large-time limit of the recurrence-free survival function $S(t)$.

The drift term of the Fokker-Planck equation (2.16) alone is not sufficient to reproduce the phase transition (3.6). The latter is the result of a combination of drift *and* diffusion in the presence of two opposite absorbing boundaries. For $\mu_A/\lambda > 1$, the initial delta peak $\rho(\vec{x}, 0) = \delta(\vec{x} - \vec{x}_0)$ travels toward the cure state $(0, 0)$. Even though it broadens as a result of diffusion, the broadening is not sufficient to reach the far-away recurrence boundary, since the diffusion tensor is linear in the (x, y) coordinates (see Eq. (2.26)) and thus decreases as the peak approaches the cure state. However, for $\mu_A/\lambda < 1$, the peak travels toward the recurrence boundary and broadening increases as the peak approaches it, as a result of the linear growth of the diffusion tensor with the (x, y) coordinates. In that case, the broadening is sufficient to allow partial absorption of the peak by the cure state, as long as μ_A/λ is sufficiently close to the critical value $\mu_A/\lambda = 1$. The result (3.3) is only slightly different

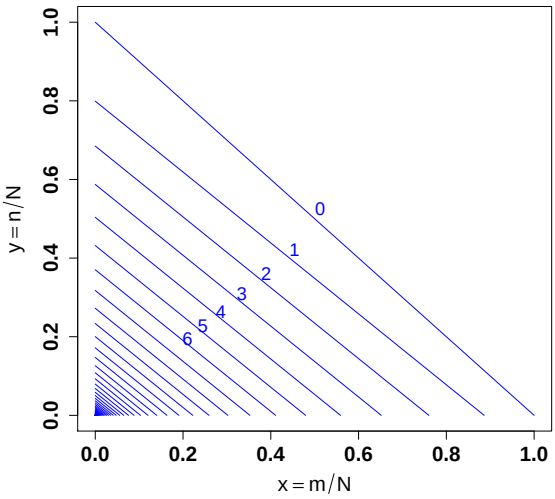


Figure 3.2: Level curves of the leading-order outer solution of the mean recurrence time equation (see Eqs. (3.15) and (3.16)) for $\mu_A/\lambda = 0.5$ and $\nu \equiv \xi/\eta = 2.5$. At large N, these curves are approximately valid outside boundary layers that exist close to the reflecting barriers $x_0 = 0$ and $y_0 = 0$. The sizes of these boundary layers vanish in the limit $N \to \infty$. The values of $T_{rec}^{(1)}$ at the first few level curves are given in the figure in units of the doubling time $1/\lambda$. The spacings between adjacent lines (which were plotted for $1/\lambda$ increments of $T_{rec}^{(1)}$) approach zero at the cure state $(0, 0)$, where the mean recurrence time diverges logarithmically. Figure reproduced from [1].

from the stationary solution of the M/M/1/N queue with absorbing boundary states (see Eq. (2.11). Further details are given in Appendix B). As will be shown in Section 3.3, our solution agrees with simulations of the discrete-state QBD process.

3.2 Approximate solution of the mean recurrence time equation

In this section, we will solve the mean recurrence time equation (2.44) analytically to leading (zeroth) order in $1/N$. This approximation is valid outside boundary layers near the reflecting barriers $z_0 = 0$ and $z_0 = w_0$, the sizes of which vanish as $N \to \infty$. The leading order solution can be obtained by neglecting the second derivative terms in Eq. (2.44). This results in the first order PDE

$$[\eta z_0 - \xi(w_0 - z_0)] \frac{\partial T_{rec}^{(1)}}{\partial z_0} - (\lambda - \mu_A)(w_0 - z_0) \frac{\partial T_{rec}^{(1)}}{\partial w_0} = 1, \tag{3.7}$$

which can be solved by the method of characteristics, as follows.

Along the characteristic curves (which are parametrized by a parameter r), the PDE (3.7) becomes the set of ODEs

$$\begin{aligned}\frac{dz_0}{dr} &= \eta z_0 - \xi(w_0 - z_0)\\ \frac{dw_0}{dr} &= -(\lambda - \mu_A)(w_0 - z_0)\\ \frac{dT_{rec}^{(1)}}{dr} &= 1\end{aligned} \tag{3.8}$$

This last equation immediately gives $r = T_{rec}^{(1)}$ (here we can choose the constant of integration to be zero, since it can be absorbed into the constants s_1 and s_2 in Eqs. (3.9) below).

Solving the first two ODEs yields

$$\begin{aligned}z_0 &= s_1\, e^{\Lambda_- T_{rec}^{(1)}} + s_2\, e^{\Lambda_+ T_{rec}^{(1)}},\\ w_0 &= \frac{1}{\xi}(\eta + \xi - \Lambda_-)\, s_1\, e^{\Lambda_- T_{rec}^{(1)}} + \frac{1}{\xi}(\eta + \xi - \Lambda_+)\, s_2\, e^{\Lambda_+ T_{rec}^{(1)}},\end{aligned} \tag{3.9}$$

where s_1 and s_2 are constants of integration and

$$\Lambda_\pm = a \pm b, \tag{3.10}$$

where

$$a \equiv \frac{1}{2}(\eta + \xi + \mu_A - \lambda) \tag{3.11}$$

and

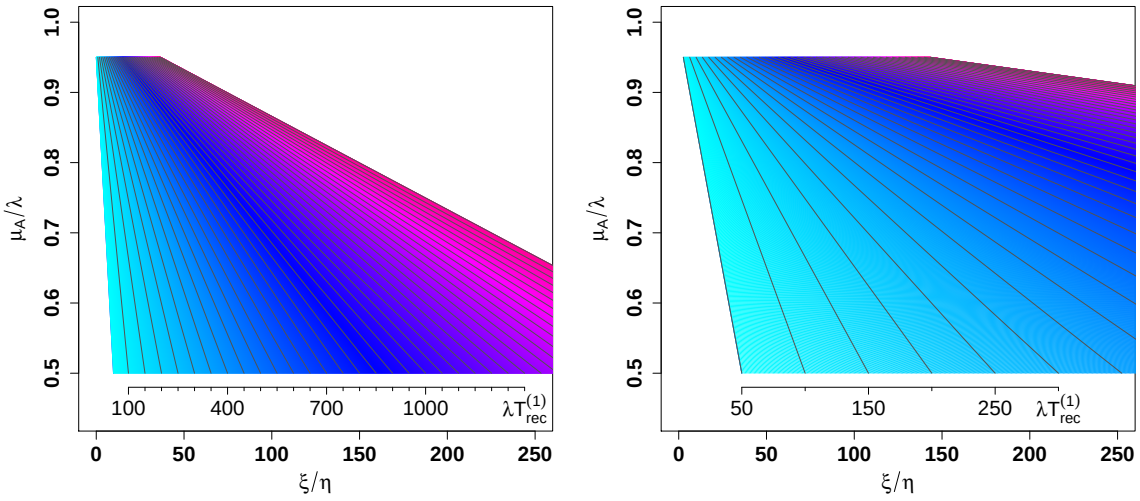
$$b \equiv \left[a^2 + \eta(\lambda - \mu_A)\right]^{1/2}. \tag{3.12}$$

By imposing the absorbing boundary condition $T_{rec}^{(1)} = 0$ at $w_0 = 1$, we find the relation

$$s_1 = \frac{\xi - (\eta + \xi - \Lambda_+)s_2}{\eta + \xi - \Lambda_-} \tag{3.13}$$

Using this relation in Eqs. (3.9), it follows that

$$\begin{aligned}(\eta + \xi - \Lambda_-)z_0 - \xi e^{\Lambda_- T_{rec}^{(1)}} &= s_2\left[(\eta + \xi - \Lambda_-)e^{\Lambda_+ T_{rec}^{(1)}} - (\eta + \xi - \Lambda_+)e^{\Lambda_- T_{rec}^{(1)}}\right],\\ \xi w_0 - \xi e^{\Lambda_- T_{rec}^{(1)}} &= s_2\left[(\eta + \xi - \Lambda_+)e^{\Lambda_+ T_{rec}^{(1)}} - (\eta + \xi - \Lambda_+)e^{\Lambda_- T_{rec}^{(1)}}\right].\end{aligned} \tag{3.14}$$



(a) Level curves for $N_0 = 7$, $m_0 = 3$.

(b) Level curves for $N_0 = 50$, $m_0 = 20$.

Figure 3.3: Level curves in parameter space of the leading-order outer solution of the mean recurrence time (MRT) equation (see Eqs. (3.15) and (3.16)) for two different initial conditions, with $N = 100$. The ruler at the bottom of each plot gives the values of the MRT in units of the doubling time $1/\lambda$. Figures reproduced from [1].

Eliminating the constant of integration s_2 from Eqs. (3.14) and simplifying, we finally arrive at the solution given by Eqs. (3.15) and (3.16).

The solution is

$$T_{rec}^{(1)} = -\frac{1}{b} \ln u, \tag{3.15}$$

where u is the only root of the transcendental equation

$$\begin{aligned} u^{a/b} \left\{ -\xi w_0 \left[(\eta + \xi - a)(1 - u^2) + b(1 + u^2) \right] + \right. \\ \left. + z_0 \left[(\eta + \xi - a)^2 - b^2 \right] (1 - u^2) \right\} + 2\xi b u = 0, \end{aligned} \tag{3.16}$$

where a and b given by Eqs. (3.11) and (3.12).

The finite-N solution of Eq. (2.44) converges to the leading-order approximation given by Eqs. (3.15) and (3.16) pointwise, but not uniformly. Indeed, the approximate solution above does not satisfy the homogeneous Neumann boundary conditions at $z_0 = 0$ or $z_0 = w_0$ (see Table 2.1). Near each reflecting barrier, there is a boundary layer within which the zeroth-order approximation in $1/N$ fails. This is seen in the form of the mean recurrence

time (MRT) level curves given by Eqs. (3.15) and (3.16), which are straight lines that are not parallel to the recurrence boundary $w_0 = 1$, except asymptotically in the limit $T_{rec}^{(1)} \to 0$ (see Fig. 3.2). In reality, close to each reflecting barrier (within some distance that vanishes in the limit $N \to \infty$), the exact (finite-N) level curves bend toward the boundary, where these curves end at a right angle. At fixed initial conditions and as a function of the parameters μ_A/λ and ξ/η, the shapes of the MRT level curves given by Eqs. (3.15) and (3.16) in parameter space are shown in Fig. 3.3 for two different initial conditions, namely $N_0 = 7$, $m_0 = 3$ and $N_0 = 50$, $m_0 = 20$.

The PDE (2.44), along with its boundary conditions given in Table 2.1, is a singular perturbation problem that should be handled by special perturbation methods, because the small parameter $1/N$ premultiplies the second-derivative terms in the backward operator (2.40). Further inspection shows that Eq. (2.44) is structured in such a way that the problem can be treated by boundary-layer theory [90]. Since the solution of Eq. (2.44) obtained by dropping the diffusion ($1/N$) terms in the backward operator (2.40) is only valid *outside* the boundary layers that exist near the reflecting boundaries $z_0 = 0$ and $z_0 = w_0$ (the sizes of which vanish as $N \to \infty$), in the language of boundary layer theory the solution given by Eqs. (3.15) and (3.16) is called the "outer solution" of the boundary value problem. The "inner solutions", on the other hand, require proper rescaling of the variables before these solutions can be expanded asymptotically in the parameter $1/N$; in this case, the second-derivative terms in Eq. (2.40) cannot be neglected inside each boundary layer, since they become comparable to the first-derivative terms within each layer.

An approximate composite solution that would be valid everywhere can in principle be obtained by the method of matched asymptotic expansions, which requires solving Eq. (2.44) both inside and outside the boundary layers [90]. In this work, however, only the leading-order outer solution is given (Eqs. (3.15) and (3.16) above). In practice, the outer solution itself is already a remarkable and useful result, even to lowest order in $1/N$, which agrees reasonably well with simulations of the model (as will be shown in Section 3.3 below),

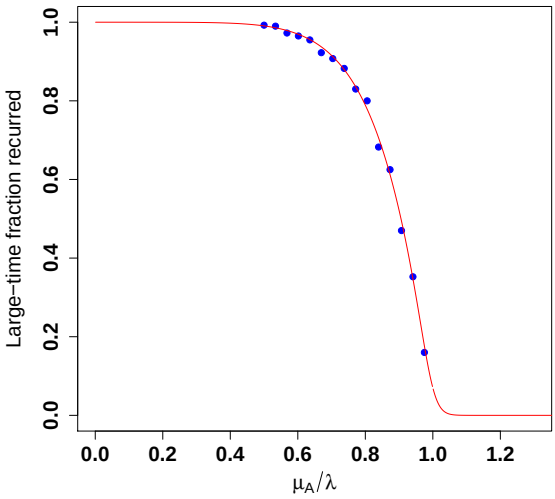


Figure 3.4: Large-time fraction of patients for which the tumor recurred, obtained in a simulation of the QBD model with 400 hypothetical patients. The parameters are $\mu_D = 0$, $\nu \equiv \xi/\eta = 5$, $m_0 + n_0 = 7$, $m_0 = 3$ and $N = 100$. The red curve shows the analytical result given by Eq. (3.3). Figure reproduced from [1].

except for small corrections that can in principle be calculated using the boundary-layer method.

3.3 Simulations

The discrete-state QBD process was simulated for several values of the model parameters and initial conditions using a pseudo random number generator. The results from the simulations were then compared to the large-time stationary solution given by Eq. (3.3), and also to the outer solution given by Eqs. (3.15) and (3.16). The time step was chosen to be such that the transition probabilities would always remain sufficiently small within the range of the transition rates used in the simulations, even for transitions between states with large numbers $m, n \sim N$. For a given set of parameters, each simulation tracked both the fraction of patients for which the tumor recurred and the recurrence times (from which the MRT was estimated) for an ensemble of 400 hypothetical patients. The detectable tumor size was set to $N = 100$.

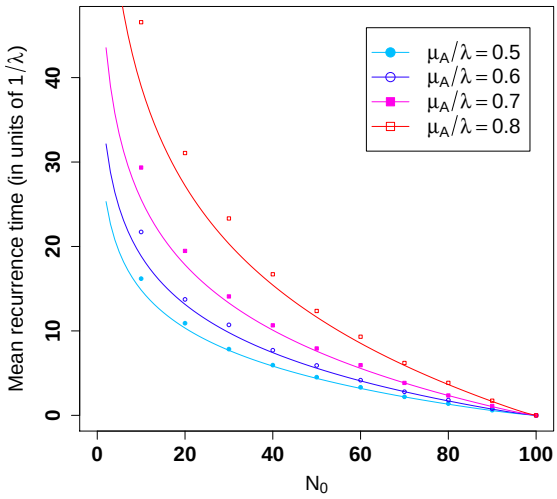
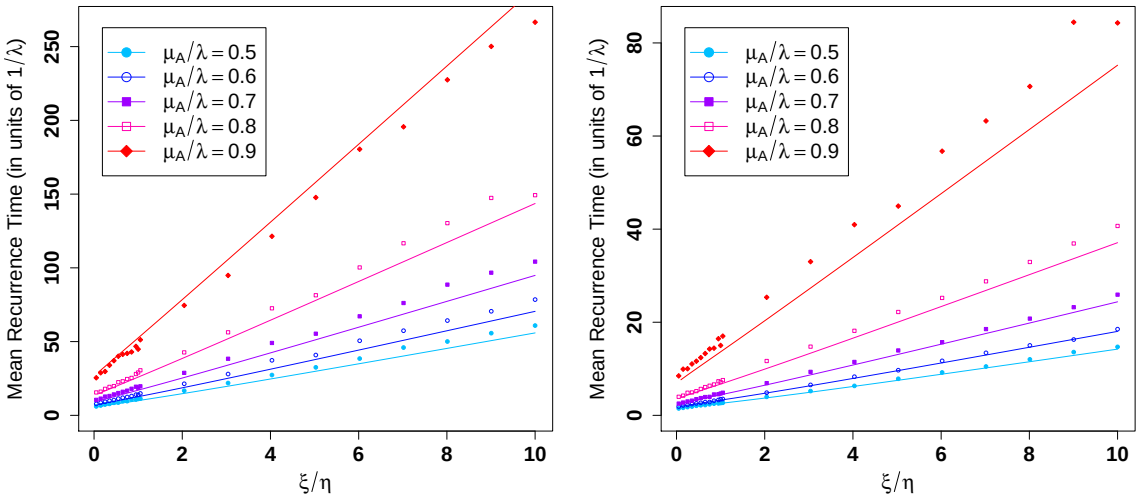


Figure 3.5: Mean recurrence time (MRT) obtained in simulations *vs.* initial total number N_0 along the line $m_0 = n_0 = N_0/2$, for different values of μ_A/λ with $\nu \equiv \xi/\eta = 2.5$ and $N = 100$. The simulations were run for 400 hypothetical patients. The smooth curves represent the leading-order outer solution of the MRT equation at different values of μ_A/λ (see Eqs. (3.15) and (3.16)). We note the logarithmic divergence at the cure state $N_0 = 0$. Figure reproduced from [1].

For a simulation with $\nu \equiv \xi/\eta = 5$ and initial condition $m_0 + n_0 = 7$, $m_0 = 3$, Fig. 3.4 shows a plot of the large-time fraction of patients for which the tumor recurred against μ_A/λ. The figure also shows that the result of the simulation agrees with the analytical formula given by Eq. (3.3). Fig. 3.5 shows a plot of the MRT obtained in simulations against the initial total number of tumor foci (N_0) along the line $m_0 = n_0 = N_0/2$, for 4 different values of μ_A/λ, with $\nu \equiv \xi/\eta = 2.5$. These results are compared to the smooth curves obtained from the leading-order outer solution of the MRT equation (Eqs. (3.15) and (3.16)). At the cure state $N_0 = 0$, the MRT diverges logarithmically and the discrepancy between simulations and the outer solution increases as N_0 approaches zero, due to the boundary-layer structure (the vicinity of the cure state is the region where the boundary layers along the reflecting barriers overlap).

For both initial conditions $N_0 = 7$, $m_0 = 3$ and $N_0 = 50$, $m_0 = 20$, Fig. 3.6 shows the MRT obtained in simulations against the ratio $\nu \equiv \xi/\eta$ at 5 different values of μ_A/λ,



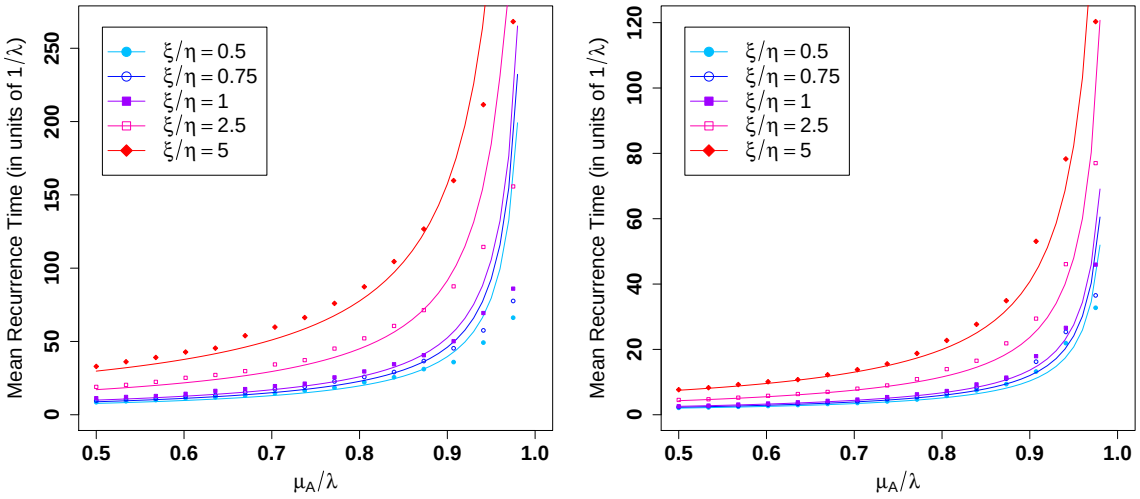
(a) $N_0 = 7$, $m_0 = 3$ (b) $N_0 = 50$, $m_0 = 20$

Figure 3.6: Mean recurrence time (MRT) obtained in simulations *vs.* the ratio $\nu \equiv \xi/\eta$ for two different initial conditions, with $N = 100$. Each set of points on each panel corresponds to a different value of μ_A/λ and the lines represent the leading-order outer solution of the MRT equation (see Eqs. (3.15) and (3.16)). The results represent an average over 400 hypothetical patients. Figures reproduced from [1].

and compares it to the curves obtained from the leading-order outer solution given by Eqs. (3.15) and (3.16). As expected, the higher the value of μ_A/λ, the larger is the discrepancy, since the finite-size effect is greatest near the critical point $\mu_A/\lambda = 1$.

It is worth noting that the ratio $\nu \equiv \xi/\eta$ can in principle be measured in experiments by estimating the relative times that the cells in a tumor spend in dormant versus active phases of the cell cycle, for example, through reconstruction of cell cycle dynamics from single-cell transcriptome data [91–93].

Fig. 3.7 shows the MRT obtained in simulations against μ_A/λ for the initial conditions $N_0 = 7$, $m_0 = 3$ and $N_0 = 50$, $m_0 = 20$, at different values of the ratio $\nu \equiv \xi/\eta$. These results are compared to the smooth curves obtained from Eqs. (3.15) and (3.16). To leading order, the MRT given by the outer solution diverges at the critical point $\mu_A/\lambda = 1$, as shown by Eqs. (3.15) and (3.16). The larger discrepancy near the critical point can be explained by a finite size effect in the phase transition given by Eq. (3.3), which effectively shifts the



(a) $N_0 = 7$, $m_0 = 3$ (b) $N_0 = 50$, $m_0 = 20$

Figure 3.7: Mean recurrence time (MRT) obtained in simulations *vs.* μ_A/λ for two different initial conditions, with $N = 100$. Each set of points on each panel corresponds to a different value of the ratio $\nu \equiv \xi/\eta$ and the smooth curves represent the leading-order outer solution of the MRT equation (see Eqs. (3.15) and (3.16)). The simulations were run for 400 hypothetical patients. At leading order, the MRT given by the outer solution diverges at the critical point $\mu_A/\lambda = 1$. Figures reproduced from [1].

critical point to a value slightly higher than $\mu_A/\lambda = 1$ (see plot in Fig. 3.1b).

We note that the continuum limit of the discrete-state master equation is a good approximation, since it agrees reasonably well with the simulations even to leading (zeroth) order in $1/N$. This is a remarkable feature of our model, since the diffusion limit sometimes fails for other types of birth-and-death processes, especially ones that involve non-linear transition rates [94].

3.4 Moments of the tumor-size distribution

Let $\mathbb{E}_t[\vec{z}]$ be the first moment of $\vec{z} = (z, w)$, restricted to the sample space Ω_ε of tumors at risk of recurrence, i.e.,

$$\mathbb{E}_t[\vec{z}] = \frac{1}{p_{risk}(\vec{z}_0, t)} \int_{\Omega_\varepsilon} \vec{z}\, \rho(\vec{z}, \vec{z}_0, t)\, d^2z \tag{3.17}$$

where

$$p_{risk}(\vec{z}_0, t) = \int_{\Omega_\varepsilon} \rho(\vec{z}, \vec{z}_0, t)\, d^2z. \tag{3.18}$$

From Eqs. (2.39) and (2.40), it follows that $\mathbb{E}_t[\vec{z}]$ is given by the equation

$$\frac{\partial}{\partial t}\left[\, p_{risk}(\vec{z}_0, t)\, \mathbb{E}_t[\vec{z}]\,\right] = \mathcal{L}_b\left[\, p_{risk}(\vec{z}_0, t)\, \mathbb{E}_t[\vec{z}]\,\right] \tag{3.19}$$

If the recurrence boundary at $w = 1$ is neglected, the solution of (3.19) is given by

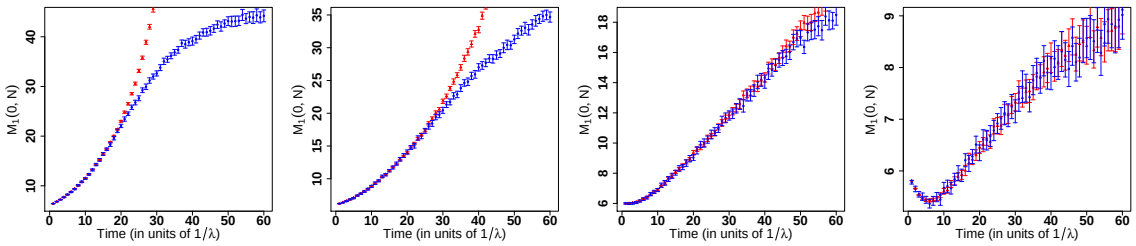
$$p_{risk}(\vec{z}_0, t)\, \mathbb{E}_t[\vec{z}] = e^{t\mathcal{L}_b}\, \vec{z}_0 \tag{3.20}$$

For $n = 0, 1, 2, \cdots$, it is easy to show by induction that

$$\begin{pmatrix} \mathcal{L}_b^n\, z_0 \\ \mathcal{L}_b^n\, w_0 \end{pmatrix} = \begin{pmatrix} -(\eta + \xi) & \xi \\ -(\lambda - \mu_A) & \lambda - \mu_A \end{pmatrix}^n \begin{pmatrix} z_0 \\ w_0 \end{pmatrix} \tag{3.21}$$

Using Eq. (3.21) to compute the right-hand side of Eq. (3.20), we find that in the absence of the recurrence boundary, the first moment of the tumor-size distribution is given by

$$\mathbb{E}_t[w] = [p_{risk}(\vec{z}_0, t)]^{-1}\, \frac{1}{b} e^{-at} \left\{ w_0 \left[b \cosh(bt) + (\eta + \xi - \alpha) \sinh(bt) \right] - z_0(\lambda - \mu) \sinh(bt) \right\}, \tag{3.22}$$



(a) $\mu_A/\lambda = 0.8$ (b) $\mu_A/\lambda = 0.9$ (c) $\mu_A/\lambda = 1.0$ (d) $\mu_A/\lambda = 1.1$

Figure 3.8: First moment of the tumor size distribution for the initial condition $m_0 = n_0 = 3$ and with $\xi/\eta = 2.5$ and $N = 100$, at different values of μ_A/λ. The blue dots are the first moment from the simulations, and the triangle-shaped red dots are the first moment given Eq. (3.22), in which the recurrence boundary is neglected.

where a and b are given by Eqs. (3.11) and (3.12).

The second-order moments (restricted to the sample space Ω_ε of tumors at risk of recurrence) $\mathbb{E}_t[z^2]$, $\mathbb{E}_t[zw]$ and $\mathbb{E}_t[w^2]$ are given by

$$\frac{\partial}{\partial t}[\, p_{risk}(\vec{z}_0, t)\ \mathbb{E}_t[\mathbf{u}]\,] = \mathcal{L}_b[\, p_{risk}(\vec{z}_0, t)\ \mathbb{E}_t[\mathbf{u}]\,] \tag{3.23}$$

where

$$\mathbf{u} \equiv \begin{pmatrix} z^2 \\ zw \\ w^2 \end{pmatrix}. \tag{3.24}$$

If the recurrence boundary at $w = 1$ is neglected, the solution of (3.23) is given by

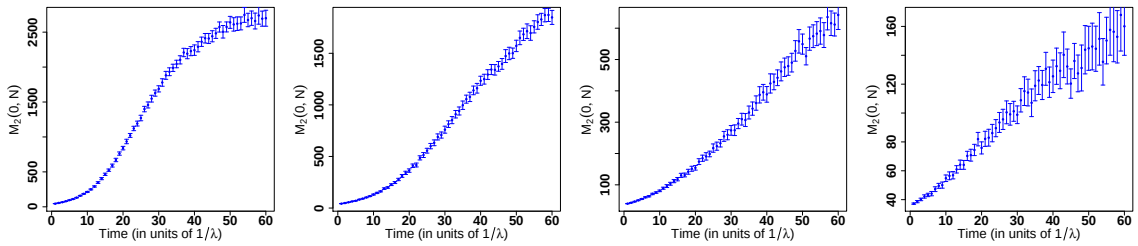
$$p_{risk}(\vec{z}_0, t)\ \mathbb{E}_t[\mathbf{u}] = e^{t\mathcal{L}_b}\ \mathbf{u_0}, \tag{3.25}$$

where

$$\mathbf{u_0} \equiv \begin{pmatrix} z_0^2 \\ z_0 w_0 \\ w_0^2 \end{pmatrix}. \tag{3.26}$$

It is easy to show that

$$\mathcal{L}_b\ \mathbf{u_0} = \mathbf{A u_0} + \mathbf{v_0}, \tag{3.27}$$



(a) $\mu_A/\lambda = 0.8$ (b) $\mu_A/\lambda = 0.9$ (c) $\mu_A/\lambda = 1.0$ (d) $\mu_A/\lambda = 1.1$

Figure 3.9: Second moment of the tumor size distribution for the initial condition $m_0 = n_0 = 3$ and with $\xi/\eta = 2.5$ and $N = 100$, at different values of μ_A/λ.

where

$$\mathbf{A} = \begin{pmatrix} -2(\eta+\xi) & 2\xi & 0 \\ 0 & -2a & \xi \\ 0 & -2(\lambda-\mu_A) & 2(\lambda-\mu_A) \end{pmatrix} \tag{3.28}$$

and

$$\mathbf{v_0} = \begin{pmatrix} \frac{1}{N}(\eta-\xi)z_0 + \frac{1}{N}\xi w_0 \\ 0 \\ \frac{1}{N}(\lambda+\mu_A)(w_0-z_0) \end{pmatrix}. \tag{3.29}$$

It then follows by induction that for $n = 1, 2, \cdots,$

$$\mathcal{L}_b^n\ \mathbf{u_0} = \mathbf{A}^n \mathbf{u_0} + \sum_{k=0}^{n-1} \mathbf{A}^{n-1-k} \mathcal{L}_b^k\ \mathbf{v_0}. \tag{3.30}$$

Using Eq. (3.30) and the fact that $\mathbf{A} \otimes 1_b$ commutes with $\mathbf{I} \otimes \mathcal{L}_b$, it then follows by series expansion that Eq. (3.25) becomes

$$p_{risk}(\vec{z}_0, t)\ \mathbb{E}_t[\mathbf{u}] = e^{t\mathcal{L}_b}\ \mathbf{u_0} = e^{t\mathbf{A}}\mathbf{u_0} + (\mathbf{A} \otimes 1_b - \mathbf{I} \otimes \mathcal{L}_b)^{-1}(e^{t\mathbf{A}} \otimes 1_b - \mathbf{I} \otimes e^{t\mathcal{L}_b})\mathbf{v_0} \tag{3.31}$$

For the initial condition $m_0 = n_0 = 3$ and with the parameter value $\xi/\eta = 2.5$, the first two moments of the tumor size distribution were obtained in simulations for a few different values of μ_A/λ, and are plotted against time in Figs. 3.8 and 3.9. Note that the solution given by Eq. (3.22), which was derived for the special case when the recurrence boundary can be neglected, agrees with simulations for $\mu_A/\lambda > 1$, since in the limit $N \to \infty$ there is no recurrence if μ_A is greater than the critical value $\mu_{critical} = \lambda$. For $\mu_A/\lambda < 1$, Fig.

3.8 shows that Eq. (3.22) agrees with simulations for sufficiently short times, before the recurrence boundary is reached.

3.5 Fitting the model to recurrence-free survival data

Even in the context of univariate birth-and-death processes, estimating model parameters from data is, in general, a difficult problem [95]. In this section, we describe a simple procedure to fit the model to recurrence-free survival data through an example. Since the method that we describe below relies on the leading-order outer solution given by Eqs. (3.15) and (3.16), it is approximately valid (within $O(1/N)$ corrections) for initial conditions outside the boundary layers.

It is also worth noting that in our model, all tumor foci in a given individual are assumed to be equivalent in their time evolution. In a more realistic scenario, various tumor foci would start with different numbers of cells and grow at different rates, and consequently, would have different parameters associated with them. It is difficult to even estimate, let alone model, this heterogeneity. Our definition of a tumor focus can be thought of as an "average" over foci and the parameters associated with a tumor focus as the "average parameters" over foci. Since the goal of our analysis is not to make specific predictions for individual patients but rather to model overall trends, we believe that our model is useful in spite of such uncertainties.

Serous ovarian cancer data downloaded from the public database *The Cancer Genome Atlas* [89] (TCGA) were used to generate the recurrence-free survival curve shown in Fig. 3.10. Out of the 583 patients in the data set, 170 were censored, i.e., although these patients were free of recurrence by the time reported in the data set, they did not follow up afterwards, so that their recurrence/no-recurrence status beyond the time of the last follow-up remained unknown. For this reason, the Kaplan-Meier product-limit estimator [96] was used to estimate the time-dependent probability of no recurrence. A brief description of the Kaplan-Meier method is given in Appendix C.

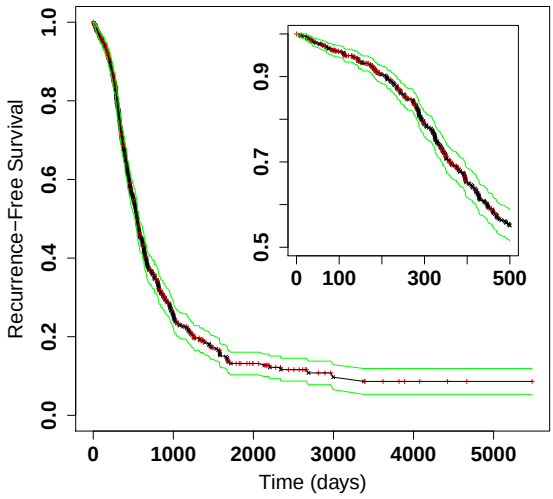
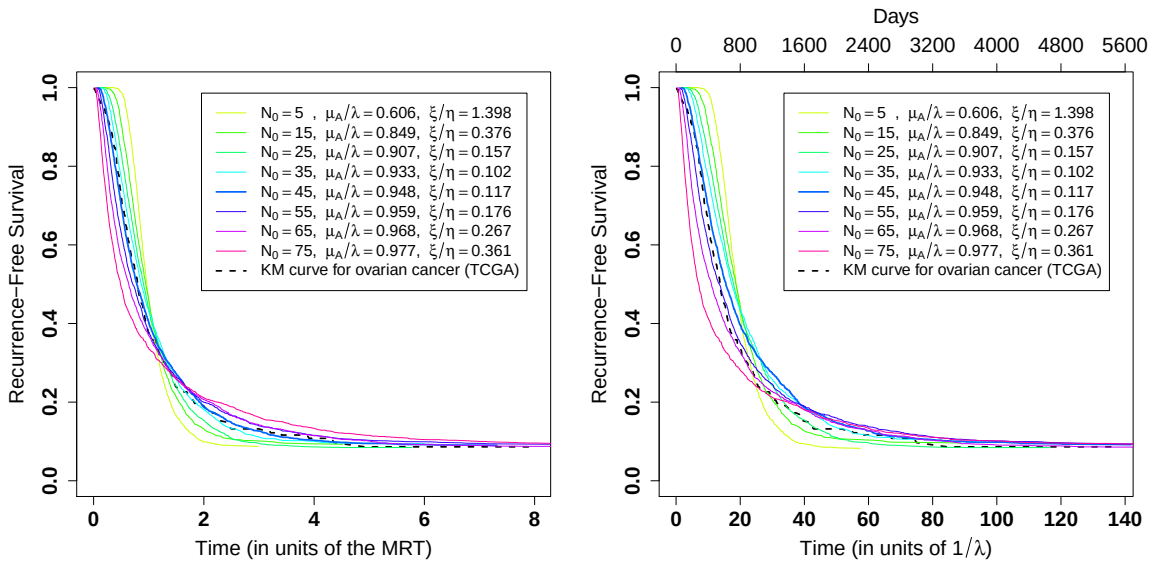


Figure 3.10: Product-limit estimate of the recurrence-free survival function (Kaplan-Meier curve [96]) for ovarian cancer in a group of 583 patients, obtained from TCGA data [89], showing Kaplan-Meier's estimate for the 90% confidence interval (green lines). The red vertical crosses (+) represent censored patients, whereas the black saltire crosses (×) represent recurrence events. The inset provides a closer view of the KM curve for the time interval between 0 and 500 days. Figure reproduced from [1].

In order to obtain the MRT from a given survival function $S(t)$, we first need to renormalize the probability of recurrence $p_{rec}(t) = 1 - S(t)$ as in Eq. (2.41), i.e., the appropriate probability measure is *conditioned* on recurrence. The renormalized survival function $\tilde{S}(t)$ is then given by

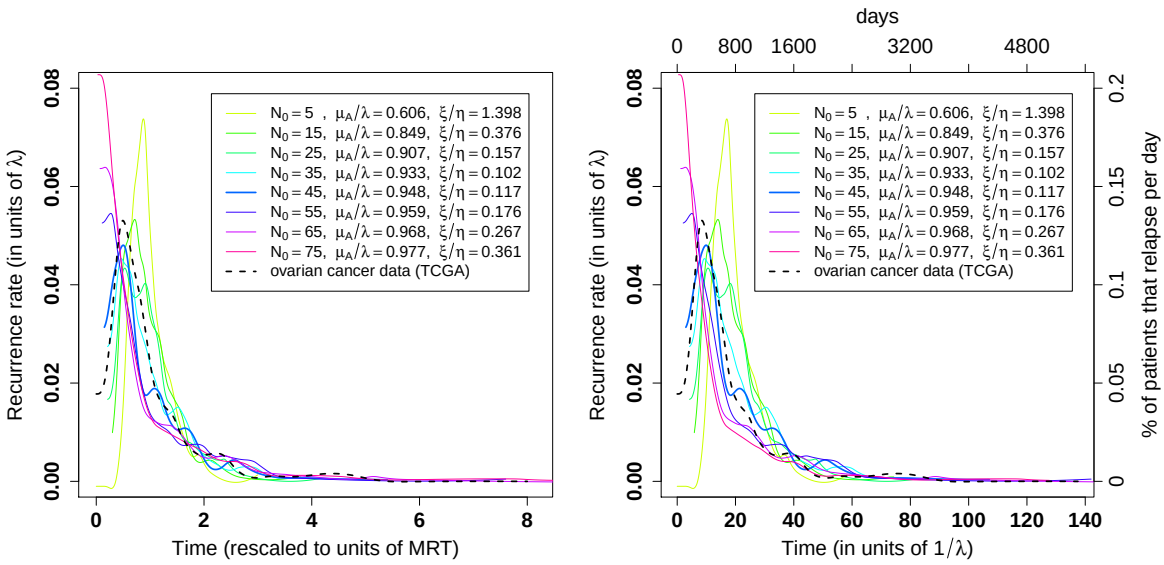
$$\tilde{S}(t) = \frac{S(t) - S(\infty)}{1 - S(\infty)}, \tag{3.32}$$

which vanishes in the limit $t \to \infty$. The MRT is then given by the area under the curve $\tilde{S}(t)$, i.e., $T_{rec}^{(1)} = \int_0^\infty \tilde{S}(t)dt$. For the serous ovarian cancer data from TCGA, we found $S(\infty) = 0.086$ (defined as the lowest value of the Kaplan-Meier estimate for $S(t)$) and an MRT of 687.5 days. For a given initial condition, estimation of the parameter $\nu \equiv \xi/\eta$ from Eqs. (3.15) and (3.16) requires that the time scale be fixed by specifying the doubling rate λ. Using the order of magnitude estimate $1/\lambda = 40$ days for the doubling time based on clinical data [97], the MRT for the ovarian cancer data in units of the doubling time was then fixed at $\lambda(687.5 \text{ days}) = 17.2$.



(a) (b)

Figure 3.11: Simulation fits of the QBD model to a Kaplan-Meier (KM) recurrence-free survival curve (dashed lines) for ovarian cancer data from The Cancer Genome Atlas [89]. Each simulated KM curve is from simulations for 2000 patients at the parameter values shown in the legends, with $N = 100$ and $m_0 = 0.4N_0$. The simulations were fit to the data using the scheme described in Section 3.5. For each curve, the value of μ_A/λ was obtained using Eqs. (3.3), (3.4) fitted to the large-time no-recurrence probability value $S(\infty) = 0.086$ obtained from the ovarian cancer data (Fig. 3.10). Given this value, ξ/η was obtained by matching $T_{rec}^{(1)}$ from Eqs. (3.15), (3.16) to the Mean Recurrence Time (MRT) value of 687.5 days from the ovarian data (see Fig. 3.10 and discussion in Section 3.5), using $1/\lambda = 40$ days. In (a) the time axis is rescaled to units of the MRT, and in (b) the time is shown in units of $1/\lambda = 40$ days on the lower axis and in days at the top. The blue curve for $N_0 = 45$ is shown in bold as the one whose shape is the best match to the clinical data. Figures reproduced from [1].



(a) With time axis in units of the MRT.

(b) With time axis in units of λ

Figure 3.12: Corresponding recurrence rates (negative time derivatives) of the curves from Fig. 3.11.

For a given choice of the initial number N_0, Eqs. (3.3) and (3.4) can be solved for the parameter μ_A/λ using the value $p_{rec}(\infty) = 1 - S(\infty) = 0.914$ obtained from the Kaplan-Meier curve. Assuming $m_0 = 0.4N_0$ (for which the outer solution gives a reasonable approximation of the MRT, unless N_0 is too small), it only remains to determine the value of the ratio $\nu \equiv \xi/\eta$ consistent with both the MRT (fixed at $T_{rec}^{(1)} = 17.2/\lambda$) and the inferred μ_A/λ. This is done by solving Eqs. (3.15) and (3.16) numerically for $\nu \equiv \xi/\eta$ (see also Eq. (2.2)).

Using the scheme described above to fix the parameters μ_A/λ and $\nu \equiv \xi/\eta$, recurrence-free survival curves were simulated for several initial values $N_0 = m_0 + n_0$, with $m_0 = 0.4N_0$ and $N = 100$, for 2000 patients. Each panel in Fig. 3.11 shows the Kaplan-Meier curves from these simulations matched to the ovarian cancer clinical recurrence data from TCGA. The fitted parameter values are shown in the legends in Fig. 3.11. For a given N_0, the value of μ_A/λ was fit to the clinical data using Eqs. (3.3), (3.4) and the large-time asymptotic value $S(\infty) = 1 - p_{rec}(\infty) = 0.086$. Using the fitted value of μ_A/λ, $\nu \equiv \xi/\eta$ was chosen to fit Eqs. (3.15), (3.16) with $T_{rec}^{(1)} \sim$ MRT(clinical), where the latter value was determined

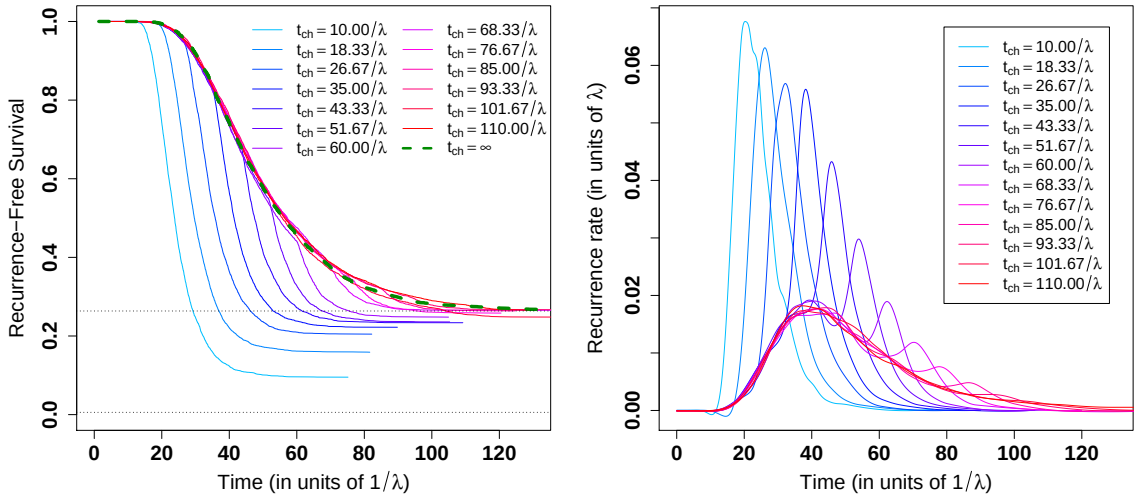
from the clinical data (from Fig. 3.10, MRT(clinical) = 687.5 days) and assuming $1/\lambda = 40$ days. The MRT for the simulation is the mean value of the recurrence time over the subset of the 2000 patients who had recurrence. The time axis can then be plotted in units of the MRT (Fig. 3.11a) or in units of $1/\lambda = 40$ days (Fig. 3.11b).

The discrepancy between Figs. 3.11a and 3.11b in the shapes of some of the simulated curves is because the parameter value ξ/η was chosen using equations Eqs. (3.15), (3.16), which are approximate to $O(1/N)$. This means that the MRT from the clinical data is only approximately matched to the MRT value from simulations using the approximation MRT(clinical) $\sim$ MRT(simulation) $\sim T_{rec}^{(1)}$. However, we see that the simulated curve that best matches the Kaplan-Meier curve ($N_0 = 45$) reproduces the recurrence rates quite well (see Fig. 3.12, where the negative time derivatives of the RFS curves were plotted both in units of MRT and in units of $1/\lambda$). The recurrence rates in Fig. 3.12 were calculated from smooth splines of the original curves with 15 degrees of freedom.

In the procedure described above for fitting the model to data, we have assumed a sharply peaked initial condition, i.e., the initial condition is a delta function centered at some specified initial number of tumor foci N_0. However, in reality this number should follow a probability distribution that would reflect the histogram of the residual-tumor size in the population under study. Moreover, different patients may have different responses to treatment (i.e., different values of μ_A/λ), as well as different ξ/η values. This means that the fitted values should be regarded as only a guide that gives insight into the possible scenarios leading to the observed recurrence-free survival curve.

3.6 Application of the model to chemotherapy simulation

The effect of chemotherapy on tumor recurrence can be simulated by switching the death rate μ_A at a time $t = t_{ch}$ (representing the end of chemotherapy) from a higher value μ_{ch}, which represents chemotherapy and lasts for times $t \leq t_{ch}$, to a baseline value $\mu_0 < \mu_{ch}$, which represents the absence of chemotherapy. Using this simple procedure, we can



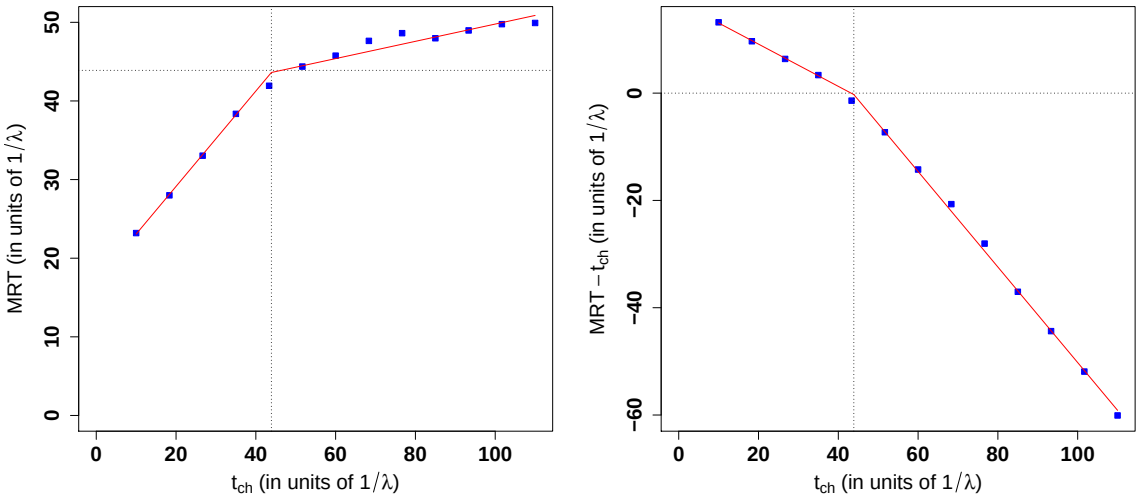
(a) (b)

Figure 3.13: Simulations of the model for $\mu_{ch}/\lambda = 0.8$, $\mu_0/\lambda = 0.4$, $\xi/\eta = 2.5$, $m_0 = n_0 = 3$ and $N = 100$, with the duration t_{ch} of chemotherapy varying from 10 to 110, in units of $1/\lambda$. (a) RFS curves obtained from simulating 4000 patients, where the horizontal line shows $S(\infty) = 1 - p_{rec}(\infty)$, the predicted asymptotic value of the large-time fraction of non-recurrent patients for $t_{ch} \to \infty$ (Eqs. (3.3), (3.4)) and the dashed green curve shows the simulated survival curve for $t_{ch} \to \infty$. Figure reproduced from [1]. (b) Recurrence rates (negative time derivatives) calculated from smooth splines of the original RFS curves with 15 degrees of freedom.

study how varying chemotherapy duration affects the MRT and the asymptotic (large time) fraction of non-recurring patients.

With $\mu_{ch} = 0.8\lambda$ and $\mu_0 = 0.4\lambda$ fixed, survival curves were simulated for t_{ch} varying from 10 to 110 (in units of $1/\lambda$) for $m_0 = n_0 = 3$, $N = 100$ and $\nu \equiv \xi/\eta = 2.5$ (Fig. 3.13). We see that increasing the duration t_{ch} of chemotherapy increases the time to recurrence (i.e., it increases the MRT) and also improves the asymptotic fraction of non-recurring patients (Fig. 3.13a). We also see (Fig. 3.14) that for these values of the parameters, although the MRT increases monotonically with t_{ch}, beyond $t_{ch} \approx 43/\lambda$ the rate of change of the MRT becomes lower than that of t_{ch}, suggesting that there is an optimum value of t_{ch} after which there is a reduction in benefit from chemotherapy. This optimum value of t_{ch} will most likely depend on all the parameters of the model: N_0, μ_{ch}, μ_0, ξ, η, and λ. We also note (Figs. 3.13a, 3.15) that the asymptotic fraction of non-recurrent patients in the simulation increases with t_{ch}, and eventually saturates at the maximum value $S(\infty) = 1 - p_{rec}(\infty)$ computed in the model, obtained using $\mu_A = \mu_{ch}$ in Eqs. (3.3), (3.4).

Recurrence can also be reduced by increasing dosage and keeping the chemotherapy time t_{ch} fixed. For $t_{ch} = 10/\lambda$ and $\mu_0 = 0.4\lambda$, survival curves were simulated for μ_{ch}/λ in the interval $1.4 - 1.6$ for $m_0 = n_0 = 3$, $N = 100$ and $\xi/\eta = 2.5$ (Fig. 3.16). We see that for these parameters, a time scale of 10 doubling times is enough for the increase in chemotherapy death rate on tumor foci to have a significant effect on long-term recurrence-free survival, even though the higher death rate μ_{ch} only acts during the finite period of time of a few doublings. Increasing μ_{ch} not only lowers recurrence rates, but also shifts them forward to later times (see 3.16b). It is worth mentioning that the benefit of changing only t_{ch} is limited by the asymptotic curve for $t_{ch} \to \infty$, in which case the large-time recurrence-free fraction is generally lower than 1 for $\mu_{ch} < \lambda$, and equal to 1 for $\mu_{ch} > \lambda$ (recall the phase transition given by Eq. (3.6)). This is to be contrasted with the effect of increasing μ_{ch} with t_{ch} fixed, in which case the long-term survival fraction can in theory be as large as desired, although in practice there is always a maximum dose tolerated by patients that limits this



(a) (b)

Figure 3.14: (a) Mean recurrence time (MRT) *vs.* chemotherapy duration t_{ch}, measured in units of $1/\lambda$. (b) Difference between the MRT and t_{ch} as a function of t_{ch} for the simulated survival curves of Fig. 3.13a. The vertical line on each plot shows the time point after which t_{ch} becomes larger than the MRT, i.e. the time point after which chemotherapy becomes less effective. Figures reproduced from [1].

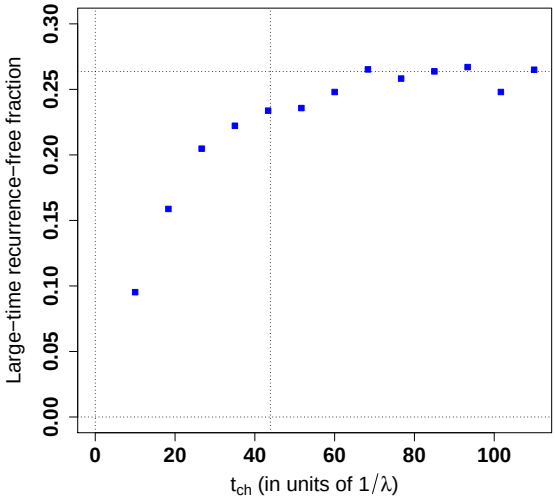
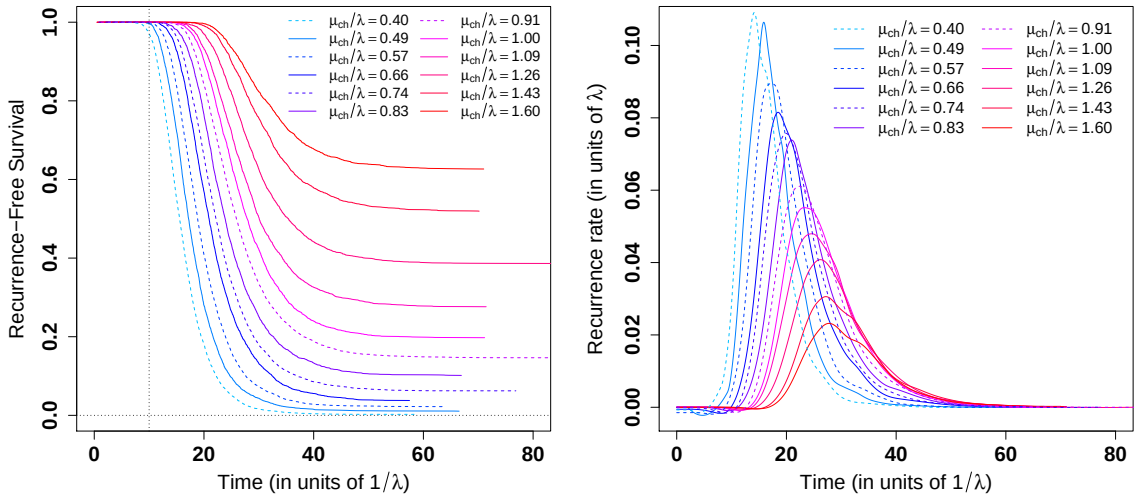


Figure 3.15: Asymptotic (large-time) non-recurrent fraction of patients versus chemotherapy duration t_{ch} for the simulated survival curves in Fig. 3.13a. The horizontal line shows the predicted asymptotic value of the non-recurrent fraction $S(\infty) = 1 - p_{rec}(\infty)$ for $t_{ch} \to \infty$, obtained theoretically by setting $\mu_A = \mu_{ch}$ in Eqs. (3.3), (3.4). Figure reproduced from [1].



(a) (b)

Figure 3.16: Simulations of the model for $t_{ch} = 10/\lambda$, $\mu_0 = 0.4\lambda$, $\xi/\eta = 2.5$, $m_0 = n_0 = 3$ and $N = 100$, with the chemotherapy death rate of the tumor (μ_{ch}) varying from 1.4λ to 1.6λ. (a) RFS curves (b) Recurrence rates (negative time derivatives) calculated from smooth splines of the original RFS curves with 15 degrees of freedom.

approach, since chemotherapeutic agents are highly toxic and also kill normal tissue, just not as efficiently as they attack rapidly growing tumors when the dose is moderate.

The version of the model presented in this section, with two additional parameters μ_{ch} and t_{ch} that define the chemotherapy regimen, will be used in Chapters 4 and 5, respectively, to fit the model to clinical trial data and make relevant predictions for different treatment scenarios.

Chapter 4

Fitting the model to data from an ovarian cancer neo-adjuvant trial with Olaparib

In this chapter, our goal is to fit the model proposed in Chapters 2 and 3 (specifically, the version introduced in Section 3.6) to the Kaplan-Meier (KM) curves from an ovarian cancer clinical trial [98], by finding a range of parameters for which simulated RFS curves agree with the data, which will then be used in Chapter 5 to predict the large-time RFS fraction for increased dosages of Olaparib (represented by the parameter μ_{ch}) and/or increased treatment times (represented by t_{ch}), within the relevant range of parameters determined from the fits.

4.1 Ovarian cancer clinical trial with Olaparib

A randomized, double-blind, phase 3 clinical trial has recently been conducted with a group of newly diagnosed, advanced high-grade ovarian cancer patients [98], to evaluate the efficacy of the drug olaparib as maintenance (adjuvant) therapy following standard treatment with cytoreductive surgery and platinum-based chemotherapy. This study will be the source of data for applying our model to predict progression-free survival in different scenarios of treatment.

The 391 patients who participated in the study had been diagnosed with stage III or IV high-grade serous or endometrioid ovarian cancer, primary peritoneal cancer, Fallopian-tube cancer or a combination of these cancers, and had a germline or somatic mutation in BRCA1, BRCA2, or both. The latter requirement is related to the mechanism of action

of olaparib, which is a poly(ADP-ribose) polymerase (PARP) inhibitor targeting tumor cells with deficiencies in homologous recombination repair. Inactivated PARP is trapped at single-strand DNA breaks by the inhibitor, which prevents the repair of these breaks, leading to an accumulation of double-strand breaks in replicating cells [99, 100]. The latter cannot be repaired accurately in tumor cells that have defects in homologous recombination repair mechanism (such as those harboring BRCA1 or BRCA2 mutations), which leads to genomic instability and ultimately, to cell death by apoptosis (if the p53 pathway is functional) or replication catastrophe.

Patients were also selected for the trial on the basis of their previous response to platinum-based chemotherapy, which was required to be complete or partial, with most patients having no evidence of disease and a normal level of the CA-125 biomarker protein after standard treatment. Of the 391 patients in the randomized trial, 260 received olaparib and 131 received placebo. In the olaparib cohort, the median duration of the intervention was 24.6 months. The recommended duration in the trial protocol was 2 years of olaparib, with dose interruption or dose reduction in the case of adverse events. In our analysis, we will assume that chemotherapy duration was fixed at 24 months.

At 3 years, the Kaplan-Meier estimate of progression-free survival was 60% in the olaparib cohort, versus 27% in the placebo cohort (control). The hazard ratio for disease progression or death, calculated using a Cox Proportional Hazards model, was 0.30, with a 95% confidence interval ranging from 0.23 to 0.41, and a P-value $P < 0.001$. The median time span of progression-free survival in the olaparib cohort was estimated to be about 3 years longer than in the placebo cohort. The conclusion of the study was that maintenance olaparib following standard treatment with cytoreductive surgery and platinum-based chemotherapy improved progression-free survival significantly for the population under study, for which the risk of disease progression or death of patients receiving olaparib was found to be 70% lower than for patients receiving placebo.

Months	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
No. at risk (Olaparib)	260	240	229	221	212	201	194	184	172	149	138	133	111	88	45	36	4	3	0
No. at risk (Placebo)	131	118	103	82	65	56	53	47	41	39	38	31	28	22	6	5	1	0	0

Table 4.1: Number of patients at risk at the beginning of each month since randomization, for the Olaparib and Placebo cohorts. Reproduced with permission from Ref. [98], Copyright Massachusetts Medical Society.

4.2 Setting up the data from the clinical study

As is unfortunately true for many clinical studies, the recurrence-free survival KM curves published in Ref. [98] for patients receiving placebo (control group) or the drug Olaparib were not accompanied by individual patient data (IPD) with recurrence events and times. However, a table specifying the number of patients at risk in one-month intervals is provided for each KM curve (see Table 4.1). This Table contains valuable additional information, because as we will show, when it is combined with data from the KM plots, it allows us to reconstruct IPD for both the placebo and Olaparib cohorts, which can then be used to compute errors and confidence intervals by the Kaplan-Meier approach [96] (for a note on this methodology, see Appendix C), thereby providing data to find allowable ranges for our parameters.

In order to reconstruct IPD from the available information, the web tool WebPlotDigitizer (https://automeris.io/WebPlotDigitizer/) was first used to accurately extract data points directly from the KM plots shown in Fig. 2A of Ref. [98]. With the data points from the published graphs at hand, as well as the data from Table 4.1, we then used the algorithm proposed in Ref. [101] to generate a table of IPD for each cohort (placebo and Olaparib), where a recurrence/no recurrence event and an event time is specified for each individual patient. The reconstruction algorithm is outlined in Appendix D, where the IPD tables for the placebo and Olaparib cohorts are also given.

Using the generated IPD data, we first verified that they generate exactly the same KM plot as the one extracted from the figure in Ref. [98] when we use the product-limit

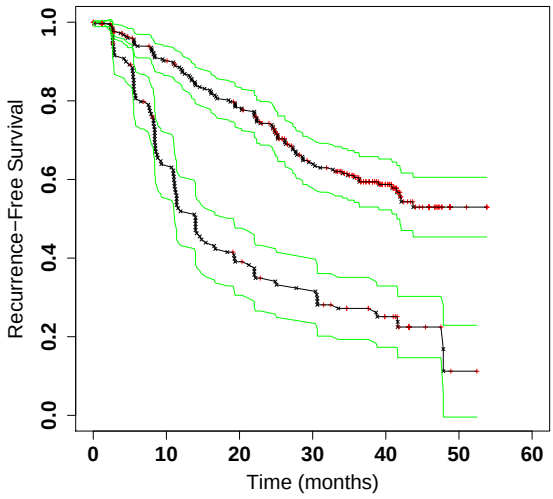


Figure 4.1: Recurrence-free survival Kaplan-Meier curves for the placebo (lower curve) and Olaparib (upper curve) cohorts of ovarian cancer patients in the study [98]. The red vertical crosses (+) represent censored patients, whereas the black saltire crosses (×) represent recurrence events. Each error band (green lines) corresponds to a 95% CI for the estimated survival probability at each recurrence time, and was calculated from reconstructed IPD using the Kaplan-Meier prescription (Eqs. (4.2) and (4.3)). The mean of each KM curve is reproduced here with permission from Ref. [98], Copyright Massachusetts Medical Society.

Kaplan-Meier formula to estimate the survival function. Both the placebo and Olaparib data sets passed this consistency check. According to the Kaplan-Meier prescription [96], given a sequence of event times (recurrence or censoring) $t'_1 \leq t'_2 \leq \cdots t'_N$, the recurrence-free survival function at time t can be estimated as

$$\hat{P}(t) = \prod_r \frac{N-r}{N-r+1}, \tag{4.1}$$

where the product is over all indices r such that each $t'_r \leq t$ is a recurrence time. The variance of the survival function can be estimated as

$$\hat{V}[\hat{P}(t)] = \hat{P}^2(t) \sum_r \frac{1}{(N-r)(N-r+1)}, \tag{4.2}$$

where, once again, the sum is over all indices r such that $t'_r \leq t$ and t'_r is a recurrence time.

Using Eq. (4.2) and assuming a Gaussian model for the error of $\hat{P}(t)$ at each given time, we can calculate the error corresponding to an α-confidence interval as

$$\text{Error}(\alpha; t) = \sqrt{\hat{V}[\hat{P}(t)]}\ \Phi^{-1}\left(\frac{1-\alpha}{2}\right), \tag{4.3}$$

where $\Phi^{-1}(q)$ is the quantile function for the standard normal distribution (i.e., the inverse of the standard normal CDF, giving the z-score value).

The reconstructed KM curves for the placebo (control) and Olaparib cohorts are shown in Fig. 4.1. These curves were plotted from the generated IPD using the Kaplan-Meier estimator (4.1), with the error band corresponding to a 95% CI given by Eqs. (4.2) and (4.3).

4.3 Fitting the model to the placebo KM curve

Our curve-fitting method starts by fixing the value of the parameter ξ/η, which represents the ratio of the transition rates at which tumor foci transition from active to dormant (ξ) and from dormant to active states (η). For each chosen value of this parameter, we determine the remaining parameters by fitting simulated RFS curves to both the placebo (control) and the Olaparib KM data.

For each fixed value of ξ/η, we used the placebo KM curve to determine the parameters λ, μ_0 and N_0. The remaining parameter μ_{ch} was then determined by fitting the model to the Olaparib KM curve. In order to fit our model to the placebo curve, we require a matching of: (A) large-time asymptotic values of the RFS fraction; (B) mean-recurrence times (MRT); (C) the shapes of the RFS curves, once the first two conditions have been met.

Before we proceed to fitting our model to the placebo curve, we need to do a preliminary curve fit (which is essentially a smoothing procedure) in order to define the asymptotic value of the KM curve, as well as its MRT, without the bias that originates from the number of patients at each given time for which the tumor has not yet relapsed. Because this number decreases over time, we see more variance in the KM estimator at later times, which means that earlier times should be given more weight in estimating the RFS fraction, proportionally to the time-dependent number of recurrence-free patients. For this preliminary smoothing

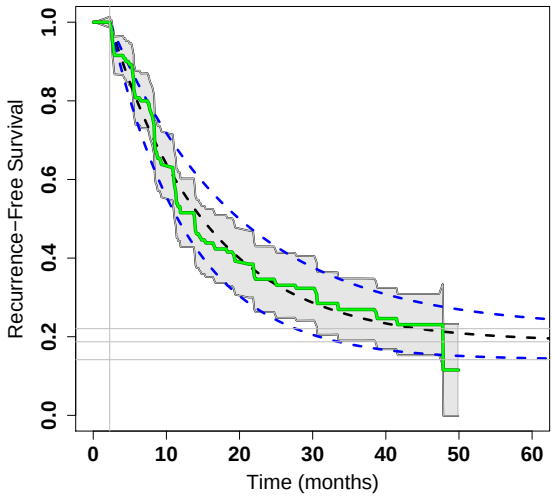


Figure 4.2: Shifted exponential fits for the mean and the 95% CI upper/lower limits of the placebo KM curve (see Eq. (4.4)), a smoothing step before the model is fitted to the placebo data.

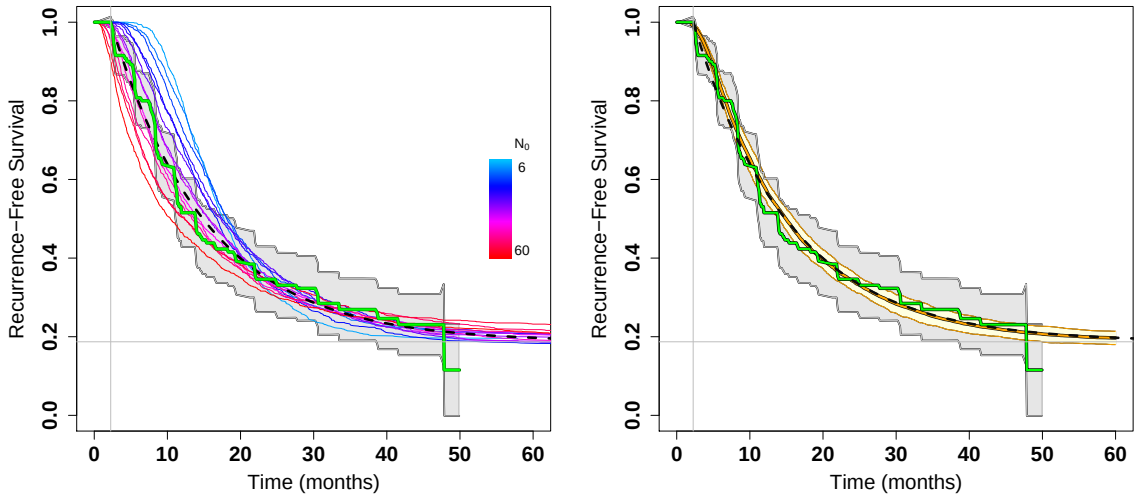
of the placebo KM curve, we choose the "shifted-exponential" function

$$\mathrm{RFS[placebo]}(t) = (1 - b_1)e^{-a_1(t-t_0)} + b_1 \ , \tag{4.4}$$

where t_0 is the point in time where the KM curve starts to drop from its initial value RFS $= 1$. Between times $t = 0$ and $t = t_0$, the value of the RFS function is constant, i.e., $\mathrm{RFS}(t) = 1$. For the placebo data at hand, we see that $t_0 = 2.25$ months. The parameters a_1 and b_1 are then determined by a weighted nonlinear regression with a diagonal weight matrix defined by $W_{ii} = \frac{\mathcal{N}_i}{\sum_i \mathcal{N}_i}$, where $\mathcal{N}_i$ is the number of recurrence-free patients in the placebo group at the i-th time point. By minimizing the weighted sum of squared residuals $\mathcal{S} = \sum_i W_{ii} r_i^2$, where r_i is the difference between the smooth curve (4.4) and the KM value at the i-th time point, we found $a_1 = 0.0757$/month and $b_1 = 0.187$. This means that an estimated 18.7% of patients receiving placebo were recurrence-free at large times.

For the smooth curve (4.4), the large-time asymptotic value is therefore given by the parameter

$$\mathrm{RFS[placebo]}(\infty) = b_1 = 0.187 \ , \tag{4.5}$$



(a) (b)

Figure 4.3: (a) Simulated placebo curves for $\xi/\eta = 1$, for several values of the initial condition N_0 and KM curve (green line) with the smooth shifted exponential fit given by Eq. (4.4) (dashed line). (b) Simulated curve of best fit, showing the 95% CI of the simulation.

whereas the mean recurrence time (MRT) is given by

$$\text{MRT[placebo]} = \int_0^\infty \frac{\text{RFS}(t) - \text{RFS}(\infty)}{1 - \text{RFS}(\infty)} dt = t_0 + 1/a_1 = 15.5 \text{ months}. \tag{4.6}$$

The smoothing procedure above was repeated for the curves that define the upper and lower limits of the 95% CI error band of the placebo KM curve. The shifted exponential fits (Eq. (4.4)) for all three curves (mean, lower and upper 95% CI limits) are shown in Fig. 4.2.

The next step in our analysis was to fix an initial condition N_0 and then determine the corresponding value of the ratio μ_0/λ by requiring that the large-time recurrence-free fraction (4.5) matches the theoretical large-time probability of no-recurrence $p_{cure}(\infty) = 1 - p_{rec}(\infty)$, where $p_{rec}(\infty)$ is given by Eqs. (3.3) and (3.4). For each fixed N_0, we can immediately solve for the ratio μ_0/λ by setting $p_{rec}(\infty) = 1 - b_1 = 0.813$. Once the ratio μ_0/λ is determined, we can determine the doubling rate λ by requiring that the MRT given by Eq. (4.6) matches the theoretical value given by Eqs. (3.15) and (3.16).

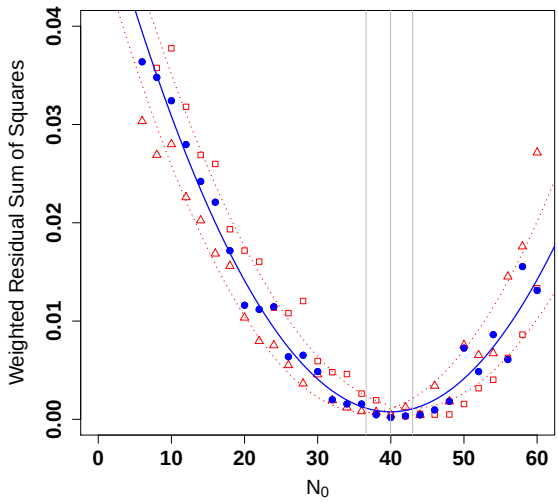
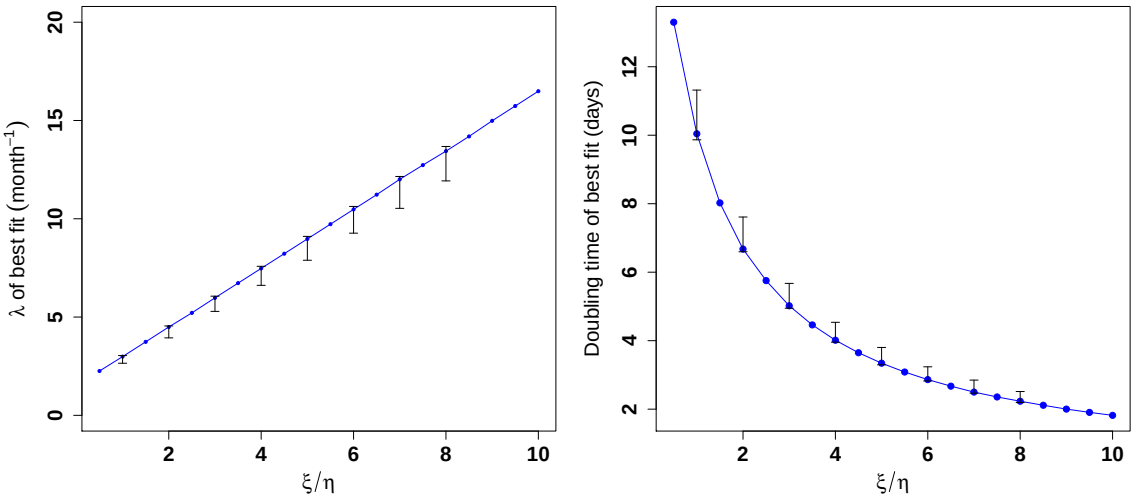


Figure 4.4: Weighted Residual Sum of Squares (RSS) for $\xi/\eta = 1$ (blue dots) fitted to a quadratic function. The red square- and triangle-shaped dots represent the weighted RSS for the lower and upper limits of the 95% CI error band.

For each choice of N_0, once the parameters λ and μ_0 are determined as described above, an RFS placebo curve can be simulated for each initial condition N_0. For $\xi/\eta = 1$, RFS curves simulated with $N = 100$ for several initial conditions N_0, as well as the placebo KM curve and the smooth curve fit given by Eq. (4.4), are all shown in one plot in Fig. 4.3a. For the specific choice of transition-ratio parameter $\xi/\eta = 1$, we see that the weighted residual sum of squares (RSS) between the simulated RFS curve and the smoothed KM placebo curve (with weights given by the fraction at risk in the simulation) has a minimum at around $N_0 = 40$, as shown in Fig. 4.4. This optimal initial condition is where the simulated RFS curves are found to best fit the overall *shape* of the KM placebo curve (see Fig. 4.3b). The location of the weighted RSS minimum is defined by fitting the RSS curve to a quadratic function. This process can be repeated for the upper and lower limits of the 95% CI error band (see Fig. 4.4), which determines an error interval around $N_0 = 40$. For $\xi/\eta = 1$, we find that N_0 can be as low as 37 and as high as 43. For our purposes, we will only consider even values of N_0, because we define the initial condition as $m_0 = n_0 = N_0/2$, i.e., half the initial tumor foci are initially dormant and half of them are initially active.



(a) (b)

Figure 4.5: (a) Relationship between the doubling rate λ and ξ/η imposed by the requirement that the MRT of the simulated placebo curves has to match the MRT of the KM curve. (b) Doubling time $(1/\lambda)$ in days against ξ/η.

The method described above can be repeated for an entire range of values of the transition-ratio parameter ξ/η, such that the optimal values of λ, μ_0 and N_0 all become functions of ξ/η. However, the values of μ_0 and N_0 turn out to be essentially constant (independent of ξ/η), as shown in Table 4.2. In the case of λ, the requirement that the MRT of each simulated placebo curve matches the MRT of the KM curve means that the time scale defined by λ has to be reajusted as a function of ξ/η, as expected from the approximately linear relationship between the MRT in $1/\lambda$ units and ξ/η (see Fig. 3.6 in Chapter 3). The relationship between λ and ξ/η for the fits is shown in Fig. 4.5, where the doubling time $1/\lambda$ in days is also shown as function of ξ/η. The error bars were calculated by repeating the fitting procedure for the lower and upper limits of the 95% CI error band of the placebo curve.

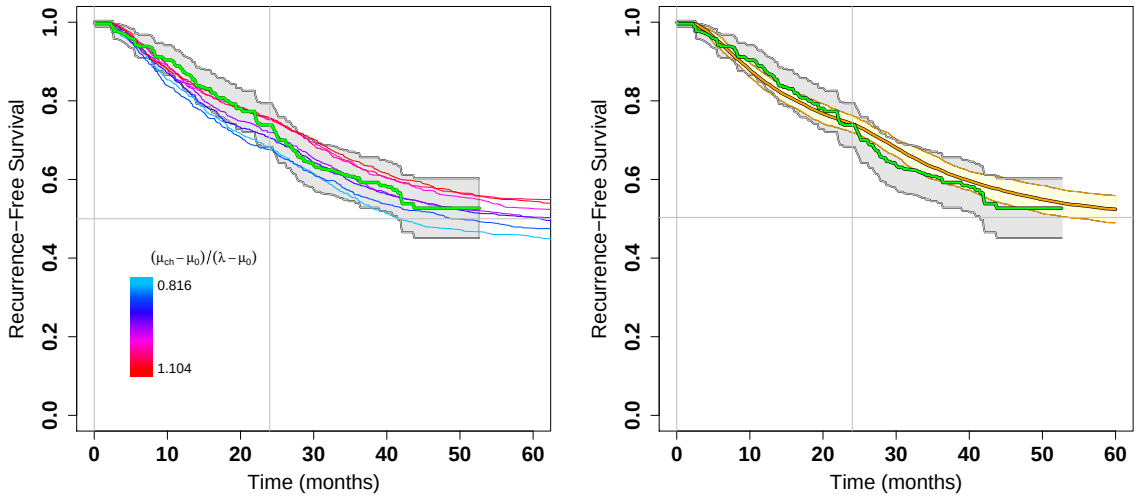
Parameter		Values							
ξ/η		1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
N_0	min	37	36	37	37	36	37	36	37
	mean	40	40	39	39	39	39	40	38
	max	43	42	42	43	42	42	43	42
μ_0/λ	min	0.957	0.956	0.956	0.957	0.956	0.956	0.957	0.956
	mean	0.961	0.961	0.960	0.960	0.960	0.960	0.961	0.958
	max	0.962	0.960	0.962	0.962	0.960	0.962	0.960	0.962
λ (month^{-1})	min	2.65	3.94	5.29	6.61	7.89	9.27	10.53	11.93
	mean	3.00	4.49	5.98	7.47	8.97	10.48	12.01	13.44
	max	3.04	4.55	6.06	7.58	9.10	10.63	12.15	13.68
$(\mu_{ch}-\mu_0)/(\lambda-\mu_0)$	min	0.816	0.816	0.816	0.816	0.864	0.864	0.816	0.768
	mean	1.017	1.020	1.048	1.027	1.007	1.034	1.018	1.055
	max	1.104	1.200	1.200	1.152	1.152	1.152	1.152	1.200

Table 4.2: Ranges of the fitting parameters for a few values of ξ/η.

4.4 Fitting the model to the Olaparib KM curve

Since all patients in the clinical trial were randomized into the Olaparib or placebo cohorts, it is a reasonable assumption that the parameters λ, μ_0 and N_0 must be the same for both groups (otherwise, the placebo cohort would not be a properly designed control). For every fixed value of the ratio ξ/η, the parameters λ, μ_0 and N_0, which have already been determined from fitting the model in the absence of chemotherapy to the placebo KM curve, will be used to simulate RFS curves for the Olaparib cohort. The tumor death-rate parameter μ_A for patients receiving the drug was initially set to a chemotherapy rate $\mu_{ch} > \mu_0$, and then switched to the baseline level $\mu_A = \mu_0$ after a time $t_{ch} = 24$ months for which the patients were given the drug. In this section, our goal will be to simulate RFS curves for different values of the parameter μ_{ch} and eventually find the value for which the RFS curve best fits the Olaparib KM curve. In principle, since the unknown parameter ξ/η is arbitrarily chosen at the very beginning, the value of μ_{ch} determined from the fit will be a function of ξ/η.

Because the doubling rate λ and the ratio ξ/η are related through the relationship shown in Fig. 4.5, an equivalent statement to our explanation above is that for every fixed value



(a) (b)

Figure 4.6: (a) Simulated RFS curves for $\xi/\eta = 1$, with different values of μ_{ch} (color scale). Only curves that essentially fall within the 95% CI of the Olaparib KM curve (shaded area) are shown. (b) RFS curve that best fits the KM Olaparib curve according to our criterion for goodness of fit, with the estimated 95% CI simulation error band. The corresponding chemotherapy parameter is $(\mu_{ch} - \mu_0) / (\lambda - \mu_0) = 1.018$.

of λ, all the other parameters in the model are determined from the fits. Essentially, λ or ξ/η define the time scale for the model according to the prescription in Fig. 4.5.

For each choice of ξ/η, RFS curves were simulated for a range of values of the parameter μ_{ch}. The next step will be to define a metric for the goodness of fit between our simulated curves and the Olaparib KM curve. Minimizing such a metric will determine the optimal value of μ_{ch}. To illustrate our procedure, we choose $\xi/\eta = 1$, for which simulated curves are plotted with the Olaparib KM data in Fig. 4.6a.

For the purpose of defining a parameter that specifies the potential efficacy of a drug (in this case, Olaparib), we will henceforth present our results in terms of the difference between the chemotherapy death rate μ_{ch} and the baseline death rate μ_0, relative to the difference between μ_0 and the critical value in the absence of chemotherapy, i.e., $\mu_{critical} = \lambda$. In other words, our reference will be $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$, which is related to the chemotherapy dose through some unknown monotonically increasing function that we hope could be determined

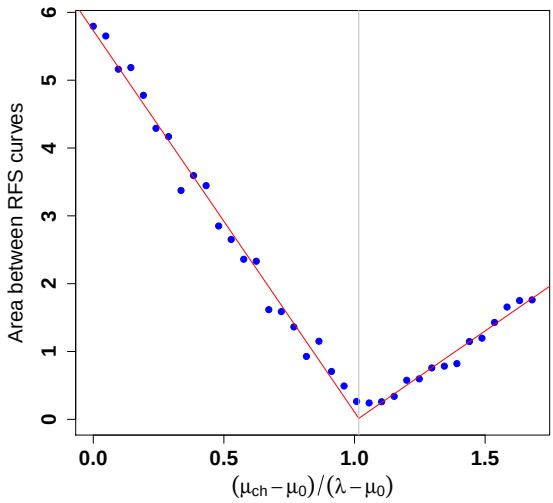


Figure 4.7: Area between the simulated RFS curves and the Olaparib KM curve within the time interval from $t = 0$ to $t_{ch} = 24$ months (i.e., within the duration of chemotherapy), for $\xi/\eta = 1$. The optimal value of the chemotherapy parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ is defined to be at the intersection of the fitted lines.

in clinical studies. Note that this parameter defines a new scale, where zero means absence of chemotherapy, whereas unity means that the chemotherapy death rate is at the critical value $\mu_{critical} = \lambda$.

Two relevant criteria need to be defined for estimating μ_{ch}, or equivalently, our new parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$: one that determines the value of the best fit and the other that determines an interval around the optimal value, reflecting the uncertainty in the Olaparib KM curve as defined by its 95% CI error band. For the best value of μ_{ch}, we simply choose the criterion that the area between the simulated RFS curves and the Olaparib KM curve within the time interval from $t = 0$ to $t_{ch} = 24$ months (i.e., within the duration of chemotherapy) has to be minimized. For $\xi/\eta = 1$, the area between the simulated curves and the KM curve is plotted against the parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ in Fig. 4.7. For each fixed value of ξ/η, the minimum of the area function was determined by dividing the set of points in its graph into two regions and performing a linear regression in each region. The optimal value of μ_{ch} is then determined as the point where the two lines intersect, as shown in Fig. 4.7. For $\xi/\eta = 1$, we find $(\mu_{ch} - \mu_0) / (\lambda - \mu_0) = 1.018$. For a range of values

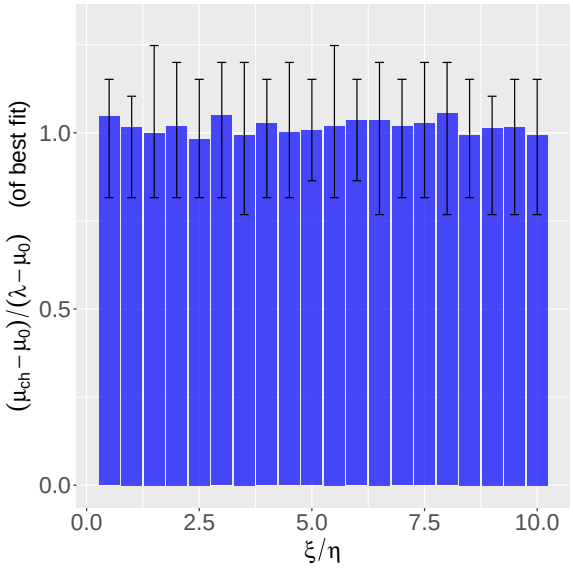
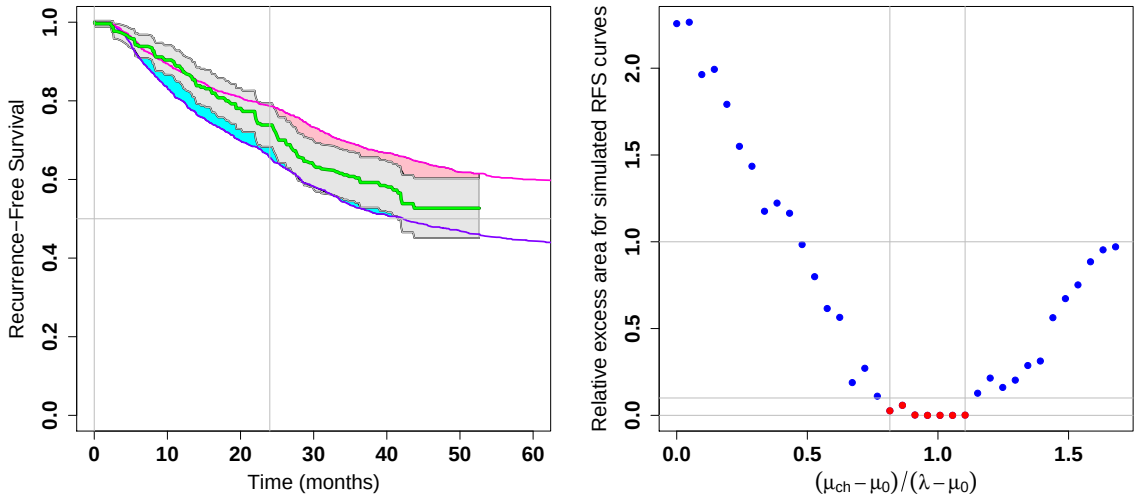


Figure 4.8: Estimated values of the parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ for different choices of ξ/η and corresponding error bars due to the uncertainty in the Olaparib KM curve.

of ξ/η, the optimal values of the parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ are shown in Table 4.2, and as a bar plot in Fig. 4.8.

The next step in the procedure was to define a range of values around the optimal value of μ_{ch} that reflects the uncertainty of the Olaparib KM curve. For this purpose, we need to define a metric for how well the simulated curves fall within the error band of the Olaparib KM curve, defined as the 95% CI error band. To illustrate our criterion, the reader is referred to Fig. 4.9a, where for two different simulated RFS curves, the *excess area* between each curve and the Olaparib KM curve error band has been shaded with different colors. Such excess area might exist below the lower 95% CI limit (as is the case with the blue curve shown in the figure) or above the upper 95% CI limit (as with the pink curve). More generally, a curve could in principle have contributions to the excess area both from below and above the error band. Our criterion for μ_{ch} to be within the range of values that are consistent with the KM curve is that the excess area for the simulated RFS curve cannot exceed 10% of the area between the upper and lower limits of the KM curve 95% CI error band, within the time range of the clinical trial (i.e., between times $t = 0$ and $t = 52.6$ months). In Fig. 4.9b, where the excess area is plotted against $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ for



(a) (b)

Figure 4.9: (a) Definition of excess area between simulated RFS curves and the Olaparib KM curve error band. (b) Relative excess area (defined as a fraction of the area between the upper and lower limits of the 95% CI error band) versus $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ for $\xi/\eta = 1$. The points that meet the error criterion of a relative excess area lower than 10% are highlighted in red.

$\xi/\eta = 1$, the points that meet our criterion form a plateau of minimal (if not zero) excess area and are highlighted in red.

The intervals for $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$, calculated using the criterion described above, are shown in Table 4.2 and in the bar plot of Fig. 4.8 for different values of the scale parameter ξ/η. Note that the optimal values from the fitting are essentially constant across different values of ξ/η and are close to the critical value (= 1). Essentially, the error bars show that for the given data, μ_{ch} is equally likely to be on either side of the critical value $\mu_{critical} = \lambda$, which we only take here as a reference, since a phase transition only exists in the limit $t_{ch} \to \infty$.

The parameters determined from the fitting procedure are summarized in Table 4.2 and the best fits for $\xi/\eta = 1$ are shown in Figs. 4.3b and 4.6b with their respective simulation error bands (defined here as a Gaussian 95% CI) determined from averaging over 10 sets of simulations, each of which consisted of 1000 simulation runs.

Chapter 5

Predicting Recurrence-Free Survival as a function of Olaparib dosing and scheduling

In this chapter, we will use our model to predict different scenarios of treatment outcome for the patients in the Olaparib clinical trial. Our main result is a prediction of the long-term recurrence-free survival (RFS) when either the duration or the dosage of Olaparib chemotherapy is increased. We will also discuss different scenarios in which the therapy may be less or more effective, namely, when the ratio of transition rates to and from dormancy (ξ/η parameter) is increased or decreased. Using the parameters determined in Chapter 4 by fitting the model to the Olaparib clinical trial data, we will also discuss briefly some of the tumor-growth dynamics that the model predicts for the patients who participated in the Olaparib study, or for any patient with a higher or lower active/dormant transition parameter ξ/η being treated with the same drug.

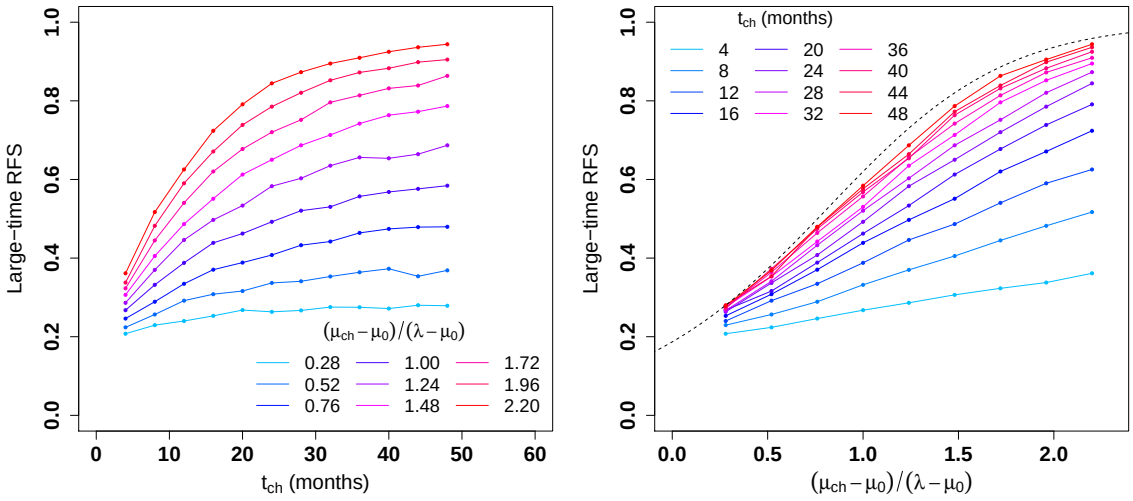
5.1 Predictions for long-term recurrence-free survival of patients in the Olaparib clinical trial

In Chapter 4, our model for chemotherapy was fitted to KM curves from a clinical trial whose goal was to identify the potential benefit of adjuvant therapy using the drug olaparib for advanced high grade ovarian cancer patients. All the parameters of the model were determined as functions of the ratio of transition rates to and from dormancy (ξ/η parameter). We found that this parameter meets the fitting criteria if it is related to the doubling rate λ through the functional relationship shown in Fig. 4.5a, because ξ/η and λ

have opposite effects on the time to recurrence, which cancel each other out when both parameters are constrained as in Fig. 4.5a. This means that we may determine all the fitting parameters equivalently as functions of either the doubling rate λ or the ratio ξ/η. Essentially, fixing λ defines a relationship between the natural time scale of the model (measured in units of the doubling time $1/\lambda$) and the time scale of the clinical KM curves (defined in months). Although this relationship remains undetermined, there are biomarkers that can help clinicians estimate the parameter λ for ovarian cancer patients, such as the cancer antigen CA125 [97]. Unfortunately, this information was not available for the patients in the Olaparib study.

Remarkably, as will be shown, our predictions for large-time RFS for the patients in the Olaparib trial turn out to be essentially independent of our choice of the time scale $1/\lambda$. Specifically, it will be shown that the dependence of the large-time RFS on the chemotherapy parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ and on the treatment time t_{ch} (measured in months) is independent of the choice of λ in the fitting procedure (or, equivalently, of the corresponding value of ξ/η given by the relationship shown in Fig. 4.5a). As will be discussed in Section 5.2, this does not mean that the large-time RFS does not depend on the parameters λ and ξ/η individually; it simply means that if λ and ξ/η are constrained as in Fig. 4.5a, the predicted large-time RFS turns out to be independent of the choice of these parameters.

For the patients in the Olaparib study, an important question is how their recurrence-free survival probability would change, had they been treated for a longer time t_{ch} or had they received a higher dose of Olaparib, which is related to our model parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ by some unknown monotonically increasing function. To answer this question, RFS curves were simulated on a 9×12 grid of points in the (μ_{ch}, t_{ch}) plane (5000 tumors for each pair of parameters), with all the other parameters fixed by the fits discussed in Chapter 4. To find the large-time RFS fraction reliably, the tumor size distribution $\mathcal{P}_k(t) = \Pr\{m+n=k; t\}$ was recorded at time $t = t_{ch}$ and the asymptotic RFS values were



(a) (b)

Figure 5.1: (a) Large-time RFS versus t_{ch} for several values of μ_{ch}, with $\xi/\eta = 1$. (b) Large-time RFS versus μ_{ch} for several values of t_{ch}, with $\xi/\eta = 1$. The dashed line is the predicted asymptotic curve for $t_{ch} \to \infty$.

estimated using the formula

$$\mathrm{RFS}(\infty) = \mathrm{RFS}(t_{ch}) - \sum_{k=1}^{N-1} \mathcal{P}_k(t_{ch}) \frac{1 - e^{-2R_0 k}}{1 - e^{-2R_0 N}}, \tag{5.1}$$

where $R_0 = (1 - \mu_0/\lambda)/(1 + \mu_0/\lambda)$.

The usefulness of Eq. (5.1) comes from the fact that the simulations are only reliable within a limited time scale, so estimating the large-time RFS from the RFS value at a finite time t_{ch} is more accurate than defining RFS(∞) to be the lowest value obtained in the simulation.

To understand Eq. (5.1), we first note that the probability of recurrence can be decomposed into the probability of recurrence before or up to time t_{ch}, and the probability of recurrence at any time after t_{ch}, starting from any of the possible tumor sizes $k = 1, \cdots, N-1$ for a patient at risk at time t_{ch}:

$$p_{rec}(\infty) = p_{rec}(t_{ch}) + \sum_{k=1}^{N-1} \mathcal{P}_k(t_{ch}) \Pr\{\text{recurrence} \,|\, m + n = k\} \tag{5.2}$$

Since the tumor is acted upon by a single death rate μ_0 after time t_{ch}, we can use our

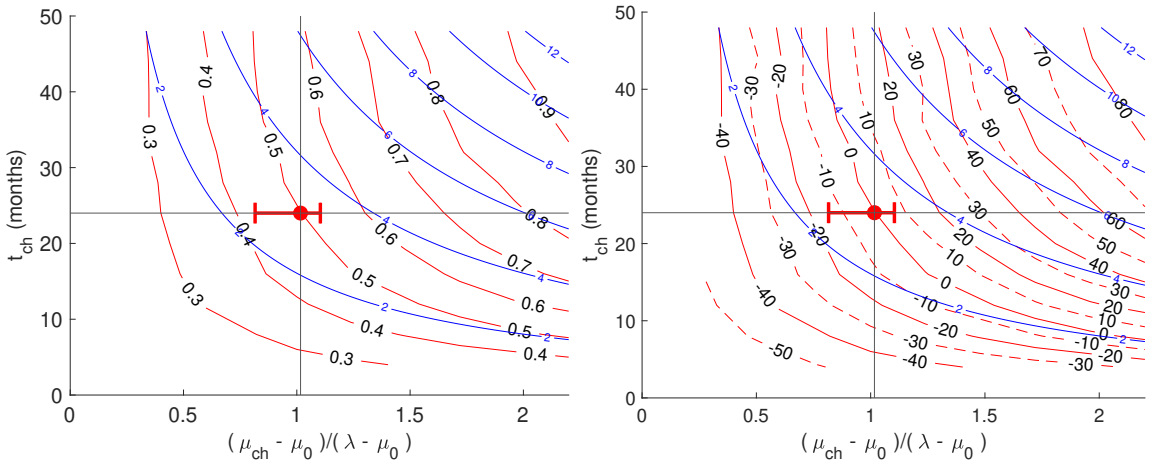


Figure 5.2: (a) Level curves of the large-time RFS fraction in the plane $(\frac{\mu_{ch}-\mu_0}{\lambda-\mu_0}, t_{ch})$, for $\xi/\eta = 1$ (red lines). (b) Corresponding percent changes relative to the best fit for the patients in the Olaparib trial (red lines), for which $t_{ch} = 24$ months and $(\mu_{ch} - \mu_0) / (\lambda - \mu_0) = 1.018$ (red circle shown in the figures, with the error bar for the chemotherapy parameter). The blue lines are the curves along which $(\frac{\mu_{ch}-\mu_0}{\lambda-\mu_0})\lambda t_{ch} = constant$.

result given by Eq. (3.3) to write $\Pr\{\text{recurrence} \,|\, m+n=k\} = (1-e^{-2R_0 k})/(1-e^{-2R_0 N})$, with $R_0 = (1-\mu_0/\lambda)/(1+\mu_0/\lambda)$, and since $\text{RFS}(t) = 1 - p_{rec}(t)$, we arrive at Eq. (5.1).

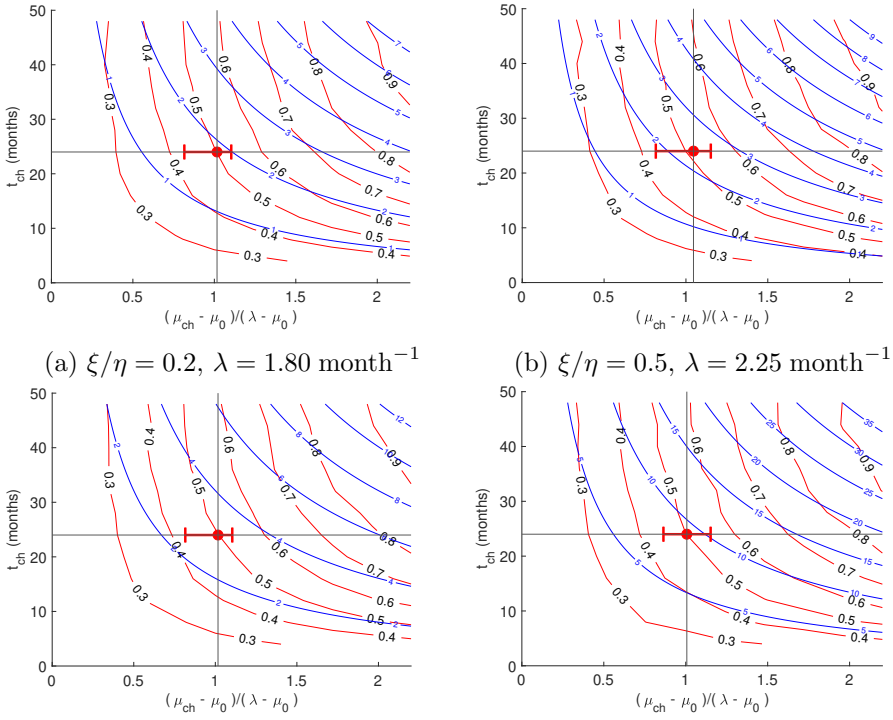
For $\xi/\eta = 1$, the large-time RFS is plotted against t_{ch} for several values of μ_{ch}, and against μ_{ch} for several values of t_{ch} in Fig. 5.1. As already noted in Section 3.6, increasing t_{ch} at a fixed death rate μ_{ch} can only improve large-time RFS up to some fixed value that depends on μ_{ch}, i.e., for fixed μ_{ch} there is an upper bound to the improvement in recurrence free survival as $t_{ch} \to \infty$. This upper bound is represented by the dashed curve in Fig. 5.1b. Also note that for a fixed value of t_{ch} (i.e. if chemotherapy is extended by a fixed amount of time), then the higher the value of μ_{ch}, the higher the predicted large-time RFS fraction. In contrast, if μ_{ch} is increased by a fixed amount, then the longer the duration of chemotherapy, the lower the relative increase in the RFS fraction, due to saturation toward the upper bound set by the limiting case $t_{ch} \to \infty$. It is not clear how much an increase in the value of $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ would be high enough that the Olaparib dose would be too toxic to be tolerated by patients. Knowledge of the relationship between our parameter $(\mu_{ch} - \mu_0) / (\lambda - \mu_0)$ and dose, as well as the dose toxicity limit, are necessary to allow our

t_{ch} (months)	$(\mu_{ch}-\mu_0)/(\lambda-\mu_0)$								
	0.28	0.52	0.76	1.00	1.24	1.48	1.72	1.96	2.20
4	–57.8	–54.5	–50.0	–45.7	–41.9	–37.7	–34.4	–31.4	–26.6
8	–53.4	–47.9	–41.3	–32.6	–24.8	–17.7	–9.6	–2.1	5.0
12	–53.1	–40.7	–32.0	–21.2	–9.4	–1.2	9.8	19.9	27.0
16	–48.6	–37.4	–24.7	–10.9	1.0	11.9	26.0	36.3	47.0
20	–45.6	–35.8	–21.1	–6.1	8.4	24.4	37.6	50.1	60.7
24	–46.5	–31.6	–17.2	0.0	18.4	32.1	46.3	59.5	71.5
28	–45.8	–30.7	–12.1	5.7	22.5	39.5	52.7	66.7	77.3
32	–44.1	–28.3	–10.2	7.7	29.0	44.9	61.7	73.1	81.8
36	–44.2	–26.1	–5.7	13.1	33.3	50.7	65.3	77.2	84.7
40	–44.8	–24.3	–3.7	15.4	32.8	55.1	68.9	79.3	87.8
44	–43.1	–28.2	–2.8	17.0	34.9	56.9	70.4	82.5	90.1
48	–43.4	–25.1	–2.6	18.7	39.5	59.8	75.4	83.8	91.7

Table 5.1: RFS percent changes for $\xi/\eta = 1$, relative to $t_{ch} = 24$ months and $(\mu_{ch}-\mu_0)/(\lambda-\mu_0) = 1.00$ (which is close to the parameter value of 1.018 for the Olaparib trial).

model to determine the best possible treatment outcome.

For $\xi/\eta = 1$, level curves of the large-time RFS fraction are plotted in the plane $(\frac{\mu_{ch}-\mu_0}{\lambda-\mu_0}, t_{ch})$ in Fig. 5.2a. For the patients of the Olaparib trial, $t_{ch} = 24$ months and from the fit with $\xi/\eta = 1$ in Chapter 4, we found $(\mu_{ch}-\mu_0)/(\lambda-\mu_0) = 1.018$. Relatively to these patients, Fig. 5.2b shows level curves for the percent changes in large-time RFS. Although the relation between the parameter $(\mu_{ch}-\mu_0)/(\lambda-\mu_0)$ and the dose of chemotherapy (or more specifically, drug concentration in the bloodstream) is unknown, if we take this parameter to represent dose, Fig. 5.2 shows that it is possible to extend chemotherapy to longer periods of time and simultaneously reduce the dose, while keeping the predicted long-term RFS unchanged. For the patients in the Olaparib study, this means following the 0% change level curve in Fig. 5.2b. However, in order to increase RFS by extending the duration of chemotherapy, the *total dose* accumulated during the time t_{ch}, which is represented here by the product $(\mu_{ch}-\mu_0)t_{ch}$, always needs to be increased. This is because the large-time RFS always increases along curves of constant total dose, $(\frac{\mu_{ch}-\mu_0}{\lambda-\mu_0})\lambda t_{ch} = constant$, if such curves are followed toward larger values of t_{ch}. The problem with extending this reasoning naively is that increasing total dose is likely to have significant side effects on the patients by exposing their normal tissues to the drug beyond



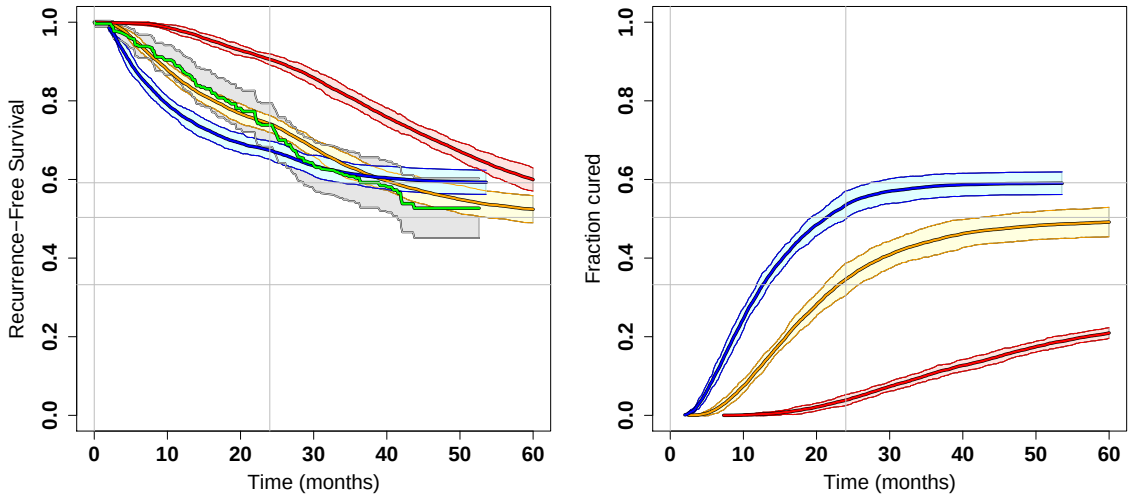
(a) $\xi/\eta = 0.2$, $\lambda = 1.80$ month^{-1} (b) $\xi/\eta = 0.5$, $\lambda = 2.25$ month^{-1}

(c) $\xi/\eta = 1$, $\lambda = 2.99$ month^{-1} (d) $\xi/\eta = 5$, $\lambda = 8.95$ month^{-1}

Figure 5.3: RFS level curves for different choices of fitting parameters ξ/η and λ (in red).

what they can tolerate. Thus, we need to establish a medically determined toxicity limit for total dose and relate it to our parameters before we can use our analysis to treat real patients in the clinic.

For the patients of the Olaparib trial, the model predicts an approximate 13% increase in RFS by extending chemotherapy from 24 to 36 months without changing the dose, and about a 19% improvement by extending to 48 months. Table 5.1 shows the percent changes in RFS relative to $t_{ch} = 24$ months and $(\mu_{ch} - \mu_0)/(\lambda - \mu_0) = 1.00$, which is close to the reference value of 1.018 for the Olaparib trial. If treatment is extended beyond 48 months, there is no longer a benefit, because the RFS saturates to the upper bound level represented by the dashed line in Fig. 5.1b, so that the level curves become essentially vertical at large times t_{ch}. Significantly higher increases in RFS are predicted for the patients in the Olaparib trial when the dose is increased for a fixed time duration, or if both the dose and the time are increased, as shown in Table 5.1. However, as noted, we cannot establish what the



(a) Recurrence-free survival

(b) Fraction cured

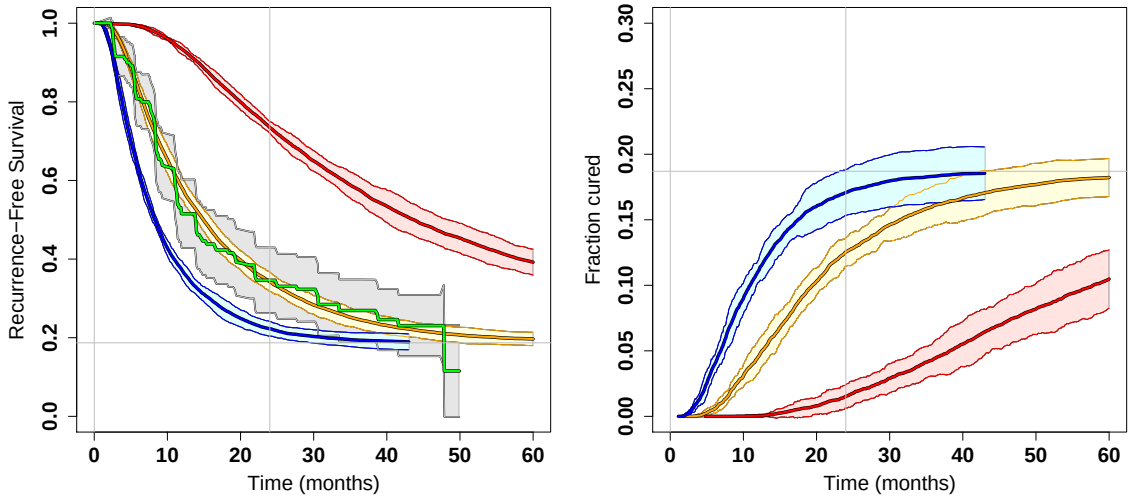
Figure 5.4: Olaparib curves for $\xi/\eta = 0.12$ (blue), $\xi/\eta = 1$ (yellow) and $\xi/\eta = 5$ (red), with their respective 95% CI error bands for the simulations. On the left, the Olaparib KM curve (green) is also plotted with its 95% CI error band. The horizontal lines represent the large-time asymptotic values for the three curves.

best predicted treatment strategy would be, unless the toxicity limit for the parameter $(\mu_{ch} - \mu_0)/(\lambda - \mu_0)$ (which is related to dose), as well as the toxicity limit for the product $(\mu_{ch} - \mu_0)t_{ch}$ (which is related to the total dose accumulated within the time t_{ch}) are known.

Although the results discussed above refer to the parameter choice $\xi/\eta = 1$ and $\lambda = 2.99$ month^{-1}, as was mentioned previously, they turn out to be independent of the choice of the time scale λ for the fitting. Indeed, we see in Fig. 5.3 that the RFS level curves for the different parameter choices $\xi/\eta = 0.2$, $\xi/\eta = 0.5$, $\xi/\eta = 5$ and $\xi/\eta = 9$, with λ given by the relation shown in Fig. 4.5a, are all very similar.

5.2 Predicting the effect of dormancy on recurrence and cure rates of patients treated with Olaparib

In Section 5.1, we noted that the large-time RFS for the Olaparib trial is independent of the choice of parameters λ and ξ/η for the fitting, as long as these parameters are related as in Fig. 4.5a. For $\lambda = 2.99$ month^{-1}, which corresponds to a doubling time of 10.0 days,



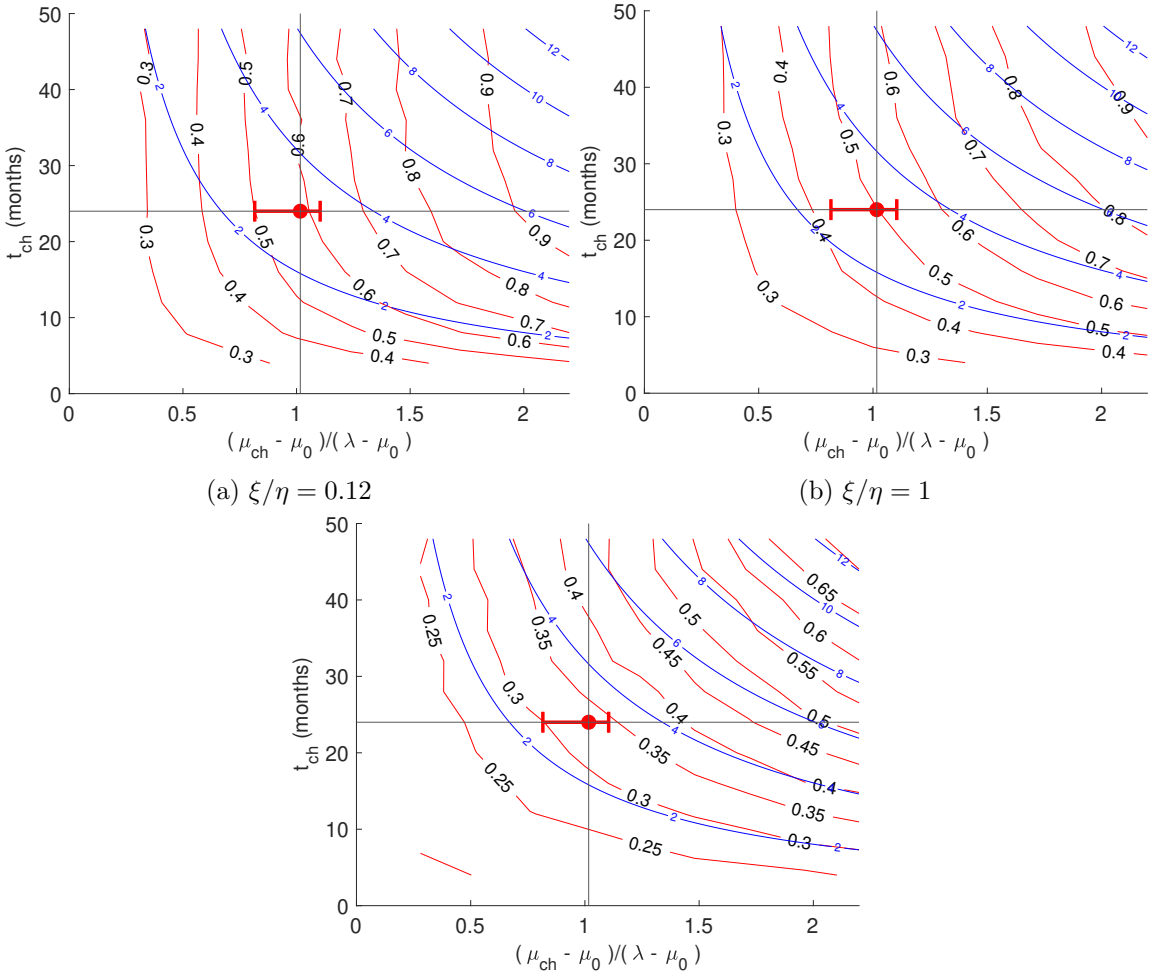
(a) Recurrence-free survival

(b) Fraction cured

Figure 5.5: Placebo curves for $\xi/\eta = 0.12$ (blue), $\xi/\eta = 1$ (yellow) and $\xi/\eta = 5$ (red), with their respective 95% CI error bands for the simulations. On the left, the placebo KM curve (green) is also plotted with its 95% CI error band. The horizontal line represents the large-time RFS value of 0.187, which is the same for all the curves.

we must choose $\xi/\eta = 1$ in order to fit the model to the Olaparib KM data. For a patient with the same doubling rate of $\lambda = 2.99$ month^{-1}, but a ξ/η parameter higher or lower than the value $\xi/\eta = 1$ required by the fit, it is evident that the recurrence or cure rates would not match those of the Olaparib clinical trial. Our next goal was to investigate how the RFS fraction and the cured fraction change when only ξ/η changes, while all the other parameters (including λ) are kept fixed at the values corresponding to the Olaparib KM data.

For the Olaparib KM curve fit with $\lambda = 2.99$ month^{-1} and $\xi/\eta = 1$, changing ξ/η to the lower value $\xi/\eta = 0.12$ or to the higher value $\xi/\eta = 5$ (while keeping λ, N_0, μ_0, μ_{ch} and t_{ch} all fixed), respectively increases and decreases long-term RFS for Olaparib chemotherapy, as expected: if tumor foci transition to the dormant state less often, they become more sensitive to elimination by chemotherapy, thus improving disease prognosis. For $\xi/\eta = 0.12$, $\xi/\eta = 1$ and $\xi/\eta = 5$, the large-time RFS fractions for Olaparib chemotherapy are 59.2%, 50.3% and 33.2%, respectively. The RFS fraction, as well as the cured fraction for tumors



(a) $\xi/\eta = 0.12$ (b) $\xi/\eta = 1$

(c) $\xi/\eta = 5$

Figure 5.6: Level curves of the large-time RFS in the plane $(\frac{\mu_{ch}-\mu_0}{\lambda-\mu_0}, t_{ch})$, where for all three values of ξ/η, the parameters λ, N_0 and μ_0 were chosen to be the same as those of the $\xi/\eta = 1$ fit.

treated with Olaparib at these three values of ξ/η are shown in Fig. 5.4, and the respective curves in the absence of chemotherapy (placebo) are shown in Fig. 5.5.

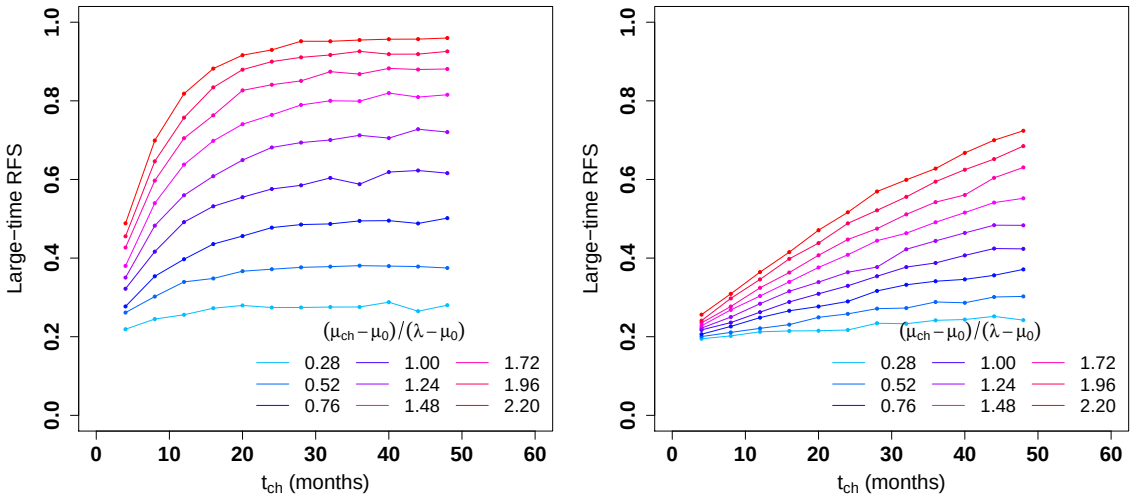
Note that the fraction of patients who get cured as a function of time cannot be determined directly from the clinical trial IPD, because the status of censored patients is unknown: censored patients are those who, for one reason or another, choose to leave the trial at some time point for reasons other than recurrence, which means that their subsequent history of relapse is unknown. However, not only can our model predict the cured fraction for a set of parameters that fits the KM curve, but it can also predict how the

cured fraction changes if one of the parameters from the fit changes, as shown in Figs. 5.4 and 5.5 for different values of ξ/η, with the parameters λ, N_0, μ_0, μ_{ch} and t_{ch} all held fixed at the same values as in the $\xi/\eta = 1$ fit.

Also note that although the placebo curves for different values of ξ/η shown in Fig. 5.5 are different from one another, their asymptotic values at large times (18.7% RFS fraction) are all the same, unlike the Olaparib curves shown in Figs. 5.4. This is because in the absence of chemotherapy, the mean recurrence time (MRT) depends on ξ/η (see Eqs. (3.15) and (3.16)), whereas the large-time RFS does not (see Eqs. (3.3) and (3.4)). This means that for different values of ξ/η and all other parameters equal, the amount of benefit from treatment with Olaparib (which is defined here as the increase in large-time RFS relative to placebo) is only determined by the Olaparib curves.

Fixing λ, μ_0 and N_0 to be the same as those of the $\xi/\eta = 1$ fit, while changing the ξ/η parameter to either $\xi/\eta = 0.12$ or $\xi/\eta = 5$, for each point of the 9×12 (μ_{ch}, t_{ch}) grid used for the analysis in Section 5.1, 5000 tumors were simulated, from which level curves of the large-time RFS were generated. The plots for $\xi/\eta = 0.12$, $\xi/\eta = 1$ and $\xi/\eta = 5$ are shown in Fig. 5.6. As expected, the large-time RFS fraction that results from increasing the dose of Olaparib or the chemotherapy duration is significantly higher for the lower parameter value $\xi/\eta = 0.12$ than for the higher value $\xi/\eta = 5$, with a considerably greater difference between $\xi/\eta = 1$ and $\xi/\eta = 5$ versus $\xi/\eta = 1$ and $\xi/\eta = 0.12$.

Indeed, Fig. 5.7 shows that increasing the dose of Olaparib (i.e., increasing μ_{ch}) by the same amount at a fixed chemotherapy duration t_{ch} results in a greater large-time RFS fraction if $\xi/\eta = 1$ changes to the lower value $\xi/\eta = 0.12$, and in a significantly smaller RFS fraction if $\xi/\eta = 1$ changes to the higher value $\xi/\eta = 5$. Similarly, extending the duration of chemotherapy at fixed dose results in a better prognosis for $\xi/\eta = 0.12$ versus $\xi/\eta = 5$: for the lower parameter value, the large-time RFS saturates more quickly as a function of t_{ch} to the upper bound defined by the limit $t_{ch} \to \infty$, as shown in Fig. 5.8. This means that for a fixed level of response, the duration of chemotherapy can be lower for patients



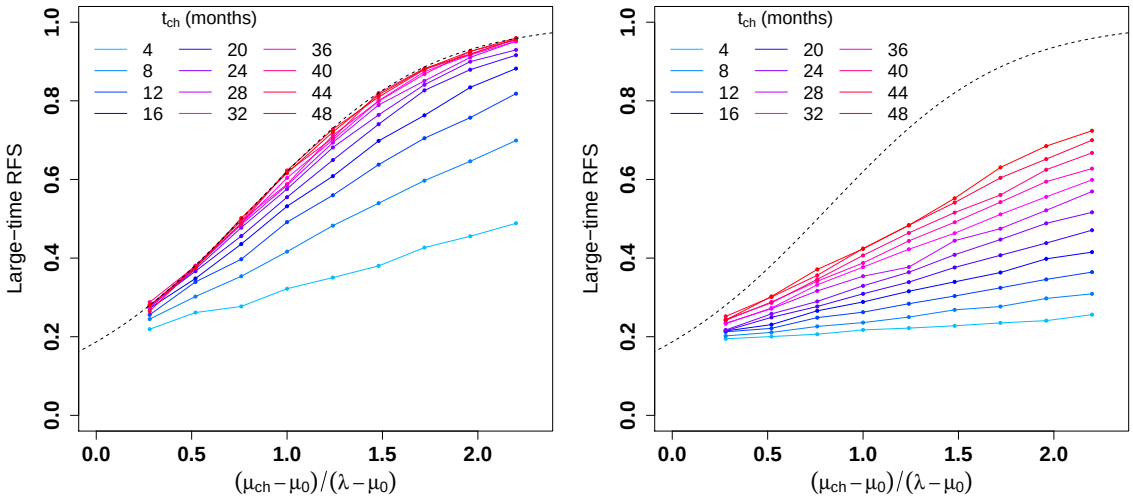
(a) $\xi/\eta = 0.12$ (b) $\xi/\eta = 5$

Figure 5.7: Large-time RFS versus t_{ch} at fixed μ_{ch}, for $\lambda = 2.99$ month^{-1} and all the other parameters the same as those of the $\xi/\eta = 1$ fit.

with a low transition rate from the active to the dormant state (and a high rate for the reverse transition) to ensure that there is enough time spent in the chemosensitive vs the chemoresistant state, so that the tumor foci are killed by chemotherapy. In contrast, if the transition rate from the active to the dormant state is high (and the rate for the reverse transition is low), the duration of chemotherapy needs to be longer to ensure that enough time is spent in the chemosensitive state.

Note that even though the RFS fraction changes to a greater value for $\xi/\eta = 0.12$ than for $\xi/\eta = 5$ as the chemotherapy duration t_{ch} is increased at fixed μ_{ch}, the amount of benefit from extending chemotherapy, defined here as the *percent* increase in large-time RFS, is in fact greater for $\xi/\eta = 5$ than for $\xi/\eta = 0.12$. This is simply because the large-time RFS saturates toward the upper bound defined by the limit $t_{ch} \to \infty$ faster for $\xi/\eta = 0.12$ than for $\xi/\eta = 5$, so that its rate of change with respect to t_{ch} also decreases faster.

In this section, we saw how patients with different transition rates between dormant and active states respond to treatment, according to our model. Although it is expected that the transition parameter may in principle vary across patients in a trial, the idea is



(a) $\xi/\eta = 0.12$ (b) $\xi/\eta = 5$

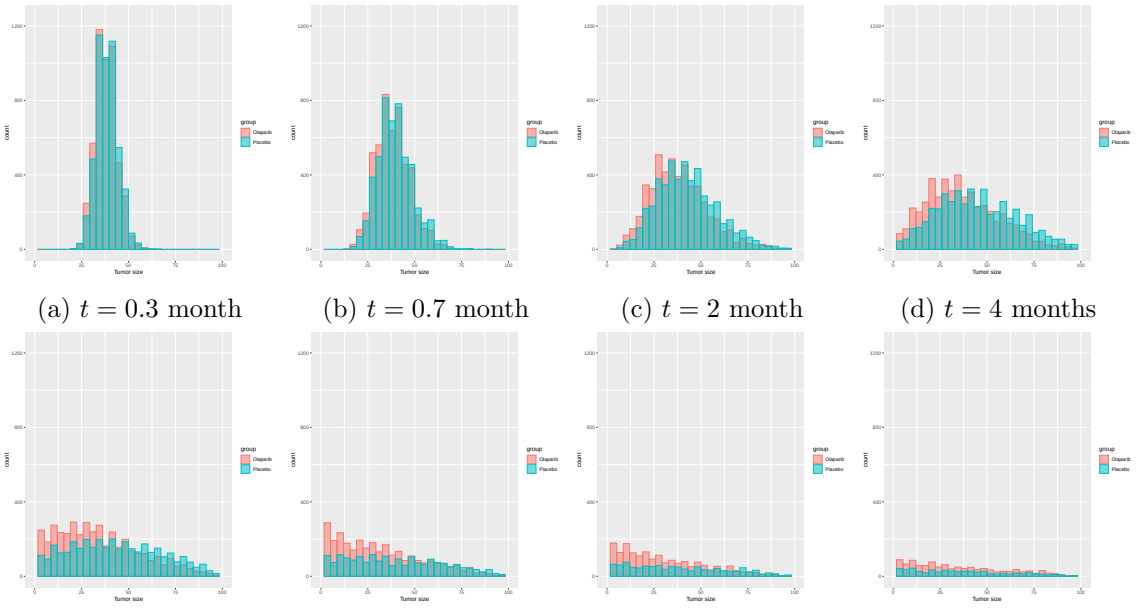
Figure 5.8: Large-time RFS versus μ_{ch} at fixed t_{ch}, for $\lambda = 2.99$ month^{-1} and all the other parameters the same as those of the $\xi/\eta = 1$ fit. The dashed line corresponds to the limit $t_{ch} \to \infty$.

to not only indicate overall trends, but also emphasize how the treatment strategy and the prognosis of individual patients may strongly depend on parameters that could be difficult to measure in a clinical setting. In the specific case of PARP inhibitor Olaparib, there is evidence from recent cell-line experiments that at low dosages it could induce senescence of ovarian cancer cells, instead of apoptosis [102]. If such effects do also occur in vivo, they are likely patient-specific and our analysis shows that they would strongly affect progression-free survival prospects.

5.3 Predicted tumor-growth dynamics for the Olaparib study and effects of dormancy on dynamics

In this section, a few snapshots of the tumor-growth dynamics are shown for the clinical data fit with $\xi/\eta = 1$, and also for the same parameter values as the $\xi/\eta = 1$ fit, but with a higher or lower ξ/η.

For the fit with $\xi/\eta = 1$ (for which the doubling time $1/\lambda$ is about 10 days), the effect



(a) $t = 0.3$ month (b) $t = 0.7$ month (c) $t = 2$ month (d) $t = 4$ months

(e) $t = 8$ months (f) $t = 16$ months (g) $t = 24$ months (h) $t = 32$ months

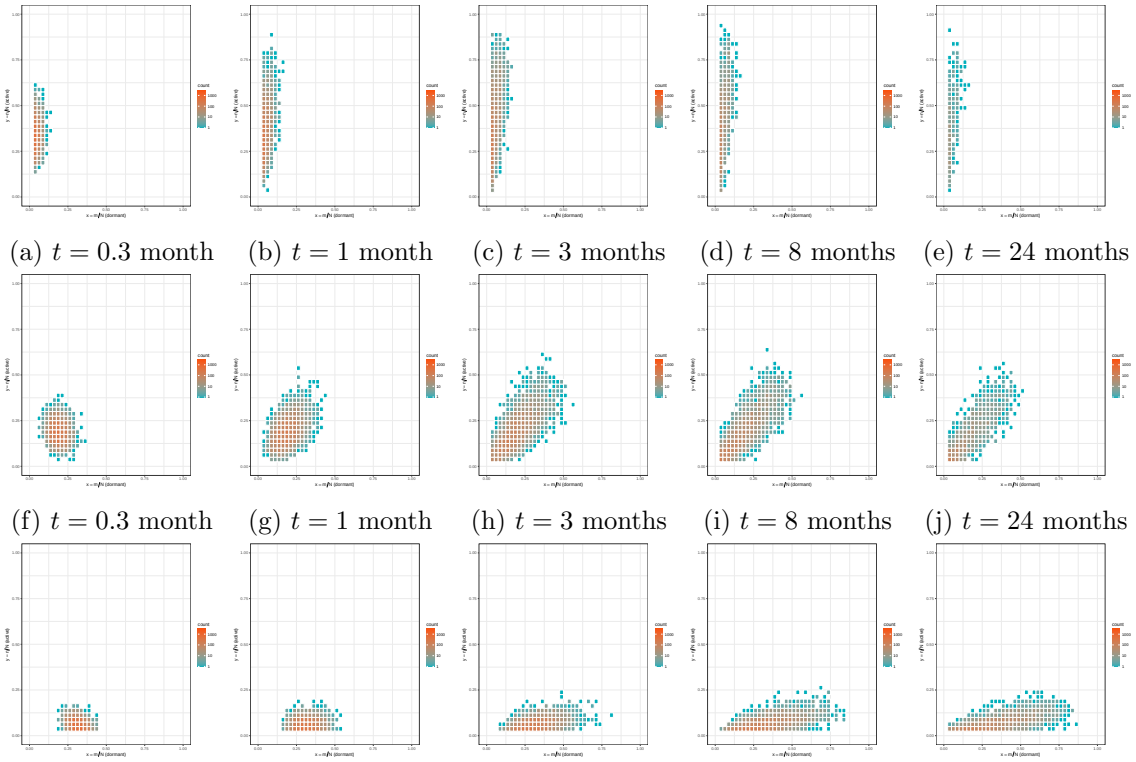
Figure 5.9: Time evolution of the tumor-size distribution for Olaparib (red) versus placebo (green), for $\xi/\eta = 1$. The histograms are defined within the reduced sample space of patients at risk of recurrence, and were generated from simulations with 5000 patients, using the parameters that fit the clinical KM curves.

of treatment with Olaparib on tumor growth can be seen in Fig. 5.9, where tumor-size distributions for Olaparib and placebo against time are compared. The upper tail of the placebo distribution is longer, which causes a higher recurrence rate by diffusion. This effect is amplified by the higher diffusion rates at the upper end of the distribution, because of the linear dependence of the diffusion coefficients on the numbers of dormant and active tumor foci.

For $\xi/\eta = 0.12$, $\xi/\eta = 1$ and $\xi/\eta = 5$ (with all the other parameters fixed by the $\xi/\eta = 1$ fit), the state-space distribution for Olaparib chemotherapy at different time points is plotted in Fig. 5.10. The color gradient of the heat maps was defined on a logarithmic scale, for best visualization. In all three cases, we see that the distribution is concentrated in a relatively small region, with the location of the peak depending upon the parameter ξ/η. We also see that as early as $t = 0.3$ month, which is of the order of the doubling time $1/\lambda = 10$ days, the initial sharp peak at $m_0 = n_0 = N_0/2$ has already shifted to the region

defined by the ratio ξ/η.

The effect of the transitions between dormant and active states on growth dynamics is seen in Fig. 5.11. For $\xi/\eta = 0.12$, diffusion is favored in the direction of the y axis (representing the number of active tumor foci), whereas for $\xi/\eta = 5$, diffusion is favored in the x axis (representing the number of dormant tumor foci). This is because the diffusion coefficient for transitions between dormant and active states is given by $\eta x + \xi y$, and the state-space distribution for $\xi/\eta = 0.12$ is peaked at low values of x and relatively high values of y, whereas for $\xi/\eta = 5$ it is the other way around. It is also seen that the broadening of the tumor-size distribution happens earlier for $\xi/\eta = 0.12$ than $\xi/\eta = 5$. Indeed, for $\xi/\eta = 0.12$ it essentially follows the broadening of the y-distribution (active tumor foci), whereas for $\xi/\eta = 5$ it follows the broadening of the x-distribution (dormant tumor foci). While the latter is only caused by diffusion due to transitions between dormant and active states, the former is also caused by diffusion due to doubling and death events, since only active tumor foci are chemosensitive. We see that recurrence and cure rates are mainly explained by diffusion, which evidently happens earlier for $\xi/\eta = 0.12$ versus $\xi/\eta = 5$. However, even though recurrence happens first for $\xi/\eta = 0.12$, the large-time recurrence-free fraction is higher for $\xi/\eta = 0.12$ than for $\xi/\eta = 5$ (see Fig. 5.4).

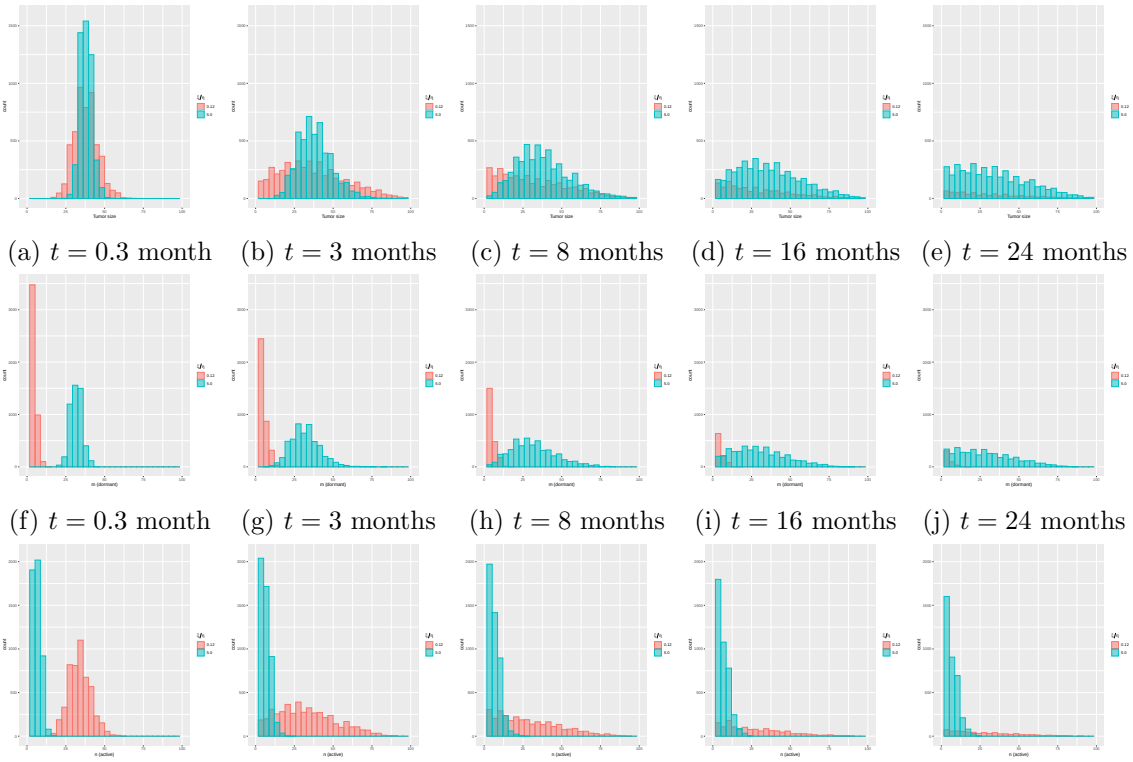


(a) $t = 0.3$ month (b) $t = 1$ month (c) $t = 3$ months (d) $t = 8$ months (e) $t = 24$ months

(f) $t = 0.3$ month (g) $t = 1$ month (h) $t = 3$ months (i) $t = 8$ months (j) $t = 24$ months

(k) $t = 0.3$ month (l) $t = 1$ month (m) $t = 3$ months (n) $t = 8$ months (o) $t = 24$ months

Figure 5.10: Heat maps of the state-space distribution in time, from simulations of Olaparib chemotherapy. (a–e) $\xi/\eta = 0.12$, (f–j) $\xi/\eta = 1$, (k–o) $\xi/\eta = 5$. The colors represent the number of patients (within an ensemble of 5000 patients) at risk of recurrence at each state (m, n) in a logarithmic scale, where m (x axis) is the number of dormant tumor foci and n (y axis) is the number of active tumor foci.



(a) $t = 0.3$ month (b) $t = 3$ months (c) $t = 8$ months (d) $t = 16$ months (e) $t = 24$ months

(f) $t = 0.3$ month (g) $t = 3$ months (h) $t = 8$ months (i) $t = 16$ months (j) $t = 24$ months

(k) $t = 0.3$ month (l) $t = 3$ months (m) $t = 8$ months (n) $t = 16$ months (o) $t = 24$ months

Figure 5.11: Histograms from simulations of Olaparib chemotherapy, for $\xi/\eta = 0.12$ (red) versus $\xi/\eta = 5$ (green). (a–e) Tumor-size distribution, (f–j) marginal distribution of the number of dormant tumor foci, (k–o) marginal distribution of the number of active tumor foci. All the histograms are defined within the reduced sample space of patients at risk of recurrence.

Chapter 6

Conclusions and outlook

In this thesis, we developed a mechanistic mathematical model that describes the stochastic dynamics of tumor recurrence through a Quasi Birth-and-Death (QBD) process. The main assumption is the presence of residual tumor foci that can transition between a chemoresistant dormant state and a chemosensitive active-growth state. In Chapter 2, we started with a continuous-time, discrete-state master equation that gives the finite-time transition probability $p_{m,n}(t)$ from an initial state (m_0, n_0) to a state (m, n), where m and n are, respectively, the numbers of dormant and active tumor foci. We then showed that for a large detectable-tumor size N, the discrete master equation can be well approximated by a drift-diffusion equation in a continuous state space. Recurrence and cure were built into the model by imposing absorbing boundary conditions at the cure state $(0, 0)$ and at the recurrence boundary defined by $m + n = N$, respectively.

Using the forward and backward Kolmogorov approaches in the continuum limit, in Chapter 2 we derived equations for the time-dependent probabilities of recurrence and cure, along with appropriate boundary conditions. The stationary solution at large times was obtained analytically in Chapter 3 (see Eq. (3.3)) and was shown to display a phase transition as a function of μ_A/λ, where μ_A is the death rate of active tumor foci and λ is their doubling rate. We also derived an equation for the mean recurrence time (MRT), which was solved analytically to leading order in $1/N$ by dropping the diffusion (second-derivative) terms in the equation, an approximation that works outside thin boundary layers along the reflecting barriers (see Eqs. (3.15) and (3.16) for this "outer solution").

The analytical results were compared to simulations of the discrete-state QBD model.

The large-time probability of recurrence obtained in simulations matched the analytical solution, whereas the MRT from the simulations showed a small discrepancy to the leading order outer solution of the MRT equation, except near the critical point $\mu_A/\lambda = 1$, where the discrepancy was larger due to finite-size effects (finite value of N). In principle, it is possible to get an improved approximation by solving the MRT equation within the boundary layers (where the variables have to be rescaled) and constructing a composite solution by the method of matched asymptotic expansions [90].

In Chapters 4 and 5, we applied the model to recurrence data from a phase 3 ovarian cancer clinical trial, in which patients were either treated with PARP inhibitor Olaparib or received placebo (control group), following standard platinum-based chemotherapy. In Chapter 4, we demonstrated that model parameters can be estimated from the clinical KM curves for the placebo and Olaparib cohorts as functions of the ratio of the transition rates to and from dormancy (ξ/η), or equivalently, as functions of the doubling time $1/\lambda$, which relates the natural time scale of the model to the time scale of the KM curves. To ensure that the model fits the placebo KM curve, the parameter ξ/η is required to satisfy a linear relationship with λ, as shown in Fig. 4.5a.

In Chapter 5, we used the parameters from the fits to predict long-term recurrence free survival (RFS) when either the duration (represented by the parameter t_{ch}) or the dose (represented by the parameter $(\mu_{ch} - \mu_0)/(\lambda - \mu_0)$) of maintenance therapy with olaparib are increased. We showed that the level curves in the (μ_{ch}, t_{ch}) plane for large-time RFS are essentially independent of the choice of time-scale parameter λ for the fits. We then investigated the influence of the parameter ξ/η on progression-free survival prospects, by changing it while keeping all the other parameters from the fit constant. We showed that while large-time RFS is independent of ξ/η in the absence of chemotherapy, it strongly depends on the latter parameter if patients receive chemotherapy during a finite amount of time. We predicted how the placebo and olaparib KM curves would change as a function of the parameter ξ/η, and also how the cured fraction worsens or improves as ξ/η increases or

decreases, respectively. Finally, we predicted how the effectiveness of increasing the duration or the dose of olaparib chemotherapy depends on the ratio ξ/η by obtaining large-time RFS level curves in the (μ_{ch}, t_{ch}) plane for different values of this parameter.

The model proposed in this thesis focused on one aspect of tumor progression that affects recurrence rates, namely the chemoresistance caused by transitions to a dormant state. In a more realistic scenario, tumors are heterogeneous and may develop resistance to drugs through selection and/or acquired mutations over the course of treatment. However, our model lays a foundation for the development of more accurate, higher-dimensional models that would account for the intricacies of clonal evolution during treatment and which would be applicable to individual patients with known genomic signatures. This would require a combination of data-driven approaches using machine learning methodologies [103] and mechanistic modeling of the kind that was explored in this thesis, but with a larger number of variables and parameters. Such models could be developed, for example, using branching processes [48, 104–106] and adapting/incorporating our approach using absorbing boundary conditions to study recurrence of tumors, with the aim of finding strategies to minimize treatment evasion in evolving tumors.

Another possible future direction would be the investigation of a different version of our model with logistic (rather than exponential) growth. This means limiting tumor growth by replacing our linear death rates $\mu_n = \mu n$ by, for example, quadratic rates $\mu_n = \mu n + \frac{\alpha}{K} n^2$, as in the so-called stochastic Verhulst model [107]. In our model, we made the assumption that the deaths of different tumor foci within a given time interval are independent events. However, this assumption is not strictly correct, because tumor growth is often limited by blood supply, or in the case of liquid tumors, by limited nutrition available for tumor cells, which causes growth rates to decrease at sufficiently large tumor sizes [56].

Appendix A

Structure of the infinitesimal transition matrix

Projecting the operator $\hat{Q} = \hat{Q}_{bulk} + \hat{Q}_{edge}$ given by Eqs. (2.6) and (2.7) on both sides between basis vectors $\langle m, n|$ and $|m', n'\rangle$, we get matrix elements with a block-tridiagonal structure in the direct-product linear space $\mathcal{D} \otimes \mathcal{A}$:

$$\mathbb{Q} = \mathbb{Q}^{bulk} + \mathbb{Q}^{edge} = \begin{pmatrix} \mathbb{Q}_{00} & \mathbb{Q}_{01} & 0 & \dots & \dots & 0 \\ \mathbb{Q}_{10} & \mathbb{Q}_{11} & \mathbb{Q}_{12} & \dots & \dots & 0 \\ 0 & \mathbb{Q}_{21} & \mathbb{Q}_{22} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \mathbb{Q}_{N-1,N} \\ 0 & 0 & \dots & \dots & \mathbb{Q}_{N,N-1} & \mathbb{Q}_{NN} \end{pmatrix} \tag{A.1}$$

From the geometry of the state space boundary (see Fig. 2.1b), it follows that the block matrices $\mathbb{Q}_{mm}$ decrease in size as m increases: $\mathbb{Q}_{mm}$ is an $(N - m + 1) \times (N - m + 1)$ matrix, since only the subspace spanned by the states $|m, 0\rangle, \dots, |m, N - m\rangle$ is accessible. The bulk part of each block $\mathbb{Q}_{mm}$ is tridiagonal and acts within the accessible subspace of $\mathcal{A}$:

$$\mathbb{Q}^{bulk}_{mm} = \begin{pmatrix} -m\eta & 0 & 0 & 0 & \dots & 0 \\ 0 & -m\gamma_D - \gamma_A & 2\mu_A & 0 & \dots & 0 \\ 0 & \lambda & -m\gamma_D - 2\gamma_A & 3\mu_A & \dots & 0 \\ 0 & 0 & 2\lambda & -m\gamma_D - 3\gamma_A & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & (N-m)\mu_A \\ 0 & 0 & 0 & \dots & \dots & -m\gamma_D - (N-m)\gamma_A \end{pmatrix}, \tag{A.2}$$

where

$$\gamma_D \equiv \mu_D + \eta\,,$$
$$\gamma_A \equiv \lambda + \mu_A + \xi\,. \tag{A.3}$$

The bulk parts of the off-diagonal blocks are the matrices

$$\mathbb{Q}^{bulk}_{m,m+1} = (m+1)\begin{pmatrix} \mu_D-\eta & & & \\ \eta & \mu_D & & \\ & \eta & \ddots & \\ & & \ddots & \mu_D \\ & & & \eta \end{pmatrix}_{(N-m+1)\times(N-m)} \tag{A.4}$$

$$\mathbb{Q}^{bulk}_{m,m-1} = \begin{pmatrix} 0 & \xi & & & & \\ & 0 & 2\xi & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & (N-m+1)\xi \end{pmatrix}_{(N-m+1)\times(N-m+2)} \tag{A.5}$$

The edge corrections for the three block matrices above are given by

$$\mathbb{Q}^{edge}_{mm} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -(N-m)\mu_A \\ 0 & \dots & & m\gamma_D + (N-m)\gamma_A \end{pmatrix}_{(N-m+1)\times(N-m+1)} \tag{A.6}$$

$$\mathbb{Q}^{edge}_{m,m+1} = \begin{pmatrix} 0 & & & & \\ 0 & 0 & & & \\ & 0 & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & -(m+1)\mu_D \\ & & & & -(m+1)\eta \end{pmatrix}_{(N-m+1)\times(N-m)} \tag{A.7}$$

$$\mathbb{Q}^{edge}_{m,m-1} = \begin{pmatrix} 0 & 0 & & & & \\ & 0 & 0 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & -(N-m+1)\xi \end{pmatrix}_{(N-m+1)\times(N-m+2)} \tag{A.8}$$

Appendix B

Solution of the M/M/1/N queue with absorbing boundary states

In this appendix, the exact solution of the $M/M/1/N$ queue with absorbing boundary states is presented. Here we follow Ref. [82], where the solution is derived. This queuing model is closely related to our QBD model, with the major differences that the state space is one-dimensional and birth and death are the only possible transitions.

Let us consider a birth-and-death process with a finite number of states $0, 1, 2, \cdots, N$ representing the size of the queue, where 0 and N are absorbing boundary states and λ_n and μ_n are, respectively, the birth rate from state n to $n+1$, and the death rate from state n to $n-1$. The absorbing boundary conditions are defined by the conditions $\lambda_0 = \mu_N = 0$. The allowed transitions are shown in Fig. B.1.

Let $X(t)$ be the stochastic process that represents the size of the queue at time t. We assume that the waiting time between the transitions $n-1 \to n$ and $n \to n+1$ is exponentially distributed with mean $1/\lambda_n$, and the waiting time between the transitions $n+1 \to n$ and $n \to n-1$ is exponentially distributed with mean $1/\mu_n$. In the limit $\xi/\lambda \to 1$ and $\eta/\lambda \to \infty$, our model reduces to this process with birth and death rates $\lambda_n = n\lambda$ and $\mu_n = n\mu$, which is the limiting case in which tumor foci are always active and never become dormant.

Our goal is to determine the transition probability

$$p_{kn}(t) \equiv Pr\{X(t) = n | X(0) = k\}, \tag{B.1}$$

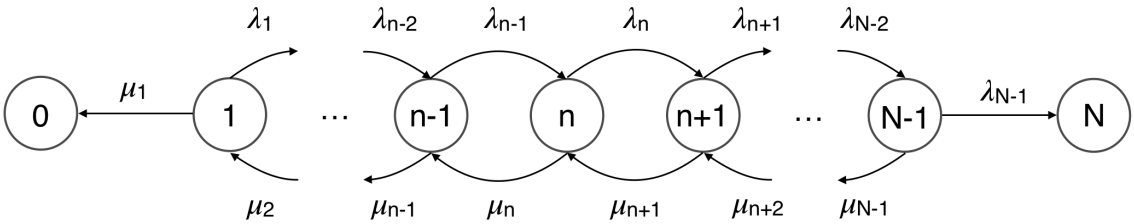


Figure B.1: Diagram of allowed transitions in the M/M/1/N queue with absorbing boundary states.

with initial condition

$$p_{kn}(0) = \delta_{nk}. \tag{B.2}$$

The transition probabilities within a small time interval Δt are given by

$$Pr\{X(t+\Delta t) = j | X(t) = i\} = \begin{cases} \lambda_i \Delta t + o(\Delta t^2) & \text{if} \quad j = i+1 \\ \mu_i \Delta t + o(\Delta t^2) & \text{if} \quad j = i-1 \\ 1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t^2) & \text{if} \quad j = i \\ o(\Delta t^2) & \text{otherwise} \end{cases} \tag{B.3}$$

From these assumptions, the following differential-difference equations can be derived for the finite-time transition probabilities:

$$\begin{cases} \dot{p}_{k0}(t) &= \mu_1 \, p_{k0}(t) \\ \dot{p}_{k1}(t) &= -(\lambda_1 + \mu_1) \, p_{k1}(t) + \mu_2 \, p_{k2}(t) \\ \dot{p}_{kn}(t) &= \lambda_{n-1} \, p_{k,n-1}(t) - (\lambda_n + \mu_n) \, p_{kn}(t) + \mu_{n+1} \, p_{k,n+1}(t) \quad (\text{if } 1 < n < N-1) \\ \dot{p}_{k,N-1}(t) &= \lambda_{N-2} \, p_{k,N-2}(t) - (\lambda_{N-1} + \mu_{N-1}) \, p_{k,N-1}(t) \\ \dot{p}_{k,N}(t) &= \lambda_{N-1} \, p_{k,N-1}(t) \end{cases} \tag{B.4}$$

Eqs. (B.4) can be solved by Laplace transformation (which converts them to algebraic equations in the s-domain), followed by matrix inversion and inverse Laplace transformation.

The solution is

$$p_n(t) = \begin{cases} \pi_0 + \mu_1\mu_2\ldots\mu_k \sum_{i=1}^{N-1} A_i\, e^{-\alpha_i t} & \text{if} \quad n = 0 \\ \mu_{n+1}\mu_{n+2}\ldots\mu_k \sum_{i=1}^{N-1} A_{ni}\, e^{-\alpha_i t} & \text{if} \quad 1 \le n \le k \\ \sum_{i=1}^{N-1} B_{ki}\, e^{-\alpha_i t} & \text{if} \quad n = k \\ \lambda_{n+1}\lambda_{n+2}\ldots\lambda_k \sum_{i=1}^{N-1} B_{ni}\, e^{-\alpha_i t} & \text{if} \quad k < n \le N-1 \\ \pi_N + \lambda_k\lambda_{k+1}\ldots\lambda_{N-1} \sum_{i=1}^{N-1} B_i\, e^{-\alpha_i t} & \text{if} \quad n = N \end{cases} \tag{B.5}$$

where

$$\pi_0 = 1 - \pi_N = \left[1 + \frac{\lambda_{N-1}}{\mu_{N-1}} + \frac{\lambda_{N-1}\lambda_{N-2}}{\mu_{N-1}\mu_{N-2}} + \cdots + \frac{\lambda_{N-1}\lambda_{N-2}\ldots\lambda_{k+1}}{\mu_{N-1}\mu_{N-2}\cdots\mu_{k+1}}\right]\beta^{-1}, \tag{B.6}$$

with

$$\beta = \left[1 + \frac{\lambda_{N-1}}{\mu_{N-1}} + \frac{\lambda_{N-1}\lambda_{N-2}}{\mu_{N-1}\mu_{N-2}} + \cdots + \frac{\lambda_{N-1}\lambda_{N-2}\ldots\lambda_1}{\mu_{N-1}\mu_{N-2}\cdots\mu_1}\right]. \tag{B.7}$$

The coefficients A_i, A_{ni}, B_i and B_{ni} are given by

$$A_i = \frac{P_{N-k-1}(-\alpha_i)}{(-\alpha_i)\prod\limits_{\substack{j=1\\ j\neq i}}^{N-1}(\alpha_j - \alpha_i)}, \tag{B.8}$$

$$A_{ni} = \frac{P_{N-k-1}(-\alpha_i)Q_{n-1}(-\alpha_i)}{\prod\limits_{\substack{j=1\\ j\neq i}}^{N-1}(\alpha_j - \alpha_i)}, \tag{B.9}$$

$$B_i = \frac{Q_{k-1}(-\alpha_i)}{(-\alpha_i)\prod\limits_{\substack{j=1\\ j\neq i}}^{N-1}(\alpha_j - \alpha_i)} \tag{B.10}$$

and

$$B_{ni} = \frac{P_{N-n-1}(-\alpha_i)Q_{k-1}(-\alpha_i)}{\prod\limits_{\substack{j=1\\ j\neq i}}^{N-1}(\alpha_j - \alpha_i)}. \tag{B.11}$$

In these equations, $P_r(\theta)$ and $Q_r(\theta)$ are polynomials satisfying the recurrence relations

$$P_r(\theta) - (\lambda_{N-r+1} + \mu_{N-r+1} + \theta)P_{r-1}(\theta) + \lambda_{N-r+1}\mu_{N-r+2}P_{r-2}(\theta) = 0 \tag{B.12}$$

and

$$Q_r(\theta) - (\lambda_r + \mu_r + \theta)Q_{r-1}(\theta) + \lambda_{r-1}\mu_r Q_{r-2}(\theta) = 0, \tag{B.13}$$

with initial conditions $P_0(\theta) = Q_0(\theta) = 1$, $P_1(\theta) = \mu_N + \theta$ and $Q_1(\theta) = \lambda_1 + \mu_1 + \theta$.

The parameters $\alpha_1, \cdots, \alpha_N$ are the eigenvalues of the matrix

$$B(0) = \begin{pmatrix} \lambda_1+\mu_1 & -\sqrt{\lambda_1\mu_2} & & \cdots & \cdots & 0 \\ -\sqrt{\lambda_1\mu_2} & \lambda_2+\mu_2 & \ddots & & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{N-2}+\mu_{N-2} & -\sqrt{\lambda_{N-2}\mu_{N-1}} & \\ 0 & 0 & \cdots & -\sqrt{\lambda_{N-2}\mu_{N-1}} & \lambda_{N-1}+\mu_{N-1} & \end{pmatrix} \tag{B.14}$$

The large-time probabilities of absorption by the boundary states π_0 and π_n are given by Eqs. (B.6) and (B.7). In connection to our model, for the special case $\lambda_n = n\lambda$, $\mu_n = n\mu$ we find

$$\pi_N = 1 - \pi_0 = \frac{1-(\mu/\lambda)^k}{1-(\mu/\lambda)^N}, \tag{B.15}$$

which corresponds to the large-time probability of recurrence of our model in the limit $\xi/\lambda \to 1$, $\eta/\lambda \to \infty$, for an initial tumor size k. It is easy to show that if N is moderately large ($N = 100$, for example, would be sufficient) and $k \gtrsim 3$, this solution is essentially the same as our stationary solution for the continuous version of the QBD model, given by Eq. (3.3).

Appendix C

The Kaplan-Meier survivorship curve

A frequent problem in the medical field and in many other fields of study is the statistical estimation of so-called *failure rates* (i.e., the frequency at which a certain *failure* event of interest occurs in a population) from incomplete data. Depending on the application, the failure event may be the death of a patient, the relapse of a disease, the failure of a system or component, etc. In this context, time-to-failure observations are often censored: for example, in disease-recurrence studies, patients may leave the study or may die from other causes, so that the recurrence/no-recurrence status of these patients is unknown beyond the time of the last follow-up. In order to efficiently estimate the fraction $P(t)$ of patients whose recurrence-free lifespans are greater than t from such incomplete records, the methodology developed by Kaplan and Meier in their seminal paper [96] is a suitable approach that makes no assumptions on the form of $P(t)$.

The data that generates a Kaplan-Meier (KM) plot is a list of event times, where each event is either labeled "1" (representing the event of interest, for example, disease recurrence) or "0" (representing censoring or loss of information). In the context of this thesis, the event of interest is recurrence of the tumor in a patient, and the survivorship function $P(t)$ should be interpreted as a time-dependent recurrence-free survival probability. In other words, $P(t)$ is defined as the probability that any evidence of the disease is absent until at least time t. The graph of $P(t)$, also known as Kaplan-Meier (KM) curve, is estimated from data using the KM non-parametric methodology, and allows researchers to compare cohorts of patients in clinical trials or studies (e.g. patients that received a drug versus patients that only received placebo).

Naively, it may be tempting to use only observations from patients whose recurrence/no-recurrence status is known within some pre-defined time interval to estimate the survivorship function $P(t)$. This is known as a reduced-sample estimate of $P(t)$. The problem with this approach is that it does not use the data efficiently by extracting all the available information, which may lead to biased estimates of $P(t)$, unless the reduced sample itself is a random sample (which mostly likely is not the case). The methodology developed by Kaplan and Meier circumvents this issue with the so-called product-limit estimate of $P(t)$, which is the unique non-parametric distribution that maximizes the likelihood of the observations.

If N is the number of observations (say, recurrence or censoring events) and the observed event times are sorted in increasing order as $t'_1 \leq t'_2 \leq \cdots t'_N$, the recurrence-free survival function at time t can be estimated as

$$\hat{P}(t) = \prod_r \frac{N-r}{N-r+1}, \tag{C.1}$$

where the product is over all indices r such that each $t'_r \leq t$ is a recurrence time.

The variance of the survival function can be estimated as

$$\hat{V}[\hat{P}(t)] = \hat{P}^2(t) \sum_r \frac{1}{(N-r)(N-r+1)}, \tag{C.2}$$

where, once again, the sum is over all indices r such that $t'_r \leq t$ and t'_r is a recurrence time.

Appendix D

Algorithm to reconstruct individual patient data from Kaplan-Meier curves

In this appendix, the algorithm used in Chapter 4 to reconstruct individual patient data (IPD) from Kaplan-Meier (KM) curves is outlined. This algorithm was proposed in Ref. [101]. The IPD tables generated from the KM plots and risk tables for the Olaparib and placebo cohorts of the ovarian cancer clinical trial are also shown on the pages to follow.

Algorithm 1: Iterative method to reconstruct IPD from Kaplan-Meier plots [101]

Input: T_k: time values extracted from KM plot ($k = 1, \cdots, n$)
S_k: survival function values extracted from KM plot
$nrisk_i$: number of patients at risk at i-th interval from provided table ($i = 1, \cdots, nint$)
$trisk_i$: time at the beginning of i-th interval in risk table
$totevents$: reported total number of events (if available)

Output: For each time interval i of the risk table, a list of censored patients $c = 1, \cdots, n\hat{cens}or_i$ and respective censor times $\hat{cent}_c$.

Define: $lower_i = \min\{k : T_k \geq trisk_i\}$
$upper_i = \max\{k : T_k \leq trisk_{i+1}\}$

Step 1:
Calculate initial guess for number censored on i-th interval:
$n\hat{cens}or_i \leftarrow (S_{lower_{i+1}}/S_{lower_i}) * nrisk_i - nrisk_{i+1}$

Step 2:
Distribute the $c = 1, \cdots, n\hat{cens}or_i$ censor times $\hat{cent}_c$ evenly over i-th interval:
$\hat{cent}_c = T_{lower_i} + c * (T_{lower_{i+1}} - T_{lower_i})/(n\hat{cens}or_i + 1)$
$\hat{cen}_k \leftarrow$ number of estimated censor times $\hat{cent}_c$ within $[T_k, T_{k+1}]$

(continues...)

Step 3:

Calculate the number of events $\hat{d}_k$ at each time T_k and then the estimated number of patients at risk $\hat{n}_{k+1}$ at time T_{k+1}:

Initialize $\hat{n}_{lower_i} \leftarrow nrisk_i$

for $k = lower_i$ **to** $upper_i$ **do**

  if $i = 1$ ***and*** $k = lower_i$ **then**

    $k' \leftarrow 1$

    $\hat{d}_k \leftarrow 0$

    $\hat{S}_k^{KM} \leftarrow 1$

  else

    $\hat{d}_k \leftarrow \hat{n}_k * (1 - S_k / \hat{S}_{k'}^{KM})$

    $\hat{S}_k^{KM} \leftarrow \hat{S}_{k'}^{KM} * (1 - \hat{d}_k / \hat{n}_k)$

  end if

  $\hat{n}_{k+1} \leftarrow \hat{n}_k - \hat{d}_k - \hat{cen}_k$

  if $\hat{d}_k \neq 0$ **then**

    Update the last index $k' \leq k$ where an event occurred:

    $k' \leftarrow k$

end for

Step 4:

if $\hat{nrisk}_{i+1} \neq nrisk_{i+1}$ **then**

  Update the estimated number of censor times in the i-th interval:

  $\hat{ncensor}_i \leftarrow \hat{ncensor}_i + (\hat{n}_{upper_i+1} - nrisk_{i+1})$

repeat

  Steps 2–4

until $\hat{nrisk}_{i+1}$ matches $nrisk_{i+1}$

Step 5:

do Steps 1–4 for $i = 1$ to $nint - 1$

Step 6:

Calculate the number censored for the last interval:

$$\hat{ncensor}_{nint} \leftarrow \min\left(\frac{T_{upper_{nint}} - T_{lower_{nint}}}{T_{upper_{nint-1}} - T_{lower_1}} * \sum_{i=1}^{nint-1} \hat{ncensor}_i \ , \ nrisk_{nint} \right)$$

do Steps 2 and **3** for the last interval

(continues...)

---

if the total number of events *totevents* is provided **then**

Step 7:

Calculate the estimated total number of events up until the beggining of the last interval:

$$\sum_{k=1}^{upper_{nint-1}} \hat{d}_k$$

if $\sum_{k=1}^{upper_{nint-1}} \hat{d}_k \geq totevents$ **then**

for $k = lower_{nint}$ **to** $upper_{nint}$ **do**

$\hat{d}_k \leftarrow 0$

$\hat{cen}_k \leftarrow 0$

$\hat{n}_k \leftarrow n_{upper_{nint-1}}$

end for

else if $\sum_{k=1}^{upper_{nint-1}} \hat{d}_k < totevents$ **then**

Step 8:

Update the estimated number of censor times in the last interval:

$$\hat{ncensor}_{nint} \leftarrow \hat{ncensor}_{nint} + (\sum_{k=1}^{upper_{nint}} \hat{d}_k - totevents)$$

repeat

Steps 2, 3 and 8

until $\sum_{k=1}^{upper_{nint-1}} \hat{d}_k = totevents$ ***or*** ($\sum_{k=1}^{upper_{nint-1}} \hat{d}_k < totevents$ ***and*** $\hat{ncensor}_{nint} = 0$)

end if

Table D.1: Generated IPD for the Olaparib cohort. Recurrence status 1 = recurred and 0 = censored.

#	Event Time (months)	Rec. Status	#	Event Time (months)	Rec. Status	#	Event Time (months)	Rec. Status
1	0.12	1	45	11.03	1	89	24.40	1
2	1.18	0	46	11.11	0	90	24.56	1
3	1.18	0	47	11.19	1	91	24.56	1
4	1.18	0	48	11.77	1	92	24.69	0
5	1.18	0	49	12.01	1	93	24.81	1
6	1.18	0	50	12.18	1	94	24.81	1
7	1.18	0	51	12.18	1	95	24.89	0
8	1.18	0	52	12.50	1	96	24.97	1
9	1.18	0	53	13.08	1	97	25.10	0
10	1.18	0	54	13.24	1	98	25.22	1
11	1.18	0	55	13.45	1	99	25.22	1
12	2.33	0	56	13.45	1	100	25.22	1
13	2.41	1	57	13.90	1	101	25.38	0
14	2.41	1	58	13.90	1	102	25.83	0
15	2.50	0	59	13.98	1	103	25.83	0
16	2.66	1	60	14.55	1	104	26.16	0
17	2.66	1	61	15.21	1	105	26.20	1
18	2.66	1	62	15.95	1	106	26.20	1
19	2.74	0	63	16.11	1	107	26.33	1
20	2.95	0	64	16.11	1	108	26.55	0
21	3.73	1	65	16.61	1	109	26.55	0
22	3.93	0	66	16.77	1	110	26.94	1
23	4.14	1	67	16.93	1	111	27.11	0
24	4.55	1	68	18.49	1	112	27.27	1
25	4.63	0	69	18.74	1	113	27.27	1
26	4.71	1	70	19.19	0	114	27.68	1
27	5.33	0	71	19.39	1	115	27.68	1
28	5.53	1	72	19.39	1	116	27.84	1
29	5.53	1	73	19.39	1	117	28.21	0
30	5.61	1	74	20.30	1	118	28.58	1
31	5.61	1	75	20.30	1	119	28.66	1
32	6.06	1	76	20.50	0	120	28.66	1
33	7.62	0	77	22.02	1	121	29.24	0
34	7.91	1	78	22.02	1	122	29.48	1
35	7.99	1	79	22.02	1	123	30.06	1
36	7.99	1	80	22.10	1	124	30.30	1
37	8.24	1	81	22.22	0	125	31.04	1
38	8.24	1	82	22.35	1	126	31.90	0
39	8.40	1	83	22.35	1	127	32.77	1
40	8.40	1	84	22.35	1	128	33.09	1
41	9.47	1	85	22.76	0	129	33.18	0
42	9.63	1	86	22.84	1	130	33.42	0
43	10.08	0	87	23.09	0	131	33.42	0
44	10.95	1	88	24.03	0	132	33.75	0

#	Event Time (months)	Rec. Status
133	33.75	0
134	34.00	0
135	34.08	1
136	34.32	0
137	34.32	0
138	34.32	0
139	34.69	0
140	34.82	1
141	35.02	0
142	35.02	0
143	35.02	0
144	35.39	0
145	35.39	0
146	35.64	0
147	35.72	1
148	35.84	0
149	36.09	0
150	36.29	0
151	36.38	1
152	36.38	1
153	36.58	0
154	36.58	0
155	36.58	0
156	36.95	0
157	36.95	0
158	37.28	0
159	37.28	0
160	37.28	0
161	37.73	0
162	37.73	0
163	37.73	0
164	37.73	0
165	38.18	0
166	38.18	0
167	38.18	0
168	38.47	0
169	38.47	0
170	38.71	0
171	38.71	0
172	38.96	0
173	39.08	1
174	39.16	0
175	39.16	0
176	39.49	0
177	39.49	0
178	39.49	0
179	39.49	0
180	39.49	0
181	39.49	0
182	39.94	0
183	39.94	0
184	39.94	0
185	39.94	0
186	39.94	0
187	39.94	0
188	40.27	0
189	40.27	0
190	40.27	0
191	40.52	0
192	40.52	0
193	40.52	0
194	40.76	0
195	40.76	0
196	40.76	0
197	40.89	1
198	41.01	0
199	41.01	0
200	41.01	0
201	41.01	0
202	41.26	0
203	41.26	0
204	41.26	0
205	41.46	0
206	41.46	0
207	41.54	1
208	41.63	0
209	41.63	0
210	41.79	0
211	41.79	0
212	41.91	0
213	41.95	1
214	42.04	0
215	42.04	0
216	42.12	1
217	42.45	0
218	42.98	0
219	43.39	0
220	43.68	0
221	43.76	1
222	44.04	0
223	44.58	0
224	45.03	0
225	45.89	0
226	45.89	0
227	45.89	0
228	45.89	0
229	45.89	0
230	45.89	0
231	45.89	0
232	45.89	0
233	45.89	0
234	46.71	0
235	46.71	0
236	46.71	0
237	47.00	0
238	47.00	0
239	47.20	0
240	47.37	0
241	47.65	0
242	47.65	0
243	47.65	0
244	48.76	0
245	48.76	0
246	48.76	0
247	48.76	0
248	48.76	0
249	48.76	0
250	48.76	0
251	48.76	0
252	48.76	0
253	48.76	0
254	48.76	0
255	48.76	0
256	48.76	0
257	51.02	0
258	53.81	0
259	53.81	0
260	53.81	0

Table D.2: Generated IPD for the placebo cohort. Recurrence status 1 = recurred and 0 = censored.

#	Event Time (months)	Rec. Status	#	Event Time (months)	Rec. Status	#	Event Time (months)	Rec. Status
1	1.22	0	45	8.48	1	89	22.18	1
2	2.33	1	46	8.48	1	90	22.84	0
3	2.50	1	47	8.57	1	91	24.89	1
4	2.66	1	48	8.73	1	92	25.05	1
5	2.66	1	49	8.81	1	93	27.76	1
6	2.66	1	50	9.22	1	94	30.39	1
7	2.66	1	51	9.30	1	95	30.55	1
8	2.66	1	52	10.70	1	96	30.63	1
9	2.70	0	53	10.95	1	97	30.63	1
10	2.74	1	54	10.95	1	98	30.63	1
11	2.82	1	55	11.03	1	99	31.49	0
12	2.82	1	56	11.03	1	100	32.48	0
13	2.91	1	57	11.11	1	101	33.50	1
14	4.22	1	58	11.11	1	102	34.65	0
15	4.46	1	59	11.19	1	103	34.65	0
16	4.96	1	60	11.27	1	104	37.61	0
17	5.08	0	61	11.40	1	105	37.61	0
18	5.37	1	62	11.40	1	106	38.67	1
19	5.37	1	63	11.40	1	107	38.84	1
20	5.45	1	64	11.40	1	108	39.86	0
21	5.45	1	65	11.60	1	109	39.86	0
22	5.53	1	66	11.93	1	110	41.01	0
23	5.53	1	67	13.82	1	111	41.30	0
24	5.53	1	68	13.98	1	112	41.54	0
25	5.53	1	69	13.98	1	113	41.63	1
26	5.61	1	70	13.98	1	114	41.63	1
27	5.78	1	71	13.98	1	115	41.75	0
28	5.78	1	72	13.98	1	116	43.14	0
29	6.52	1	73	14.14	1	117	43.14	0
30	6.91	0	74	14.55	1	118	43.14	0
31	7.54	1	75	14.80	1	119	43.14	0
32	7.66	1	76	15.38	1	120	43.14	0
33	7.75	1	77	16.52	1	121	43.14	0
34	7.91	1	78	16.69	1	122	43.14	0
35	8.07	1	79	18.33	1	123	43.14	0
36	8.16	0	80	19.11	0	124	43.14	0
37	8.24	1	81	19.23	1	125	43.14	0
38	8.24	1	82	19.31	1	126	45.44	0
39	8.32	1	83	19.39	1	127	47.53	0
40	8.40	1	84	20.34	0	128	47.86	1
41	8.40	1	85	21.28	1	129	47.86	1
42	8.40	1	86	22.02	1	130	48.88	0
43	8.40	1	87	22.02	1	131	52.46	0
44	8.40	1	88	22.02	1			

Bibliography

[1] L. M. Santana, S. Ganesan, and G. Bhanot. A quasi birth-and-death model for tumor recurrence. *J. Theor. Biol.*, 480:175–191, 2019. URL https://doi.org/10.1016/j.jtbi.2019.07.017.

[2] R. A. Weinberg. *The Biology of Cancer*. Garland Science, Taylor & Francis Group, New York, USA, 2014.

[3] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 44 (5):646–674, 2011. URL https://doi.org/10.1016/j.cell.2011.02.013.

[4] K. C. Cheng and L. A. Loeb. Genomic instability and tumor progression: Mechanistic considerations. *Adv. Cancer Res.*, 60:121–156, 1992. URL https://doi.org/10.1016/S0065-230X(08)60824-6.

[5] M. Kastan and J. Bartek. Cell-cycle checkpoints and cancer. *Nature*, 432:316–323, 2004. URL https://doi.org/10.1038/nature03097.

[6] J. Massagué. G1 cell-cycle control and cancer. *Nature*, 432:298–306, 2004. URL https://doi.org/10.1038/nature03094.

[7] B. Vogelstein and K. W. Kinzler. The multistep nature of cancer. *Trends Genet.*, 9 (4):138–141, 1993. URL https://doi.org/10.1016/0168-9525(93)90209-Z.

[8] B. Vogelstein and K. Kinzler. Cancer genes and the pathways they control. *Nat. Med.*, 10:789–799, 2004. URL https://doi.org/10.1038/nm1087.

[9] J. H. Goldie and A. J. Coldman. *Drug Resistance in Cancer – Mechanisms and Models*. Cambridge Univ. Press, New York, USA, 2009.

[10] A. Levine, J. Momand, and C. Finlay. The p53 tumour suppressor gene. *Nature*, 351: 453–456, 1991. URL https://doi.org/10.1038/351453a0.

[11] C. J. Sherr and F. McCormick. The rb and p53 pathways in cancer. *Cancer Cell*, 2 (2):103–112, 2002. URL https://doi.org/10.1016/S1535-6108(02)00102-2.

[12] G. Selivanova and K. G. Wiman. p53: A cell cycle regulator activated by dna damage. *Adv. Cancer Res.*, 66:143–180, 1995. URL https://doi.org/10.1016/S0065-230X(08)60253-5.

[13] S. Jackson and J. Bartek. The dna-damage response in human biology and disease. *Nature*, 461:1071–1078, 2009. URL https://doi.org/10.1038/nature08467.

[14] T. D. Halazonetis, V. G. Gorgoulis, and J. Bartek. An oncogene-induced dna damage model for cancer development. *Science*, 319(5868):1352–1355, 2008. URL https://doi.org/10.1126/science.1140735.

[15] G. Martin. The hunting of the src. *Nat. Rev. Mol. Cell Biol.*, 2:467–475, 2001. URL https://doi.org/10.1038/35073094.

[16] D. Stehelin, H. Varmus, J. Bishop, et al. Dna related to the transforming gene(s) of avian sarcoma viruses is present in normal avian dna. *Nature*, 260:170–173, 1976. URL https://doi.org/10.1038/260170a0.

[17] W. Hayward, B. Neel, and S. Astrin. Activation of a cellular onc gene by promoter insertion in alv-induced lymphoid leukosis. *Nature*, 290:475–480, 1981. URL https://doi.org/10.1038/290475a0.

[18] P. Vogt. Retroviral oncogenes: a historical primer. *Nat. Rev. Cancer*, 12:639–648, 2012. URL https://doi.org/10.1038/nrc3320.

[19] R. S. Herbst. Review of epidermal growth factor receptor biology. *Int. J. Radiat. Oncol. Biol. Phys.*, 59(2):S21–S26, 2004. URL https://doi.org/10.1016/j.ijrobp.2003.11.041.

[20] J. Downward, Y. Yarden, E. Mayes, et al. Close similarity of epidermal growth factor receptor and v-erb-b oncogene protein sequences. *Nature*, 307:521–527, 1984. URL https://doi.org/10.1038/307521a0.

[21] H. Zhang, A. Berezov, Q. Wang, et al. Erbb receptors: from oncogenes to targeted cancer therapies. *J. Clin. Investig.*, 117:2051–2058, 2007. URL https://doi.org/10.1172/JCI32278.

[22] S. A. Aaronson. Growth factors and cancer. *Science*, 254(5035):1146–1153, 1991. URL https://doi.org/10.1126/science.1659742.

[23] J. Desgrosellier and D. Cheresh. Integrins in cancer: biological implications and therapeutic opportunities. *Nat. Rev. Cancer*, 10:9–22, 2010. URL https://doi.org/10.1038/nrc2748.

[24] M. Waterfield, G. Scrace, N. Whittle, et al. Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus. *Nature*, 304:35–39, 1983. URL https://doi.org/10.1038/304035a0.

[25] R. F. Doolittle, M. W. Hunkapiller, L. E. Hood, et al. Simian sarcoma virus oncogene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. *Science*, 221(4607):275–277, 1983. URL https://doi.org/10.1126/science.6304883.

[26] A. Karnoub and R. Weinberg. Ras oncogenes: split personalities. *Nat. Rev. Mol. Cell Biol.*, 9:517–531, 2008. URL https://doi.org/10.1038/nrm2438.

[27] A. G. Knudson. Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.*, 68(4):820–823, 1971. URL https://doi.org/10.1073/pnas.68.4.820.

[28] C. J. Sherr. Principles of tumor suppression. *Cell*, 116(2):235–246, 2004. URL https://doi.org/10.1016/S0092-8674(03)01075-4.

[29] J. Hayes, P. P. Peruzzi, and S. Lawler. Micrornas in cancer: biomarkers, functions and therapy. *Trends. Mol. Med.*, 20(8):460–469, 2014. URL https://doi.org/10.1016/j.molmed.2014.06.005.

[30] B. Alberts et al. *Molecular Biology of the Cell*. Garland Science, Taylor & Francis Group, New York, USA, 2015.

[31] M. Malumbres and M. Barbacid. To cycle or not to cycle: a critical decision in cancer. *Nat. Rev. Cancer*, 1:222–231, 2001. URL https://doi.org/10.1038/35106065.

[32] E. Baehrecke. How death shapes life during development. *Nat Rev Mol Cell Biol*, 3: 779–787, 2002. URL https://doi.org/10.1038/nrm931.

[33] B. Vogelstein, D. Lane, and A. Levine. Surfing the p53 network. *Nature*, 408:307–310, 2000. URL https://doi.org/10.1038/35042675.

[34] D. Lane and L. Crawford. T antigen is bound to a host protein in sy40-transformed cells. *Nature*, 278:261–263, 1979. URL https://doi.org/10.1038/278261a0.

[35] D. I. H. Linzer and A. J. Levine. Characterization of a 54k dalton cellular sv40 tumor antigen present in sv40-transformed cells and uninfected embryonal carcinoma cells. *Cell*, 17(1):43–52, 1979. URL https://doi.org/10.1016/0092-8674(79)90293-9.

[36] C. Sherr. The ink4a/arf network in tumour suppression. *Nat. Rev. Mol. Cell Biol.*, 2:731–737, 2001. URL https://doi.org/10.1038/35096061.

[37] T. Abbas and A. Dutta. p21 in cancer: intricate networks and multiple activities. *Nat. Rev. Cancer*, (9):400–414, 2009. URL https://doi.org/10.1038/nrc2657.

[38] J. Plati, O. Bucur, and R. Khosravi-Far. Dysregulation of apoptotic signaling in

cancer: Molecular mechanisms and therapeutic opportunities. *J. Cell. Biochem.*, 104: 1124–1149, 2008. URL https://doi.org/10.1002/jcb.21707.

[39] B. Chabner and T. Roberts. Chemotherapy and the war on cancer. *Nat. Rev. Cancer*, 5:65–72, 2005. URL https://doi.org/10.1038/nrc1529.

[40] American Cancer Society. *Cancer Facts & Figures 2019*. American Cancer Society, Atlanta, 2019. URL https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf.

[41] P. M. Altrock, L. L. Liu, and F. Michor. The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer*, 15:730–745, 2015. URL https://doi.org/10.1038/nrc4029.

[42] A. Swierniak, A. Polanski, and M. Kimmel. Optimal control problems arising in cell-cycle-specific cancer chemotherapy. *Cell Prolif.*, 29:117–139, 1996. URL https://doi.org/10.1046/j.1365-2184.1996.00995.x.

[43] A. Wu, D. Liao, V. Kirilin, et al. Cancer dormancy and criticality from a game theory perspective. *Cancer Converg.*, 2(1), 2018. URL https://doi.org/10.1186/s41236-018-0008-0.

[44] J. S. Sartakhti, M. H. Manshaei, and M. Sadeghi. Mmp-timp interactions in cancer invasion: An evolutionary game-theoretical framework. *J. Theor. Biol.*, 412:17–26, 2017. URL https://doi.org/10.1016/j.jtbi.2016.09.019.

[45] N. Komarova. Stochastic modeling of drug resistance in cancer. *J. Theor. Biol.*, 239 (3):351–366, 2006. URL https://doi.org/10.1016/j.jtbi.2005.08.003.

[46] N. L. Komarova and D. Wodarz. Stochastic modeling of cellular colonies with quiescence: An application to drug resistance in cancer. *Theor. Popul. Biol.*, 72(4): 523–538, 2007. URL https://doi.org/10.1016/j.tpb.2007.08.003.

[47] J. Foo and Michor F. Evolution of resistance to anti-cancer therapy during general dosing schedules. *J. Theor. Biol.*, 263(2):179–188, 2010. URL https://doi.org/10.1016/j.jtbi.2009.11.022.

[48] T. Antal and P. L. Krapivsky. Exact solution of a two-type branching process: models of tumor progression. *J. Stat. Mech. Theory Exp.*, (P08018), 2011. URL https://doi.org/10.1088/1742-5468/2011/08/P08018.

[49] A.J. Coldman and J.H. Goldie. A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Math. Biosci.*, 65(2):291–307, 1983. URL https://doi.org/10.1016/0025-5564(83)90066-4.

[50] B. G. Birkhead, E. M. Rankin, S. Gallivan, L. Dones, and R. D. Rubens. A mathematical model of the development of drug resistance to cancer chemotherapy. *Eur. J. Cancer Clin. Oncol.*, 23:1421–1427, 1987. URL https://doi.org/10.1016/0277-5379(87)90133-7.

[51] J. C. Panetta and J. Adam. A mathematical model of cycle-specific chemotherapy. *Math. Comput. Modelling*, 22(2):67–82, 1995. URL https://doi.org/10.1016/0895-7177(95)00112-F.

[52] J. C. Panetta. A mathematical model of breast and ovarian cancer treated with paclitaxel. *Math. Biosci.*, 146(2):89–113, 1997. URL https://doi.org/10.1016/S0025-5564(97)00077-1.

[53] E. Gaffney. The application of mathematical modelling to aspects of adjuvant chemotherapy scheduling. *J. Math. Biol.*, 48:375–422, 2004. URL https://doi.org/10.1007/s00285-003-0246-2.

[54] V. T. DeVita and P. S. Schein. The use of drugs in combination for the treatment of cancer: rationale and results. *N. Engl. J. Med.*, 288(19):998–1006, 1973. URL https://doi.org/10.1056/NEJM197305102881905.

[55] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. L. Ebos, L. Hlatky, and P. Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Comput. Biol.*, 10(e1003800), 2014. URL https://doi.org/10.1371/journal.pcbi.1003800.

[56] L. Norton, R. Simon, H. D. Brereton, and A. E. Bogden. Predicting the course of gompertzian growth. *Nature*, 264:542–545, 1976. URL https://doi.org/10.1038/264542a0.

[57] J. West and P. K. Newton. Chemotherapeutic dose scheduling based on tumor growth rates provides a case for low-dose metronomic high-entropy therapies. *Cancer Res.*, 77 (23):6717–6728, 2017. URL https://doi.org/10.1158/0008-5472.CAN-17-1120.

[58] S. E. Clare, F. Nakhlis, and J. C. Panetta. Molecular biology of breast cancer metastasis: The use of mathematical models to determine relapse and to predict response to chemotherapy in breast cancer. *Breast Cancer Res.*, 2(6):430–435, 2000. URL https://doi.org/10.1186/bcr90.

[59] M. W. Retsky, D. E. Swartzendruber, R. H. Wardwell, and P. D. Bame. Is gompertzian or exponential kinetics a valid description of individual human cancer growth? *Med. Hypotheses*, 33(2):95–106, 1990. URL https://doi.org/10.1016/0306-9877(90)90186-I.

[60] M. W. Retsky, D. E. Swartzendruber, P. D. Bame, and R. H. Wardwell. Computer model challenges breast cancer treatment strategy. *Cancer Investig.*, 12(6):559–567, 1994. URL https://doi.org/10.3109/07357909409023040.

[61] M. W. Retsky, R. Demicheli, D. E. Swartzendruber, P. D. Bame, R. H. Wardwell, G. Bonadonna, J. F. Speer, and P. Valagussa. Computer simulation of a breast cancer metastasis model. *Breast Cancer Res. Treat.*, 45:193–202, 1997. URL https://doi.org/10.1023/A:1005849301420.

[62] R. Demicheli, R. Miceli, C. Brambilla, L. Ferrari, Moliterni A., M. Zambetti, P. Valagussa, and G. Bonadonna. Comparative analysis of breast cancer recurrence risk for patients receiving or not receiving adjuvant cyclophosphamide, methotrexate, fluorouracil (cmf). data supporting the occurrence of 'cures'. *Breast Cancer Res. Treat.*, 53:209–215, 1999. URL https://doi.org/10.1023/A:1006134702484.

[63] M. Collado and M. Serrano. Senescence in tumours: evidence from mice and humans. *Nat. Rev. Cancer*, 10:51–57, 2010. URL https://doi.org/10.1038/nrc2772.

[64] J. F. Speer, V. E. Petrosky, M. W. Retsky, and R. H. Wardwell. A stochastic numerical model of breast cancer growth that simulates clinical data. *Cancer Res.*, 44:4124–4130, 1984. URL https://cancerres.aacrjournals.org/content/44/9/4124.

[65] P. Alexander. 2nd gordon hamilton fairley lecture. need for new approaches to the treatment of patients in clinical remission, with special reference to acute myeloid leukaemia. *Br. J. Cancer*, 46:151–159, 1982. URL https://doi.org/10.1038/bjc.1982.178.

[66] Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Adjuvant chemotherapy in oestrogen-receptor-poor breast cancer: patient-level meta-analysis of randomised trials. *Lancet*, 371(9606):29–40, 2008. URL https://doi.org/10.1016/S0140-6736(08)60069-0.

[67] D. F. Hayes, A. D. Thor, L. G. Dressler, et al. Her2 and response to paclitaxel in node-positive breast cancer. *N. Engl. J. Med.*, 357:1496–1506, 2007. URL https://doi.org/10.1056/NEJMoa071167.

[68] N. W. Wilkinson, G. Yothers, S. Lopa, J. P. Costantino, N. J. Petrelli, and N. Wolmark. Long-term survival results of surgery alone versus surgery plus 5-fluorouracil and leucovorin for stage ii and stage iii colon cancer: Pooled analysis of nsabp c-01

through c-05. a baseline from which to compare modern adjuvant trials. *Ann. Surg. Oncol.*, 17:959–966, 2010. URL https://doi.org/10.1245/s10434-009-0881-y.

[69] N. Gautam. *Analysis of Queues: Methods and Applications*. CRC Press, Philadelphia, PA, 2012.

[70] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA, 1999.

[71] R. A. Usmani. Inversion of jacobi's tridiagonal matrix. *Computers Math. Applic.*, 27 (8):59–66, 1994. URL https://doi.org/10.1016/0898-1221(94)90066-3.

[72] Y. Huang and W. F. McColl. Analytical inversion of general tridiagonal matrices. *J. Phys. A: Math. Gen.*, 30:7919–7933, 1997. URL https://doi.org/10.1088/0305-4470/30/22/026.

[73] J. P. Kharoufeh. *Level-dependent quasi-birth-and-death processes*. Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, Inc., Hoboken, NJ, 2011. URL https://doi.org/10.1002/9780470400531.eorms0460.

[74] V. Ramaswami and P. G. Taylor. Some properties of the rate operators in level dependent quasi-birth-and-death processes with a countable number of phases. *Stochastic Models*, 12:143–164, 1996. URL https://doi.org/10.1080/15326349608807377.

[75] L. Bright and P. G. Taylor. Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Stochastic Models*, 11(3):497–525, 1995. URL https://doi.org/10.1080/15326349508807357.

[76] L. S. T. Ho, J. Xu, F. W. Crawford, V. N. Minin, and M. A. Suchard. Birth/birth-death processes and their computable transition probabilities with biological applications. *J. Math. Biol.*, 76:911–944, 2018. URL https://doi.org/10.1007/s00285-017-1160-3.

[77] J. Abate and W. Whitt. Transient behavior of the m/m/1 queue via laplace transforms. *Adv. Appl. Prob.*, 20:145–178, 1988. URL https://doi.org/10.2307/1427274.

[78] J. Abate and W. Whitt. Transient behavior of the m/m/1 queue: Starting at the origin. *Queueing Syst.*, 2(41), 1987. URL https://doi.org/10.1007/BF01182933.

[79] J. Abate and W. Whitt. Approximations for the m/m/1 busy-period distribution. In J. W. Cohen, O. J. Boxma, and R. Syski, editors, *Queueing Theory and its Applications – Liber Amicorum*, pages 149–191. North-Holland, Amsterdam, 1988.

[80] P. Leguesdron, J. Pellaumail, G. Rubino, and B. Sericola. Transient analysis of the m/m/1 queue. *Adv. Appl. Prob.*, 25:702–713, 1993. URL https://doi.org/10.2307/1427531.

[81] O. P. Sharma and B. D. Bunday. A simple formula for the transient state probabilities of an m/m/1/∞ queue. *Optimization*, 40:79–84, 1997. URL https://doi.org/10.1080/02331939708844299.

[82] O. P. Sharma and N. S. K. Nair. Transient analysis of finite state birth and death process with absorbing boundary states. *Stoch. Anal. Appl.*, 14(5):565–589, 1996. URL https://doi.org/10.1080/07362999608809457.

[83] H. Risken. *The Fokker-Planck equation: methods of solution and applications*. Springer, Berlin, 1984.

[84] S. Iyer-Biswas and A. Zilman. First-passage processes in cellular biology. *Adv. Chem. Phys.*, pages 261–306, 2016. URL https://doi.org/10.1002/9781119165156.ch5.

[85] G. H. Weiss. First passage time problems in chemical physics. *Adv. Chem. Phys.*, 13: 1–18, 1967. URL https://doi.org/10.1002/9780470140154.ch1.

[86] T. Chou and M. R. D'Orsogna. First passage problems in biology. In R. Metzler, G. Oshanin, and S. Redner, editors, *First-Passage Phenomena and Their Applications*, pages 306–345. World Scientific, 2014. URL https://doi.org/10.1142/9789814590297_0013.

[87] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, Berlin, 1985.

[88] S. Redner. *A Guide to First-Passage Processes*. Cambridge Univ. Press, Cambridge, UK, 2001.

[89] The cancer genome atlas. https://cancergenome.nih.gov and https://portal.gdc.cancer.gov.

[90] M. H. Holmes. *Introduction to Perturbation Methods*. Springer-Verlag, New York, 1995.

[91] Z. Liu, H. Lou, K. Xie, H. Wang, N. Chen, O. M. Aparicio, M. Q. Zhang, R. Jiang, and T. Chen. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nat. Commun.*, 8(22), 2017. URL https://doi.org/10.1038/s41467-017-00039-z.

[92] Z. Ji and H. Ji. Tscan: Pseudo-time reconstruction and evaluation in single-cell rna-seq analysis. *Nucleic Acids Res.*, 44(13):e117, 2016. URL https://doi.org/10.1093/nar/gkw430.

[93] M. S. Kowalczyk, I. Tirosh, D. Heckl, et al. Single cell rna-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.*, 25:1860–1872, 2015. URL https://doi.org/10.1101/gr.192237.115.

[94] C. R. Doering, K. V. Sargsyan, and L. M. Sander. Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the fokker-planck

approximation. *Multiscale Model. Simul.*, 3(2):283–299, 2005. URL https://doi.org/10.1137/030602800.

[95] F. W. Crawford, L. S. T. Ho, and M. A. Suchard. Computational methods for birth-death processes. *WIREs Comput Stat.*, (e1423), 2018. URL https://doi.org/10.1002/wics.1423.

[96] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, 53(282):457–481, 1958. URL https://doi.org/10.2307/2281868.

[97] L. Y. Han, V. Karavasilis, T. v. Hagen, et al. Doubling time of serum ca125 is an independent prognostic factor for survival in patients with ovarian cancer relapsing after first-line chemotherapy. *Eur. J. Cancer*, 46(8):1359–1364, 2010. URL https://doi.org/10.1016/j.ejca.2010.02.012.

[98] K. Moore, N. Colombo, Scambia G., et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N. Engl. J. Med.*, 379(26):2495–2505, 2018. URL https://doi.org/10.1056/NEJMoa1810858.

[99] M. J. O'Connor. Targeting the dna damage response in cancer. *Mol. Cell*, 60(4):547–560, 2015. URL https://doi.org/10.1016/j.molcel.2015.10.040.

[100] H. Farmer, N. McCabe, C. Lord, et al. Targeting the dna repair defect in brca mutant cells as a therapeutic strategy. *Nature*, 434:917–921, 2005. URL https://doi.org/10.1038/nature03445.

[101] P. Guyot, A. Ades, M.J. Ouwens, et al. Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves. *BMC Med. Res. Methodol.*, 12(9), 2012. URL https://doi.org/10.1186/1471-2288-12-9.

[102] Z. Wang, J. Gao, J. Zhou, H. Liu, and C. Xu. Olaparib induced senescence under

p16 or p53 dependent manner in ovarian cancer. *J. Gynecol. Oncol.*, 30(2):e26, 2019. URL https://doi.org/10.3802/jgo.2019.30.e26.

[103] G. Caravagna, Y. Giarratano, D. Ramazzotti, et al. Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nat. Methods*, 15:707–714, 2018. URL https://doi.org/10.1038/s41592-018-0108-x.

[104] R. Durrett. Branching process models of cancer. In *Branching Process Models of Cancer*, volume 1.1, pages 1–63. Springer, Cham, 2015. URL https://doi.org/10.1007/978-3-319-16065-8_1.

[105] I. Bozic and M. A. Nowak. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc. Natl. Acad. Sci. U.S.A.*, 111(45):15964–15968, 2014. URL https://doi.org/10.1073/pnas.1412075111.

[106] I. Bozic, J. G. Reiter, B. Allen, T. Antal, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife*, 2:e00747, 2013. URL https://doi.org/10.7554/eLife.00747.

[107] J. Grasman and R. HilleRisLambers. On local extinction in a metapopulation. *Ecol. Model.*, 103(1):71–80, 1997. URL https://doi.org/10.1016/S0304-3800(97)00079-3.