Qualitative and Quantitative Approaches to Modelling Tyrosine Kinase Inhibitor Treatment and Resistance in Non-Small Cell Lung Cancer



Emma Carrick Smith
Wadham College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2021

To Mum, Dad, William, Robbie and Jonny

Statement of Originality

This thesis is submitted to the University of Oxford in fulfilment of the requirements of the degree of Doctor of Philosophy. No part of this thesis has already been submitted for any other qualification or degree at the University of Oxford or any other institution. This thesis is my own work and does not include work done through collaboration. All model development and analysis in chapters 4 and 5 were carried out by myself, although the in vitro and in vivo data, from which the models were developed, were collected by an experimentalist at AstraZeneca. All contributions that are not my own are clearly marked in the text.

Acknowledgements

I would like to thank my supervisors, Helen Byrne and Eamonn Gaffney. Helen, thank you for all your support and encouragement. Your contributions to my work and your tireless enthusiasm have allowed me to become a confident researcher. Eamonn, thank you for your positivity over the past few years. You have always seen the best in my work and have shown me the context of it throughout. The DPhil was hard to face at times due to the pandemic and I want to thank you both for going above and beyond with your support, academically and on a personal level. It has been inspirational and a privilege to work with you both in such a pivotal period of my life. I would also like to thank James Yates, my AstraZeneca supervisor. I enjoyed sharing ideas with you in our meetings and working with AstraZeneca has been very motivational. I would like to thank Ruth Baker for her contribution in research directions at confirmation. I thoroughly enjoyed exploring the directions you suggested.

Thanks to the Doctoral Training Centre and SABS for preparing me for my research and opening my eyes to the world of drug discovery and, mathematical and experimental biology. I thank EPSRC and AstraZeneca for their financial support that has allowed me to pursue this project.

Thank you to my friends at Oxford. Róisín, you have been a friend from the start and have supported me throughout, helping me through the highs and lows. You never fail to make me laugh and have helped me more than you know. I hope we are able to work closely with each other in the future. Josh, you have been such a huge inspiration to me. You have taught me to see the bigger picture when I was struggling. Our film clubs got me through lockdown and I thank you for never failing to make me laugh. Clara, you massive legend, I couldn't ask for a better friend. Living with you in Oxford has provided some of the best moments of my life. We have had so much fun together and you have made my time at Oxford so much better.

Thank you to Jonny, I am so lucky to have met you at Oxford and I am so excited for all the times together we have ahead of us. You have been my rock throughout this whole journey and I am grateful for you. I want to finish by saying a huge thank you to my family. Will and Robbie, I couldn't ask for better brothers. Mum and Dad, you have provided endless support and love. Your help throughout this process has been immeasurable and I am so very inspired by you both.

List of abbreviations

ABM	Agent-based model
ACF	Autocorrelation function
AIC	Akaike information criterion
ALK	Anaplastic lymphoma kinase
ATP	Adenosine triphosphate
BIC	Bayesian information criterion
BRAF	Rapidly accelerated fibrosarcoma oncogene homolog B
CML	Chronic myeloid leukaemia
CFDNA	Circulating free DNA
DAISY	Differential Algebra for Identifiability of SYstems
EFAST	Extended Fourier amplitude sensitivity test
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
FDA	Food and drug association
ILD	Interstitial lung disease
IQR	Interquartile range
KRAS	Kristen rat sarcoma gene
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MET	Mesenchymal epithelial transition factor
MCMC	Markov chain Monte Carlo
M-H	Metropolis-Hastings
MTD	Maximum tolerated dose
NLLS	Non-linear least squares
NSCLC	Non-small cell lung cancer
ODE	Ordinary differential equation
ORR	Objective response rate
PCR	Polymerase chain reaction
PFS	Progression free survival
PK-PD	Pharmacokinetic-pharmacodynamic
PPC	Posterior predictive check
PPD	Predictive posterior distribution
QVC	Quartile variation coefficient
SSA	Stochastic simulation algorithm
TKI	Tyrosine kinase inhibitor
TMB	Tumour mutation burden
VDT	Volume doubling time

Abstract

Non-small cell lung cancer (NSCLC) affects 80-85% of lung cancer patients. Tyrosine kinase inhibitors (TKIs) are targeted therapies that bind to NSCLC cells with specific mutations and inhibit their growth. However, responses to these therapies generally fail due to the emergence of drug resistance and the onset of adverse events which can include; interstitial lung disease (ILD), edema, hypothyroidism, nausea and diarrhoea. Mathematical models can increase our understanding of the biological mechanisms that determine whether a tumour is likely to respond or become resistant to treatment. When combined with experimental studies, such models can provide mechanistic insight into factors which influence the evolution to resistance and how specific drugs act. They can also aid experimental design and generate testable hypotheses concerning, for example, optimal dose regimens, scheduling protocols and combination therapies. In this thesis, we develop mathematical models of tumour growth and treatment response, and use them to determine tumour-specific factors that influence drug resistance. Further, by combining the models with preclinical data, we infer the biological properties of different NSCLC cell lines and characterise their responses to different TKIs.

In the first part of the thesis, we develop a mathematical model that describes the dynamics of four cell populations that differ in their levels of resistance to two TKIs. The model accounts for mutations, competition for resources and cellular interactions and is formulated as a system of time-dependent, ordinary differential equations. We use the model to show how tumour responses to TKI combinations depend on the rates at which the tumour cells mutate and proliferate. We then reformulate the model to account for stochastic effects that encapsulate the random nature of mutation events that occur with low probabilities. We find considerable variation in the predictions generated by the deterministic and stochastic modelling frameworks, highlighting the need for careful consideration of the modelling approach when predicting drug resistance and treatment response.

The second part of the thesis focuses on integrating mathematical models with experimental data in order to uncover the mechanisms driving NSCLC cells' responses to TKI treatment. We develop a suite of increasingly complex models that describe in vitro cell growth data. We use parameter estimation, information criteria/goodness-of-fit metrics and parameter identifiability analysis to investigate the identifiability of the model parameters given the preclinical data and to guide model selection. We extend the selected model to describe the response of NSCLC cells to TKI treatment. While this model is structurally identifiable, the existence of practically non-identifiable model parameters given the in vitro treatment data leads to uncertainty in the predictions of hidden variables. We use synthetic data to evaluate how experimental design (i.e. what variables are measured) affects certainty in parameter estimates and to generate hypotheses about the likely mechanisms of action of TKIs on NSCLC cells.

To summarise, in this thesis, we develop a series of mathematical models that increase our understanding of the biological mechanisms that drive the evolution to resistant tumours and determine tumour response to treatment. Our results highlight the benefit of theoretical approaches for distinguishing the growth dynamics of different NSCLC cell lines, providing mechanistic insight into the effects of different drugs on different cell lines and suggesting strategies for improving experimental design.

Contents

1 Introduction **1**

1.1 Biological review . . . 2

1.1.1 Overview of non-small cell lung cancer (NSCLC) . . . 2

1.1.2 Mutations driving NSCLC . . . 3

1.1.3 Drug resistance mechanisms in NSCLC . . . 4

1.1.4 Epidermal growth factor receptor TKIs . . . 7

1.1.5 Combination therapies applying TKIs . . . 10

1.2 Mathematical models of tumour growth, drug resistance and treatment response in cancer . . . 11

1.2.1 Simple tumour growth models . . . 12

1.2.2 Multi-compartment models of heterogeneous tumours . . . 15

1.2.3 Stochastic models of drug resistance and treatment response . . . 16

1.2.4 In vitro models of TKI response . . . 19

1.2.5 Sensitivity analysis . . . 21

1.2.6 Parameter inference . . . 21

1.2.7 Summary . . . 25

1.3 Thesis outline, aims and objectives . . . 25

2 A deterministic model of tumour growth and response to combination TKI treatment **30**

2.1 Introduction . . . 30

2.2 Untreated tumour growth model . . . 31

2.2.1 Model development . . . 31

2.2.2 Population-dependent birth rates . . . 34

2.2.3 Initial conditions . . . 37

2.2.4 Parameters . . . 37

2.3 Methods . . . 39

2.3.1 Global sensitivity analysis . . . 40

2.4 Results of untreated tumour growth . . . 40

2.4.1 Typical model simulations 41
2.4.2 Sensitivity analysis 41
2.4.3 Quantifying pre-existing resistance 45
2.4.4 Effect of cellular interactions on tumour composition over time . 47
2.5 Treated tumour growth model 49
2.5.1 Treatment model development 49
2.5.2 Treatment parameters 52
2.5.3 Combination treatment protocols 52
2.5.4 Typical simulations of treatment protocols 55
2.6 Results of first-line treatment response 56
2.6.1 Effect of key parameter values on first-line treatment response . . 56
2.7 Discussion 61

Appendices **66**
2.A Extended Fourier Amplitude Sensitivity Test (eFAST) 66
2.B Convergence testing on eFAST 71

3 A stochastic model of tumour growth and response to combination TKI treatment **73**
3.1 Introduction 73
3.2 Stochastic model 75
3.2.1 Event propensities 76
3.2.2 Parameters 78
3.2.3 Stochastic process simulation 78
3.3 Methods 82
3.3.1 Stochastic simulation algorithm 82
3.4 Results of untreated tumour growth with stochasticity 85
3.4.1 Typical stochastic simulations of untreated tumour growth 85
3.4.2 Quantifying variability in tumour cell dynamics 88
3.4.3 Quantifying variability in resistance at diagnosis 90
3.5 Results of treatment response with stochasticity 93
3.5.1 Typical stochastic simulations of treated tumour cell dynamics . 94
3.5.2 First and second-line treatment response 95
3.6 Discussion 98

Appendices **101**
3.A Convergence testing for identification of the optimal tolerance parameter, ϵ in the τ-leap method with adaptive τ calculation 101

4 Modelling preclinical tumour growth data **104**
4.1 Introduction . . . 104
4.2 Experimental data . . . 108
4.2.1 In vitro monoculture growth curves . . . 108
4.2.2 In vivo mouse xenograft tumour volume growth curves . . . 111
4.3 In vitro tumour growth models . . . 112
4.3.1 Model development . . . 112
4.3.2 Initial conditions . . . 113
4.3.3 Parameters . . . 114
4.3.4 Typical simulations of candidate ODE models . . . 115
4.4 Methods . . . 116
4.4.1 Bounds on model parameters for in vitro data . . . 117
4.4.2 Bounds of model parameters for in vivo data . . . 118
4.4.3 Parameter estimation . . . 119
4.4.4 Akaike and Bayesian information criterion for model comparison 122
4.4.5 Structural and practical identifiability . . . 122
4.4.6 Markov chain Monte Carlo (MCMC) . . . 128
4.5 Results . . . 131
4.5.1 Structural identifiability analysis . . . 131
4.5.2 Model calibration . . . 132
4.5.3 Practical identifiability analysis . . . 136
4.5.4 Model reduction . . . 139
4.5.5 Parameter estimation and model checks . . . 142
4.5.6 Practical identifiability of the model fit to in vivo data . . . 146
4.6 Discussion . . . 149

Appendices **153**
4.A Structural identifiability analysis . . . 153
A Richards growth equation . . . 153
B Generalised-logistic model . . . 153
4.B Markov chain Monte Carlo results, diagnostic testing and identifiability analysis . . . 154

5 Modelling preclinical data of NSCLC response to TKIs **161**
5.1 Introduction . . . 161
5.2 In vitro dose-response curves . . . 164
5.2.1 In vitro experiment protocol . . . 164
5.2.2 Discussion on in vitro data sets . . . 165

5.3 Model development 167
5.3.1 Initial conditions 169
5.4 Methods 169
5.4.1 Parameter bounds 170
5.4.2 Monte Carlo Simulations for calculating 95% confidence bounds of the model fit to data 172
5.5 Results 173
5.5.1 Richards growth equation fit to individual dose-response curves . 173
5.5.2 Steady-state analysis 176
5.5.3 Structural identifiability 176
5.5.4 Parameter estimation and practical identifiability 179
5.5.5 Implications of predicting unobserved variable trajectories 186
5.5.6 Relationships between estimated model parameters using profile likelihoods 189
5.5.7 Model reduction 190
5.5.8 Practical identifiability analysis and parameter estimation of the reduced treatment model 194
5.5.9 Experimental design 198
5.6 Discussion 199

Appendices **203**
5.A Supplementary results 203

6 Conclusions and Future Work **209**
6.1 Research summary and contextualisation 209
6.2 Future directions 216

References **218**

List of Figures

1.1.1 Left) Diagram of a tyrosine kinase inhibitor binding to a cancer cell. In the absence of drug, growth factors such as epidermal growth factor (EGF) bind to the receptor and signals are transduced across the cell surface, upregulating anti-apoptotic and cell proliferation pathways. Tyrosine kinase inhibitors bind to the receptor and block these signals (diagram credit: Cancer Research UK). Right) Diagram of the evolution of TKI treatment and resistance mechanisms that result in their failure (diagram credit: Murtuza et al. [101]). 7

2.2.1 Schematic of our mathematical model of tumour growth in the absence of treatment. Four cell populations are considered. Sensitive cells (green cells), are denoted by $S(t)$; cells that are sensitive to drug two and resistant to drug one (blue cells) are denoted by $R_1(t)$; cells that are sensitive to drug one and resistant to drug two (yellow cells) are denoted by $R_2(t)$; fully drug-resistant cells (red cells) are denoted by $R_{12}(t)$. The parameter μ denotes the proportion of cells that mutate per cell division from (S to R_1), (R_2 to R_{12}), (S to R_2) and (R_1 to R_{12}). S cells divide to produce S cells, ($S \to S + S$), R_1 cells ($S \to S + R_1$) and R_2 cells ($S \to S + R_2$) at rates, $b_S(1-2\mu)$, $b_S\mu$ and $b_S\mu$, respectively. S cells die at rate, d_S and the net growth rate of S cells is $r_S = b_S - d_S$. Similarly, R_1 cells divide to produce R_1 cells ($R_1 \to R_1 + R_1$) and R_{12} cells ($R_1 \to R_1 + R_{12}$) at rates, $b_{R_1}(1-\mu)$ and $b_{R_1}\mu$, respectively. R_1 cells die at rate, d_{R_1} and have a net growth rate of $r_{R_1} = b_{R_1} - d_{R_1}$. Similarly for R_2 cells. R_{12} cells divide to produce R_{12} cells at rate, $b_{R_{12}}$ and die at rate, $d_{R_{12}}$ and have a net growth rate, $r_{R_{12}} = b_{R_{12}} - d_{R_{12}}$. 32

2.2.2 Tumour doubling time (days) for a range of σ values. For a given value of σ, we simulate equations (2.2.1)-(2.2.4) where the birth and death rates are defined by equations (2.2.12)-(2.2.15). The doubling time is the time taken for the total tumour population to increase from (0.5×10^8) to 10^8 cells. Parameters: all parameters are fixed at the default values in Table 2.2.1, except for the carrying capacity, $K = 1.3 \times 10^{12}$ and σ, which is varied between 1.02 and 1.1. 38

2.4.1 Typical model solutions showing how cell numbers change over time for an untreated tumour. Key: line styles represent cell types; $S(t)$= block lines; $R_1(t)$=dashed lines; $R_{12}(t)$=dotted lines. Colours represent parameter values used where black, dark grey and light grey represent high, mid and low values of the varied parameter, respectively. Subplots: (a) Model solutions with default model parameters given in Table 2.2.1. (b) Model solutions with β values in [0.5, 1, 1.5] where light grey lines represent $\beta = 0.5$; dark grey; $\beta = 1$ and black; $\beta = 1.5$. The same colour scheme holds for the following plots and their respective parameter values; (c) δ values in [0.5, 1, 1.5]; (d) γ values in [0.5, 1, 1.5]; (e) μ values in $[10^{-8}, 10^{-6}, 10^{-4}]$; (f) σ values in [1.02, 1.05, 1.08]. In a given subplot, for non-varied parameters, default parameter values given in Table 2.2.1 are used. 42

2.4.2 Results from global sensitivity analysis, using eFAST, applied to the untreated tumour growth model. We implement eFAST with the number of samples, $N_S = 5000$ and the number of resamples, $N_R = 10$ (see appendix for details of eFAST). (a) Total-order sensitivity index, S_{Ti} for model parameters, over time. (b) First-order sensitivity index, S_i for model parameters, over time. In plots (a) and (b), lines represent mean sensitivity index of the sample ($N_R = 10$ samples in total) and the shaded regions represent two standard deviations about the mean sensitivity index for a given parameter. The legend corresponding to plots (a) and (b) is given to the right of subplot (b). In plots (a) and (b), when a parameter is no longer significant, the line corresponding to its sensitivity index is terminated. (c) First and total-order sensitivity indices for model parameters at diagnosis (10^8 cells). An asterisk, * signifies significance with a p-value < 0.01 compared to the distributions of first and total-order sensitivity indices corresponding to the dummy parameter (distributions are made up of $N_R = 10$ values in total). Error bars represent two standard deviations about the mean of the sensitivity indices. Model parameters, summarised in Table 2.2.1, are sampled from a uniform distribution with ranges also given in Table 2.2.1. 44

2.4.3 Predicted proportion of fully drug-resistant cells, $p_{R_{12}}$ at diagnosis (heatmap plots) and over time (lineplots): Heatmap plots: tumours are grown until the total tumour population reaches 10^8 cells (diagnosis) at which time, $p_{R_{12}}$ is measured. Dark blue and yellow heatmap colours represent low and high values of $p_{R_{12}}$, respectively, in which two key parameters from μ, δ, γ and σ, are varied simultaneously in a pairwise manner. White represents cell proportions less than 10^{-8}. The horizontal colour bar presents the range of $p_{R_{12}}$ values $[10^{-8}, 1]$. Line plots: describe $p_{R_{12}}$ over time in which one key parameter is varied in a single plot. The ranges of the parameters, μ, δ, γ and σ are given in Table 2.2.1. Yellow and blue lines in the line plots correspond to the maximum and minimum parameter values in the predefined range (see Table 2.2.1), respectively. When parameters are not varied, they take the default values given in Table 2.2.1. 46

2.4.4 Series of plots showing how the proportion of tumour cells changes over time as the parameters, β, γ and δ vary. Key: colours represent the proportions of cell types: p_S=green; p_{R_1}=blue; $p_{R_{12}}$=red. In each subplot we consider three values of δ: dotted lines = $\delta = 0.5$; dashed lines = $\delta = 1$; block lines = $\delta = 1.5$. Columns correspond to different values of $\gamma \in [0.5, 1, 1.5]$. Rows correspond to different values of $\beta \in [0.5, 1, 1.5]$. All tumours are grown for a period of 10 years. The legend is given in the top right figure. Parameters: $\alpha = 0.11(\text{day}^{-1})$, $\mu = 10^{-4}$, $K = 1.3 \times 10^{12}$ cells and $\sigma = 1.03$. 48

2.5.1 Typical simulations of tumour cell dynamics with first-line treatment applied at diagnosis. (a) Protocol A, (b) Protocol B. Key: S cells are presented by blue lines; R_1; green, R_2; yellow and R_{12}; orange. We note that R_1 and R_2 cells have the same dynamics, hence we observe only the solutions corresponding to R_1 cells (green). The cyan shaded regions indicate when treatment is applied. Black dashed lines represent the time of diagnosis, black solid line represents time of progression on or after (for protocol B only) first-line treatment. Parameters: the default parameter values given in Table 2.2.1 are used. Descriptions of protocols are given in Table 2.5.1. 55

2.6.1 Predicted time to progression (20% increase in total tumour population size from diagnosis), in years from the start of treatment (heatmap plots) and $p_{R_{12}}$ values over time (diagonal lineplots). Heatmaps: Heatmaps represent progression times as two key parameters in (μ, δ, γ, σ) are varied simultaneously. Below the diagonal presents results from tumours undergoing protocol A. Above the diagonal presents results from tumours undergoing protocol B. The colourbar presents a range of progression times, $[0, 25]$ years. Lineplots: $p_{R_{12}}$ over time in which one key parameter is varied in a single plot. Dashed lines indicate tumours undergoing high dose, pulsed treatment (Protocol B) and block lines indicate tumours undergoing continuous, moderate treatment (Protocol A). Ranges of key parameter values are given in Table 2.2.1 with yellow lines representing the maximum parameter value and blue, the minimum. When parameters are not varied, parameters take the default values given in Tables 2.2.1. 58

2.6.2 Predicted $p_{R_{12}}$ at progression (heatmap plots) and $p_{R_{12}}$ over time (lineplots). Heatmaps: Heatmap plots represent $p_{R_{12}}$ values at progression as two key parameters in the set (μ, δ, γ, σ) are varied simultaneously. Heatmaps below the diagonal present results from tumours undergoing protocol A. Heatmaps above the diagonal present results from tumours undergoing protocol B. The colourbar presents a range $p_{R_{12}}$ values at progression of $[10^{-8}, 1]$. White regions represent $p_{R_{12}} < 8.3 \times 10^{-9}$ $(1/(1.2 \times 10^8))$, when no R_{12} cells are present. Lineplots: $p_{R_{12}}$ over time in which one key parameter is varied in a single plot. Dashed lines indicate tumours undergoing high dose, pulsed TKI treatment (protocol B) and block lines represents tumours undergoing continuous moderate TKI treatment (protocol A). Ranges of key parameter values are given in Table 2.2.1 with yellow lines representing the maximum parameter value and blue, the minimum. When parameters are not varied they take the default values given in Table 2.2.1. 59

2.A.1 First and total-order sensitivity indices of simplified model parameters for $p_{R1}(1865)$. * presents significance at p-value < 0.01 when compared to the dummy parameter. Grey represents S_{Ti}, black represents S_i. $N_S = 4000$, $N_R = 5$. 70

2.B.1 Results of global sensitivity analysis applied to untreated tumour growth model in the absence of drug: a) and b) First and total-order sensitivity indices for model parameters at diagnosis for increasing sample sizes, N_S with $N_R = 10$ (number of resamples), respectively. Block lines represent the sample mean and shaded regions represent two standard deviations about the mean sensitivity index for a given parameter. Parameters are varied uniformly from a uniform distribution whose ranges are given in Table 2.2.1. 71

3.2.1 Schematic of our stochastic model of tumour growth and treatment response. Four cell populations are considered. Sensitive cells (green cells), are denoted by $S(t)$; cells that are sensitive to drug two and resistant to drug one (blue cells) are denoted by $R_1(t)$; cells that are sensitive to drug one and resistant to drug two (yellow cells) are denoted by $R_2(t)$; fully drug-resistant cells (red cells) are denoted by $R_{12}(t)$. The parameter μ denotes the intrinsic probability of mutation per cell division from (S to R_1), (S to R_2), (R_1 to R_{12}) and from (R_2 to R_{12}). S cells divide to produce S cells, ($S \to S + S$), R_1 cells ($S \to S + R_1$) and R_2 cells ($S \to S + R_2$) at rates, $b_S(1 - 2\mu)$, $b_S\mu$ and $b_S\mu$, respectively. S cells die at rate, d_S and the net growth rate of S cells is $r_S = b_S - d_S$. Similarly, R_1 cells divide to produce R_1 cells ($R_1 \to R_1 + R_1$) and R_{12} cells ($R_1 \to R_1 + R_{12}$) at rates, $b_{R_1}(1-\mu)$ and $b_{R_1}\mu$, respectively. R_1 cells die at rate, d_{R_1} and have a net growth rate of $r_{R_1} = b_{R_1} - d_{R_1}$. Similarly for R_2 cells. R_{12} cells divide to produce R_{12} cells at rate, $b_{R_{12}}$ and die at rate, $d_{R_{12}}$ and have a net growth rate, $r_{R_{12}} = b_{R_{12}} - d_{R_{12}}$. S cells are affected by drugs one and two at rate, $d_S(\theta_1 + \theta_2)$. R_1 cells are affected by drug two at rate, $d_{R_1}\theta_2$ and R_2 cells are affected by drug two at rate, $d_{R_2}\theta_1$. Propensity functions associated to cellular reactions presented in this schematic are summarised in Table 3.2.1 along with death rates associated to cellular competition. 76

3.4.1 Typical stochastic simulations (10 realisations) and the deterministic model solution describing untreated tumour cell dynamics for a range of values of μ and δ. Thick lines with black borders represent the solution of the ODE. Thin lines represent stochastic realisations. The y-scale is plotted using the symmetric logarithmic scale which is logarithmic for cell populations with numbers greater than 1 and linear for cell populations less than 1. Hence, populations with no cells are plotted. Key: blue represents S cells, orange; R_1 cells, green; R_2 cells and pink; R_{12} cells. Parameters: all model parameters are fixed at the default values stated in Table 3.2.2, with the exception of μ and δ; their values are indicated on each subplot. 86

3.4.2 The standard deviation about the median ($\times 100$) of 7000 stochastic simulations plotted as a percentage for values of μ and δ. Each row represents a different value of μ ($\mu = 10^{-8}$ in top row). For each subplot, values of $\delta = 0.5$, $\delta = 1$ and $\delta = 1.5$ are presented as dotted, dashed and block lines, respectively. Parameters: for non-varied model parameters, default values are given in Table 3.2.2 are employed. Ranges of values of μ and δ are also given in Table 3.2.2. For $\mu = 10^{-8}$ and cell line R_{12}, no values are plotted as the median remained zero for all time. 89

3.4.3 Histograms of values of $p_{R_{12}}$ at diagnosis for values of μ in $[10^{-8}, 10^{-6}, 10^{-4}]$ and $\delta = 1.5$. 91

3.4.4 (a) Probability of R_{12} cells present at diagnosis across simulations in which diagnosis occurs for a range of μ and δ values. (b) Predicted $p_{R_{12}}$ values generated using the ODE model. (e)-(f) Statistics of the proportions of pre-existing fully drug-resistant cells ($p_{R_{12}}$) for a set of stochastic realisations; (c) median and (d) standard deviation/median, (e) standard deviation, (f) Quartile coefficient of variation, $QVC = (Q3 - Q1)/(Q3 + Q1)$ where $Q1$ and $Q3$ denote the upper and lower quartiles of a given distribution of $p_{R_{12}}$ values. Blank squares correspond to probabilities equal to zero and means and median values equal to zero. In (a), white squares indicate cases where the probability is equal to zero. In (b), white squares indicate cases where the value of the ODE is less than 10^{-8} (i.e. no R_{12} cells present). In (c)-(d), white squares indicate cases where the median is equal to zero. 92

3.5.1 A selection of 10 stochastic simulations and ODE solutions of treated tumour growth dynamics for 3 values of the mutation probability, μ. Treatment with protocol A commences at diagnosis and is followed by protocol B when progression occurs on protocol A. Key: Blue represents S cells, orange; R_1 cells, green; R_2 cells and pink; R_{12} cells. For a given cell line, block thick lines represent the solution of the deterministic model (see chapter 2). Thin lines represent a stochastic simulations. Parameters: for non-varied model parameters, default values are given in Table 3.2.2 are employed. Ranges of values of μ are also given in Table 3.2.2 and values of μ are indicated in subplot titles. 94

3.5.2 Box plots of progression times (from diagnosis) and $p_{R_{12}}$ values at progression on first and second-line treatment. Box plots: orange lines represent the median. The box indicates the interquartile range (IQR) ranging from the lower quartile (25th percentile) to the upper (75th percentile) quartile. The bars indicate 1.5× the upper and lower quartiles and black circles indicate values greater or less 1.5× the upper and lower quartiles. Green crosses indicate the mean of the distribution of values recorded from 2623 stochastic simulations. We include progression times and $p_{R_{12}}$ values for untreated tumours generated by the ODE and present these predicted values using cyan squares. Parameters: for non-varied parameters, default model parameters presented in Table 3.2.2 are employed. 96

3.5.3 Progression times on first-line treatment (from the time of diagnosis) versus the proportion of R_{12} values at diagnosis from 2623 stochastic simulations. Blue crosses indicate results from the stochastic framework. Magenta crosses indicate the solution from the ODE. Parameters: for non-varied parameters, default model parameters presented in Table 3.2.2 are employed. 97

3.5.4 The duration of time from the start of second-line treatment to progression on, or after second-line treatment versus the proportion of R_{12} values at the start of second-line treatment, generated from 2623 stochastic simulations. Blue crosses indicate results from the stochastic framework. Magenta crosses indicate the solution from the ODE. Parameters: for non-varied parameters, default model parameters presented in Table 3.2.2 are employed. 97

3.A.1 Kolmogorov-Smirnov (K-S) statistics over time for tumour cell populations. 7000 realisations of the SSA and τ-leap method are simulated where a range of values of ϵ are considered in the τ-leap method. Key: coloured lines; different values of ϵ, black horizontal dashed line; K-S statistic value corresponding to the null hypothesis. 103

4.1.1 Schematic of the workflow in this chapter. 106

4.2.1 Plots of in vitro growth curves of the 4 cell lines: PC9, T790M, SCD4 and SCB5 (see Table 4.2.1). Marker shapes: circles, crosses and squares are representative of experiment repeats, $N = 1$, 2 and 3, respectively. The legend presents the marker shape corresponding to each experimental replicate. 110

4.2.2 In vivo control experimental mouse xenograft tumour volume data (units of mm^3). a) Raw in vivo tumour volume data. Marker shapes: diamonds, stars and triangles represent PC9, T790M and 1:1 SCD4:SCB5 tumours, respectively. Five experimental repeats were carried out for each cell line. b) Mean in vivo tumour volume data for each cell line (maker shape) and standard deviation (bars) across five experimental repeats. 112

4.3.1 Illustrative simulations of the candidate ODE models defined by equations (4.3.1)-(4.3.5) (see Table 4.3.1). Key: Colour black; solution of Richards growth equation with varying line styles corresponding to a range of values of β, dark blue; the exponential growth model, grey; logistic growth model, red; Gompertz model, cyan; generalised-logistic model. Common parameters, $r = 0.045$, $K = 50000$. Parameter values specific to the generalised logistic model: $\alpha = 0.99$, $\beta = 2$, $\gamma = 1.1$. Parameter values specific to Richards growth equation: β varies for $\beta \in [0.5, 3, 5, 15]$. We fix $y_0 = 1500$ cells for all model solutions. 116

4.5.1 Best-fit model solutions of candidate models (see Table 4.3.1 for models) fit to the in vitro data for the four cell lines (PC9, T790M, SCD4 and SCB5). Pink dots represent the data. Block black lines represent the best fit solution of the generalised-logistic model, and similarly for the following models: black dashed line; Gompertz, black dashed-dotted line; Richards growth equation, blue block line; exponential, black dotted line; logistic. A summary of the estimated parameters for all models are given in Table 4.5.1 for each cell line. 132

4.5.2 One-dimensional parameter profile likelihoods for the generalised-logistic model and the Richards growth equation (see Table 4.3.1) when fitted to each in vitro cell line data set. The x-axis describes the $log_{10}(\theta_i)$ profiled parameter values. Colours represent different model parameters (see figure legend). Coloured diamonds represent the $log_{10}(\hat{\boldsymbol{\theta}})$ value of the optimal parameter value. Black dashed lines represent the 95% confidence threshold and is plotted at the value $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$ where $icdf$ denotes the inverse cumulative density function. Coloured lines represent the interpolated profile likelihoods for each model parameter. 136

4.5.3 Synthetic data sets (orange dots) with noise levels, 1%, 5% and 20% and solutions of the Richards growth equation model fit to each data set (black lines). 138

4.5.4 Violin plots of estimated parameter values and relative standard errors (standard error/estimate) of model parameters, r, β and y_0 estimated by fitting the Richards growth equation to each of the the 10 synthetic data sets at different noise levels. The synthetic data is generated from the Richards growth equation with true model parameters: $r = 0.04$ (hour^{-1}), $\beta = 3$ (dimensionless) and $y_0 = 1500$ (cells). Horizontal block lines in each violin plot represent the mean estimated values and relative errors across the 10 fits for a given noise level. 139

4.5.5 One-dimensional profile likelihoods of r, y_0 and β when fitting the Richards growth equation to synthetic data for 0%, 5% and 20% noise. Colour block lines indicate profiles of parameters for different noise levels. Coloured dashed lines indicate the 95% confidence level plotted at $\chi^2(\hat{\boldsymbol{\theta}})+icdf(\chi^2_1, \alpha)$ for different noise levels. 140

4.5.6 Two-dimensional profile likelihoods for parameters r and β ($\chi^2_{PL}(r, \beta)$) plotted as contour plots. Dark green and light green regions represent low and high $\chi^2(\theta)$ values, respectively. Black dashed lines represent the 95% confidence level calculated as $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_2, \alpha)$, where $icdf(\chi^2_2, \alpha)$ represents the $95th$ percentile for the χ^2 distribution with 2 degrees of freedom at the 95% confidence level. Crosses represent the coordinates of the optimal parameter values of r and β. For each cell line, the likelihood-based confidence region is infinitely extended for β demonstrating that an increase in β does not markedly alter the goodness-of-fit of the Richards equation to the in vitro data. 141

4.5.7 One-dimensional profile likelihoods of the reduced Richards growth equation model with fixed values of $\beta = \beta_F$ ($\beta_F = 3.315$). Black dashed lines represent the 95% confidence level calculated (see Figure 4.5.2 for details). The endpoints of the likelihood-based confidence intervals lie at the intersection between the likelihood and the 95% confidence level threshold and are indicated by vertical grey dashed lines (see Table 4.5.7). For all cell lines, r and y_0 is practically identifiable and we say that the model is identifiable. 142

4.5.8 Richards growth equation model, with fixed β, fitted to the in vitro data. (a)-(d) Solutions of the Richards growth equation model fitted to the in vitro data. Black dashed line; best-fit model solutions found using non-linear least squares, black block line; median of posterior predictive distribution, blue shaded region; 95% credible interval of the model (between the 2.5th and 97.5th percentile of the posterior predictive distribution) found using MCMC. Red dots represent the data. (e)-(h) Corresponding residual plots to subplots (a)-(d). Cyan crosses represent residuals for NLLS. Black dots represent residuals for the median of the posterior predictive distribution. Black horizontal lines indicate residual values equal to zero. 145

4.5.9 One-dimensional parameter profile likelihoods of the Richards growth equation model, with estimated β and carrying capacity, K. Horizontal black dashed lines represent the 95% confidence level, plotted at $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$. Red dots represent profile likelihood values. Black block lines represent the interpolated profile likelihood. Cyan crosses represent optimal estimated parameter values. 147

4.5.10 Best-fit model solutions of the Richards growth equation fitted to the in vivo data for each cell line, for fixed $\beta = 1$. Dots represent raw in vivo tumour volume data. Block lines represent the best fit solution for $\beta = 1$. Shaded regions represent 95% confidence bands, calculated by sampling from a multivariate normal distribution with model parameter estimates (see Table 4.5.9) as the means of the distribution. We employed each sample of parameter values (150000 in total) in the Richards growth equation. At each time point, we calculated the median and the 2.5th and 97.5th quantiles of the distribution of model solutions, yielding 95% confidence intervals of the model fit to the data. 147

4.B.1 Autocorrelation function plots of combined Markov chain samples (10 independent Markov chains in total) for the Richards growth equation with fixed β. A thinning of 1/100 and a burn in of 5000 samples were used to increase the independence of the samples. Rapid decay of the autocorrelation function for all parameters and all datasets is observed, demonstrating the independence of the MCMC samples and convergence of the MCMC chains. Lag (x-axis) describes the distance between two samples in the MCMC sequence. The autocorrelation drops to a value close to zero within a lag of approximately 20 for all parameter and datasets, a value that is much smaller than the number of MCMC samples (up to 150,000 for SCD4 and SCB5 data sets), demonstrating the samples of the MCMC quickly become independent and are not correlated. These results suggest that the model is identifiable when β is fixed. 156

4.B.2 Trace plots of 3 randomly selected, independent Markov chains for each data set and for the Richards growth equation with fixed β. The Markov chains are of length 100,000 for PC9 and T790M data sets and of length 150,000 for SCD4 and SCB5 datasets. Each chain is initiated with initial parameter values which are sampled from a neighbourhood of 0.1ϵ about the optimal estimated parameter values, where $\epsilon \sim U(0, 1)$. The Markov chains appear to converge for all model parameters and datasets. Key: colours represent different chains. These results suggest that the model is identifiable when β is fixed. 157

4.B.3 Posterior distributions (histograms on the diagonal) and joint posterior distributions (contour plots) of the Richards growth equation model parameters when β is fixed. For each in vitro dataset, 10 MCMC chains were run independently, with lengths 100,000 for the PC9 and T790M data sets and of lengths 150,000 for the SCD4 and SCB5 data sets. The independence of samples and the convergence of the MCMC chains were checked by the autocorrelation function and the trace plots. A Gelman-Rubin statistic of $\hat{R} \leq 1.01$ was calculated for all model parameters and all cell lines, confirming convergence of the MCMC chains. Black dashed lines indicate the optimal r and y_0 parameter values, estimated using non-linear least squares method and for σ, correspond to the initial value of the proposed noise level $\sigma = 1000$ in the prior parameter distribution, employed in our MCMC. Key: orange curves on diagonal plots; the kernel-density estimation for the histograms. For the contour plots: dark and light regions correspond to high and low densities, respectively. 159

4.B.4 Visual posterior predictive check (PPC). 30 samples of the posterior predictive distribution are plotted, where random Gaussian noise is added at each time point with mean zero and variance equal to the σ value for a given sample of the Markov chain (see section 4.4 for details on how the PPC is performed). These results demonstrate the model is appropriate for the data for all datasets. Key: red dots; data, grey lines; model solutions sampled from the posterior predictive distribution, with noise. 160

5.1.1 Schematic of the workflow in this chapter. 162

5.2.1 In vitro treatment data. (a)-(h) In vitro dose-response curves of the 4 cell lines; PC9, T790M, SCD4 and SCB5. Key: colours represent different drug concentrations in micromolar (μM). Shapes represent experiment repeats: circles; $N = 1$, crosses; $N = 2$, squares; $N = 3$. Drug concentration, c is fixed throughout the experiment. 165

5.2.2 In vitro dose-response curves of PC9 cells in 0.123μM gefitinib. Key: shapes represent experiment repeats: circles; $N = 1$, crosses; $N = 2$, squares; $N = 3$. 165

5.3.1 Hill equation dynamics for a range of values of the Hill coefficient, m (equation 5.3.3). Parameters: $\lambda = 0.1$, $IC50 = 5$. 168

5.5.1 Best-fit solutions of the reduced Richards growth equation, with fixed $\beta = \beta_F$ (see equation (4.3.2) in Table 4.3.1 in chapter 4), fitted to drug sensitive PC9 cells in gefitinib. For each drug concentration, c, parameters r and y_0 are estimated. Key: coloured dots; data, black lines; model solution describing $y(t)$ over time. Values of c are indicated in the subplot titles. 174

5.5.2 Best-fit solutions when the reduced Richards growth equation, with fixed $\beta = \beta_F$ (see equation (4.3.2) in Table 4.3.1 in chapter 4), is fitted to drug sensitive SCD4 cells in gefitinib. Figure details the same as in Figure 5.5.1.174

5.5.3 Estimated values of r (left hand y-axis) and corresponding minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ (right hand y-axis) plotted with respect to c for PC9 and SCD4 gefitinib data sets when fitting the reduced Richards growth equation, with fixed β, to individual dose-response curves. Key: black dots; estimated values of r and bars represent the standard error of each estimate, cyan diamonds; $\chi^2(\hat{\boldsymbol{\theta}})$ values. 174

5.5.4 Model fits for the PC9 gefitinib data set. The model is fitted to pooled dose-response curves for all drug concentrations. Key: blue dots; data, blue block lines; best-fit solution of model describing the total population over time, $N(t)$, orange dashed lines; best-fit solution describing live cell numbers, $y(t)$; green dotted lines; best-fit solution describing dead cell population, $z(t)$. The drug concentrations, c are indicated in each subplot title. 181

5.5.5 One-dimensional parameter profile likelihoods for treatment model parameters for each drug-sensitive data set. Key: magenta crosses; optimal parameter values (see Table 5.5.1), black dashed horizontal lines; 95% confidence level calculated as $icdf(\chi^2_1, \alpha)$ and plotted at the value $\chi^2(\hat{\boldsymbol{\theta}})+icdf(\chi^2_1, \alpha)$, grey vertical dashed lines; lower and upper likelihood-based confidence interval bounds, red dots; calculated profile likelihood values, black block line; cubic interpolation of the profile likelihood values. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 184

5.5.6 Model predictions of total ($N(t)$), live ($y(t)$) and dead ($z(t)$) cell populations describing PC9 cells in $c = 0.123\mu$M gefitinib for two profiled values of k, $k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1} which are fixed. For each value of k, the model is fitted to the pooled dose-response curves and the other model parameters are estimated (estimates of these parameters are given in Table 5.5.5). For fixed $k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1}, similar values of $\chi^2(\hat{\boldsymbol{\theta}})$ were calculated ($\chi^2(\hat{\boldsymbol{\theta}}) = 11.93$ for $k = 10^{-5}$ hour^{-1} and $\chi^2(\hat{\boldsymbol{\theta}}) = 11.46$ for $k = 0.031$ hour^{-1}). Key: blue dots; data, blue lines; best-fit solution of the total tumour population, orange; live cells and green; dead cells. Parameters: see Table 5.5.5. . . 187

5.5.7 Estimated values of practically identifiable model parameters along profiles of practically non-identifiable model parameters, calculated using profile likelihoods (see Figure 5.5.5 in section 5.5.4.2). The x-axis represents values of the profiled practically non-identifiable model parameter. For each profiled parameter value, we record the estimated values of the practically identifiable model parameters from the profile likelihoods which we plot on the y-axis. (a) Estimated values of practically identifiable parameters, r, $IC50$ and λ along the m profile for PC9 cells in osimertinib, (b) Values of r, $IC50$ and λ along the k profile for PC9 cells in osimertinib, (c) values of r and λ along the $IC50$ profile for T790M cells in osimertinib. Key: block lines; estimated values of $IC50$, dotted lines; λ; dashed-dotted lines; r. 190

5.5.8 Best-fit solutions when the Hill equation (equation (5.3.3)) is fitted to estimated values of $\tilde{h}$ ($\tilde{h}$ is estimated by applying step 1 which we describe at the beginning of this section) for the PC9 gefitinib data set. Key: coloured block lines; m fixed with values in $\{0.5, 1, 2, 3, 4\}$, black dashed-dotted lines; m is estimated. We also provide, in the Figure legend, for a given fit, the estimated value of λ and $IC50$ along with the estimated or fixed value of m and the goodness-of-fit value, $\chi^2(\hat{\boldsymbol{\theta}})$. 192

5.5.9 Right hand y-axis: estimated values of λ and $IC50$ for varying k, obtained using steps 1 and 2 from subsection 5.5.7 for fixed values of $k \in \{0.0001, 0.001, 0.01, 0.1\}$. Key: black dots (joined with black lines); estimated values of λ with bars representing standard errors of the estimates, cyan (joined with cyan lines); estimated values of $IC50$ with standard errors as bars. Left hand y-axis: the mean of the $\chi^2(\hat{\boldsymbol{\theta}})$ values, where each $\chi^2(\hat{\boldsymbol{\theta}})$ value is calculated by performing step 1 from section 5.5.7, for a given dose-response curve with a fixed value of k. Key: pink dots; mean $\chi^2(\hat{\boldsymbol{\theta}})$ value. 193

5.5.10 One-dimensional profile likelihoods for parameters r, $IC50$ and λ when m and k are fixed with $k = 10^{-5}$ hour^{-1} and values of m fixed at values given in Table 5.5.8. Key: horizontal black dashed lines; 95% confidence level plotted at $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$, vertical grey dashed lines; likelihood-based confidence interval bounds, cyan crosses; optimal parameter values, red dots; profile likelihood values, block black lines; cubic interpolation of the profile likelihood values. The units of parameters are, r (hour^{-1}), $IC50$ (μM) and λ (hour^{-1}). 194

5.5.11 Best-fit model solutions when the model is fitted to pooled dose-response curves of PC9 cells in gefitinib. Key: blue lines; total tumour cell population, orange lines; live cells, green lines; dead cells. Shaded regions indicate the 95% confidence bound of the model solution for a given cell line, calculated using the Monte Carlo Simulations Method with 5000 simulations (see Methods section of this chapter). Parameters: see Table 5.5.9. 196

5.5.12 Synthetic data describing live ($y(t)$), dead ($z(t)$) and total ($N(t)$) cell numbers over time for a range of drug concentrations. Synthetic data are generated from the full treatment model using the best-fit parameter estimates of PC9 cells in gefitinib (see Table 5.5.1). Key: green; total cell numbers, red; live, black; dead. 199

5.5.13 Profile likelihoods for the treatment model parameters when fitted to the pooled synthetic data sets consisting of live and dead cell numbers. Key: cyan crosses; optimal parameter estimates, black horizontal dashed lines; 95% confidence threshold calculated as the $min(\chi^2_{PL}(\theta_i)) + icdf(\chi^2, \alpha)$, black block lines; cubic interpolation of the profile likelihood. 199

5.A.1 Estimated values of r (left hand y-axis) and corresponding minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ (right hand y-axis) plotted, with respect to c for the dose-response curves in the PC9 osimertinib data set (a), T790M data sets (b)-(c), the SCB5 data sets (d)-(e) and the SCD4 osimertinib data set (f). Values of r and $\chi^2(\hat{\boldsymbol{\theta}})$ are calculated by fitting the reduced Richards growth equation, with fixed β, to individual dose-response curves. Key: black dots; estimated values of r and bars represent the standard error of each estimate, cyan diamonds; $\chi^2(\hat{\boldsymbol{\theta}})$ values. 203

5.A.2 Model fits for the PC9 osimertinib data set. The model is fit to pooled dose-response curves for all drug concentrations. Key: blue dots; data, blue block lines; best-fit solution of model describing the total population over time, $N(t)$, orange dashed lines; best-fit solution describing live cell numbers, $y(t)$; green dotted lines; best-fit solution describing dead cell population, $z(t)$. The drug concentration, c are indicated in each subplot title. 204

5.A.3 Model fits for the T790M osimertinib data set. The model is fit to pooled dose-response curves for all drug concentrations. Key: see Figure 5.A.2 for plotting details. 205

5.A.4 Bar plots of first and total-order sensitivity indices of model parameters recorded at fixed times ($t = 10$, $t = 60$ and $t = 120$ hours) for fixed drug concentrations $c \in \{0.01, 0.1, 1, 3\}$. Different colours represent parameters (see Figure legend). eFAST control parameters: number of samples per resample; 4000, number of resamples; 8. Convergence testing was carried out to determine these control parameter values although not presented (see chapter 2 for details of eFAST). 206

5.A.5 Best-fit solutions when the Hill equation (equation (5.3.3)) is fitted to estimated values of $\tilde{h}$ ($\tilde{h}$ is estimated by applying step 1 in subsection 5.5.7) for the PC9 and T790M osimertinib data sets. Key: coloured block lines; m fixed with values in $\{0.5, 1, 2, 3, 4\}$, black dashed-dotted lines; m is estimated. We also provide, in the Figure legend, for a given fit, the estimated value of λ and $IC50$ along with the estimated or fixed value of m and the goodness-of-fit value, $\chi^2(\hat{\boldsymbol{\theta}})$. 207

5.A.6 Best-fit model solutions when the model is fitted to pooled dose-response curves of PC9 cells in osimertinib. Key: blue lines; total tumour cell population, orange lines; live cells, green lines; dead cells. Shaded regions indicate the 95% confidence bound of the model solution for a given cell line, calculated using the Monte Carlo Simulations Method with 5000 simulations (see Methods section of this chapter). Parameters: see Table 5.5.9. 208

5.A.7 Best-fit model solutions when the model is fit to pooled dose-response curves of T790M cells in osimertinib. Key: plotting details the same as in Figure 5.A.6. 208

List of Tables

2.2.1 Default model parameter values, descriptions, units and their respective uniform distribution ranges used in eFAST analysis. We include references for their source. 37

2.5.1 Summary of treatment protocols applied at diagnosis when the total tumour population reaches 10^8 cells. Protocol descriptions, amounts of drug applied in a given protocol, response, units and references are given. 54

2.A.1 Table of parameters and ranges for simple model parameters defined in equations (2.A.1)-(2.A.2). We use the same parameter notation as that presented in the main text. We also use the same parameter ranges. References for the sources of the parameter ranges can be found in Table 2.2.1. 67

3.2.1 Propensity functions for the cellular reactions associated with the four cell types. These reactions include birth without mutation, birth with mutation, and death due to natural cell death, drug exposure and interspecific and intraspecific competition. Λ_{n+1} denotes the state of the stochastic process when a particular event occurs, given $\Lambda_n = [S, R_1, R_2, R_{12}]$. 77

3.2.2 Default parameter values, descriptions and their respective units (for references motivating these parameter values, see chapter 2). All model parameters are fixed except for the mutation probability, μ and, δ, the relative growth rate of the resistant cells to the growth rate of the sensitive cells. Dim. denotes dimensionless. 79

3.A.1 Parameter values employed in the model for convergence testing of the τ-leap algorithm. Descriptions of model parameters are given in Table 3.2.2. 102

4.2.1 Summary of the properties of the four NSCLC cell lines used to generate the in vitro datasets: cell lines and their respective resistance conferring mutations to two TKIs, gefitinib and osimertinib. 109

4.3.1 Table presenting candidate models of in vitro tumour growth: generalised-logistic, Richards growth equation, logistic, Gompertz, and exponential growth. $y(t)$ denotes the number of cells at time, t. For a summary of model parameters, see Table 4.3.2. 114

4.3.2 Table of estimated and fixed parameters and their respective units, corresponding to in vitro candidate models given in Table 4.3.1 which are fitted to in vitro data. The column labelled Model refers to the model reference in which a given parameter is present. Carrying capacity, K is fixed at $K = 50000$ cells in candidate models 1-4 and all other parameters are estimated by fitting each model to the in vitro datasets. Dim. is an abbreviation for dimensionless. 115

4.4.1 Parameter bounds for candidate models (with reference numbers 1-5 as in Table 4.3.1) when fitting to the in vitro data. Candidate models are given in Table 4.3.1. Dim. denotes dimensionless. 117

4.4.2 Parameter bounds for Richards growth equation (equation (4.3.2) in Table 4.3.1) when fitted to in vivo tumour volume data and parameters r, β, y_0 and K are estimated, where y_0 denotes the initial tumour volume. 118

4.5.1 Table of estimated values of model parameters ($\pm$ standard errors) and goodness-of-fit statistics, sum of squared errors ($\chi^2(\hat{\boldsymbol{\theta}})$), AIC and BIC information criterion statistics for each candidate model (see Table 4.3.1) when fitted to the in vitro data for each cell line. Dim. denotes dimensionless. 133

4.5.2 Table of correlations between estimated model parameters for the generalised-logistic model (see equation (4.3.1) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}), α (dim.), β (dim.), γ (dim.) and y_0 (cells) where dim. stands for dimensionless. 134

4.5.3 Table of correlations between estimated model parameters for the Richards growth equation model (see equation (4.3.2) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}), β (dim.), and y_0 (cells) where dim. stands for dimensionless. 134

4.5.4 Table of correlations between estimated model parameters for the logistic growth model (see equation (4.3.3) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells). 135

4.5.5 Table of correlations between estimated model parameters for the Gompertz model (see equation (4.3.4) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells). 135

4.5.6 Table of correlations between estimated model parameters for the exponential model (see equation (4.3.5) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells). 135

4.5.7 Estimated values of parameters, r and y_0 ($\pm$ standard error) when the reduced Richards growth equation (fixed $\beta = 3.315$) is fitted to the in vitro data for each cell line. The likelihood-based confidence intervals are calculated using the profile likelihood method (CI stands for confidence interval). The goodness-of-fit values, $\chi^2(\hat{\boldsymbol{\theta}})$ are also given. Std. err stands for standard error. Likeli. stands for likelihood. 143

4.5.8 Table of correlations between estimated model parameters for the reduced Richards growth model with fixed $\beta = 3.315$ (see equation (4.3.2) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells). 143

4.5.9 Table of estimated in vivo model parameters ($\pm$ standard error) for each tumour type. Estimated parameter values include the carrying capacity, K, growth rates r and the initial tumour volume, y_0. Std. err stands for the standard error. 148

4.5.10 Table of correlations between estimated parameters of the Richards growth equation model fitted to the in vivo tumour growth data for each tumour type. 148

5.4.1 Estimated values ($\pm$ standard error) of the growth rate, r, for each cell line in the absence of drug. Estimates were found by fitting the reduced Richards growth equation, with fixed $\beta = 3.315$, to the control in vitro data sets in chapter 4. 170

5.4.2 Summary of estimated and fixed parameters for the in vitro treatment model defined by equations (5.3.1)-(5.3.3). Values of fixed parameters and parameter bounds considered for estimated model parameters are given along with their units. Except for the carrying capacity, K which is fixed at 50,000 cells and the competition strength parameter β_F which is also fixed ($\beta_F = 3.315$), all model parameters are estimated by fitting to the in vitro data sets. Bounds on the parameters are used for parameter estimation and the calculation of profile likelihoods (see subsection 5.5.4.2). 171

5.5.1 Estimated parameter values ($\pm$ standard error) and $\chi^2(\hat{\boldsymbol{\theta}})$ values of the treatment model fitted to pooled dose-response curves of the drug-sensitive data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. . 180

5.5.2 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 181

5.5.3 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 osimertinib data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 181

5.5.4 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the T790M osimertinib data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 182

5.5.5 Estimated values ($\pm$ standard error) of model parameter when profiled values of k ($k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1}) are fixed. For each fixed value of k, the model is fitted to the pooled dose-response curves and the other model parameters are estimated. These parameters are employed in the model and to generate model solutions in Figure 5.5.6. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 187

5.5.6 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set for fixed $k = 10^{-5}$ hour^{-1}. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. . 188

5.5.7 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set for fixed $k = 0.31$ hour^{-1}. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. . 188

5.5.8 Estimated parameter values ($\pm$ standard error) of Hill equation parameters, λ, $IC50$ and m for drug-sensitive data sets. Estimates were calculated by performing steps 1 and 2 in subsection 5.5.7. The units of parameters are, $IC50$ (μM), m (dim.) and λ (hour^{-1}). Dim. stands for dimensionless. 193

5.5.9 Estimated values of model parameter ($\pm$ standard error) and $\chi^2(\hat{\boldsymbol{\theta}})$ values found by fitting the reduced model, with fixed m and k, to the pooled dose-response curves for each drug-sensitive data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 195

5.5.10 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set for fixed $k = 10^{-5}$ hour^{-1} and $m = 2$. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 196

5.5.11 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 osimertinib data set for fixed $k = 10^{-5}$ hour^{-1} and $m = 2$. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 196

5.5.12 Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the T790M osimertinib data set for fixed $k = 10^{-5}$ hour^{-1} and $m = 1$. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless. 197

5.A.1 Estimated values ($\pm$ standard error) of initial cell numbers, y_0, for the in vitro PC9 data sets and the T790M osimertinib data set. Values of y_0 are estimated by fitting the treatment model to single dose-response curves with a fixed drug concentration. Values of y_0 are estimated and fixed at their estimated values prior to fitting the model to pooled dose-response curves. 204

5.A.2 Bounds for treatment model parameters defined in equations (5.3.1)-(5.3.3) when performing global sensitivity analysis of the treatment model using eFAST. 206

Chapter 1

Introduction

Cancer incidence and mortality are growing rapidly, with an estimated 10 million cancer deaths worldwide in 2020, up from 9.6 million in 2018 [12, 132]. In 2020, lung cancer was the leading cause of cancer mortality, with 1.8 million deaths (18% of cancer deaths), and the second most commonly diagnosed cancer (11.4% of new cases worldwide). The fundamental factor limiting a cure in lung cancer is its ability to develop mechanisms that enable it to evade pharmaceutical treatments, i.e. drug resistance. These mechanisms include, mutations that alter drug binding targets, the removal of drugs from cells due to overexpression of drug efflux pumps and changes to the cancer cell's ability to metabolise drugs, resulting in decreased drug activation [63, 89].

In this thesis, we focus on non-small cell lung cancer (NSCLC), a histological subtype of lung cancer affecting 80-85% of lung cancer patients [37]. The emergence of drug resistance in NSCLC is the leading cause of failure in tyrosine kinase inhibitor (TKI) treatment. TKIs are targeted therapies that bind to receptors on the surface of cancer cells and interfere with mechanisms regulating cell growth and death. Initial response to TKIs in the clinic is observed but tumour progression is inevitable due to the acquisition of drug resistance mechanisms. Common mechanisms of resistance to TKIs are mutations that alter the drug-binding target. Mathematical models can elucidate and quantify our understanding of tumour growth and the biological processes that drive tumours to become drug-resistant. We develop mathematical models that describe the growth of monotypic populations (i.e. one type of tumour cell) and mixed populations that differ in their sensitivity to TKIs. We use these models to identify factors that

influence the efficacy of treatment with TKIs. Through coupling the models with pre-clinical data sets, which allows us to fit them and estimate model parameters, we are able to infer the biological mechanisms underlying tumour cell growth and response to TKI treatment.

In the remainder of this chapter, we provide context for the research. We start by describing NSCLC and summarising treatments that are currently applied in the clinic. We continue by introducing the various oncogenic mutations that are associated with NSCLC and explain their relevance in TKI treatment. We then review drug resistance mechanisms in NSCLC and explain how TKI treatment has evolved to tackle resistance and, discuss combination treatments involving TKIs that are currently being used in the clinic. We then review the relevant mathematical literature, focussing on methodologies that have been applied in this thesis. We conclude this introduction by outlining the structure of the remainder of the thesis.

1.1 Biological review

1.1.1 Overview of non-small cell lung cancer (NSCLC)

NSCLC is an umbrella term that includes cancers such as lung adenocarcinoma (LUAD) ($\sim$ 40% of cases), lung squamous cell carcinoma (LUSC) (25-30% of cases) and large cell undifferentiated carcinoma (10-15% of cases) [130]. Lung cancer has an estimated 5-year survival rate of 15% in England, underlining the importance of studies that investigate the mechanisms responsible for treatment failure [16].

More than two thirds of NSCLC cases are diagnosed at stages III-IV [130]. At these stages, NSCLC typically presents as a large primary tumour and/or one that has already metastasised (i.e. spread to other parts of the body). Applied treatments for NSCLC depend on the stage of the disease. These treatments include surgical resection, chemotherapy, radiotherapy, targeted therapy and immunotherapy [154]. The standard first-line treatment for stage IV NSCLC is cytotoxic combination chemotherapy [149, 154]. Over the past 20 years, the development of targeted therapies has been increasingly popular for treating NSCLC with specific genetic profiles. Targeted therapies are agents that target genes and proteins in the cancer cell. These targets include specific oncogenic

mutations that drive the disease. Therefore, we focus on TKI therapy and the NSCLC-specific mutations that they are designed to target. More details about TKI therapy will be given in section 1.1.4 of this introduction.

1.1.2 Mutations driving NSCLC

NSCLC is a highly heterogeneous disease, with a higher tumour mutation burden (TMB) in patients with metastatic NSCLC [57]. Genome sequencing in NSCLC has identified hundreds of mutations and the number of mutations is shown to increase with tumour grade (level of abnormality of tumour cells compared to healthy cells) [21, 146]. Mutations in the epidermal growth factor receptor (EGFR) gene, which is a member of the erbB family of receptor tyrosine kinases which span the cancer cell membrane, are present in $15-20\%$ of NSCLC patients [107]. Binding of epidermal growth factor (EGF) to the kinase domain of the EGFR protein induces the transduction of signals across the cell surface. These signals upregulate cell proliferation and anti-apoptotic pathways and are popular targets for TKIs.

The most common mutations in the EGFR gene are exon 19 deletions and exon 21 L858R (leucine to arginine) substitution; together, they account for more than 90% of EGFR mutations [9, 141, 152]. These mutations can lead to overexpression of EGFR on the cancer cell surface, which promotes uncontrolled cell division [9, 152]. EGFR mutations are a prominent feature of the mathematical models of NSCLC that we develop in chapters 2-5. In chapter 4 we use experimental data sets, provided by our industrial sponsor, AstraZeneca, to develop mathematical models describing the growth of EGFR-mutant NSCLC cell lines. Using parameter estimation and parameter inference techniques that are commonly used in the pharmaceutical industry, we infer the biological mechanisms driving the growth dynamics of these cell lines. In chapter 5, using the same techniques on in vitro data sets, describing the response of four different NSCLC cell lines to TKI treatment, we examine how these cells respond to treatment with EGFR-TKIs and, in so doing, seek to determine the underlying mechanisms of action of first and third-generation EGFR TKIs. We discuss parameter estimation and parameter inference techniques in more detail in section 1.2.6.

Although mutations in the EGFR gene are common (10-15% in LUAD in female, never-smokers), other mutations exist in NSCLC [154]. Mutations in the Kirsten rat sarcoma gene (KRAS) (30% in LUAD in western countries) disrupt signalling which normally controls cell proliferation, resulting in increased rates of cell division [148]. Mutations in the mesenchymal epithelial transition factor proto-oncogene (MET) (7% of LUAD) induce amplifications of growth receptors on the cell surface which, in turn, drives uncontrolled cell division. Rearrangements in the anaplastic lymphoma kinase (ALK) gene (found in approximately 5% of NSCLC patients) promote uncontrolled cell growth [72]. EGFR and ALK mutations are mutually exclusive [40]. EGFR and KRAS mutations are generally mutually exclusive however, when KRAS mutations co-exist with EGFR mutations, KRAS mutations may confer resistance to EGFR TKIs [9, 40, 57, 109, 118, 142].

Genetic alterations in tumour cells that occur in the early stages of NSCLC are natural targets for TKI treatment because, identification of these mutations has shaped NSCLC treatment today [69]. Prior to treatment, gene expression profiles of patients' tumours are measured to assess the most appropriate course of treatment. Tumours with EGFR mutations in exon 19 or 21 may benefit from TKIs as an alternative to first or second-line chemotherapy [77]. Similarly, patients with ALK rearrangements benefit from ALK inhibitors [61]. Although mutations in KRAS are most common in NSCLC, no drugs that effectively target KRAS mutations have been clinically approved [130]. While TKI treatment has shown promise, response to these drugs is short-lived due to the presence of pre-existing resistant cells at the start of treatment and/or resistance acquisition during TKI treatment [55, 64]. In the next section, we describe the characteristic features of NSCLC and review the mechanisms which contribute to NSCLC resistance to therapy.

1.1.3 Drug resistance mechanisms in NSCLC

Mechanisms of drug resistance in NSCLC include, overexpression of drug efflux pumps which remove drugs from the cancer cells, changes in the cell's ability to metabolise drugs which would lead to the drugs activation and the acquisition of altered drug-binding targets [89]. Broadly speaking, two groups of mechanisms drive EGFR TKI resistance.

The first group involves alterations to the drug-target via secondary point mutations or changes in expression levels of the drug-target themselves. These mutations alter the structure of the EGFR and, in doing so, may also change binding rates and/or increase the receptor's affinity for growth factors such as EGF [141]. In addition an increase in receptor expression levels results in more drug being needed to inhibit activated receptors on the cell surface [62]. The second group of mechanisms acts by amplifying alternative pathways, including the amplification of MET and KRAS and the rapidly accelerated fibrosarcoma oncogene homolog B (BRAF), which upregulate cell growth [101].

In this thesis, we focus on secondary point mutations that increase tumour growth rates and enable them to evade treatment. Secondary point mutations that give rise to drug-resistant cells are either pre-existing or acquired during TKI treatment [55, 71]. At a cellular level, intrinsic resistance refers to resistant cells that are present before treatment and that have arisen as a result of genetic and epigenetic alterations during earlier stages of tumourigenesis [14, 62]. We will refer to this type of resistance mechanism as 'pre-existing'. Selection for pre-existing resistant clones during treatment results in treatment failure. In chronic myeloid leukaemia (CML), pre-existing clones that are resistant to the TKI, imatinib, are correlated with poor treatment outcomes [66]. Conversely, at a cellular level, acquired resistance arises when cells which are initially responsive to treatment become resistant [14, 62]. In the clinic, primary (intrinsic/pre-existing) resistance refers to failure of initial treatment response and secondary (acquired) resistance describes cases for which an initial response is observed but, over time, resistance develops [14, 62]. There is a long-standing debate about whether pre-existing drug-resistant, or, *de novo* acquired cell populations, have a greater impact on the failure of cancer treatments [71]. Whether the *de novo* acquisition of resistance conferring alterations is caused by random mutations (Darwinian resistance) or drug-induced changes is also debated. As stated by Alvarez-Arenas et al., Darwinian evolution of cancer describes the case in which cancer and drug resistance develops through a series of stochastic epigenetic and genetic changes [4]. Through a Darwinian process, tumour cells are subject to selective pressures (e.g. the tumour microenvironment and anticancer treatments) which result in resistant cells emerging within the tumour population [45].

The pathways to resistance are diverse and studies that elucidate understanding of these pathways are needed for rational drug development and protocol design [55, 71]. Further information about molecular mechanisms of drug resistance can be found in the literature [14, 62]. In chapter 3, we employ a stochastic model of tumour growth and use it to demonstrate how the number of (pre-existing) resistant tumour cells at the start of treatment affect a tumour's response to treatment. We demonstrate how the number of resistant cells in the tumour at the time of failure of first-line treatment affects a tumour's response to second-line treatment. We present further details of the modelling techniques used in chapter 3 in section 1.2.6 of this introduction and we outline the work of this chapter in more detail in section 1.3 of this introduction.

Cancer was first described as an evolutionary process by Nowell in 1976 whereby tumours are heterogeneous populations of cells with varying genetic, epigenetic and phenotypic characteristics (e.g. drug resistance) [106]. Genetic instability is a hallmark of cancer affecting DNA sequence, chromosome number and structure and drives heterogeneity in the population of cancer cells which contributes to treatment failure [14, 54, 62]. Heterogeneity can take the following forms: intratumoural heterogeneity refers to the variation of tumour cells within a single patient and is believed to contribute to drug resistance [32, 42]. Intertumoural heterogeneity refers to the heterogeneity across patients with tumours of the same histological subtype and is believed to be driven by patient-specific factors including the patient's genetic profile [32]. In chapter 2, we use a deterministic model of tumour growth to investigate the role of factors such as mutation rates and cell division rates of cancer cells on tumour composition and response to treatment.

Interactions between tumour cells and their environment influence the emergence of resistance and tumour cell heterogeneity. In vitro and in vivo mouse model studies have been carried out to investigate the effects of competition for space, differences in environmental conditions on the growth of tumour cells and their response to treatment [5, 72, 92, 108]. In chapter 2, we exploit this novel perspective, that interactions between cells influence tumour cell heterogeneity, to guide model construction and in the examination of the modelling predictions as further detailed in section 1.3.

In what follows, we explain how treatments involving EGFR TKIs have evolved over time, with new TKIs being developed to replace earlier TKIs when patients develop resistance to them.

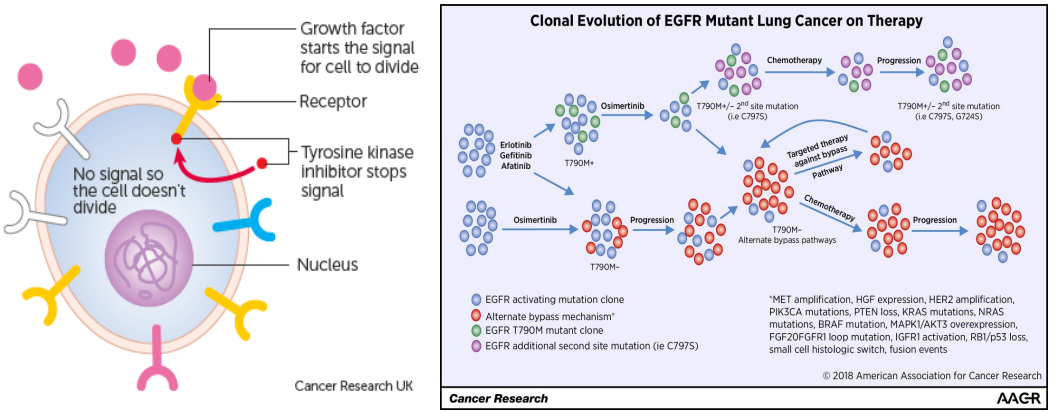


Figure 1.1.1: Left) Diagram of a tyrosine kinase inhibitor binding to a cancer cell. In the absence of drug, growth factors such as epidermal growth factor (EGF) bind to the receptor and signals are transduced across the cell surface, upregulating anti-apoptotic and cell proliferation pathways. Tyrosine kinase inhibitors bind to the receptor and block these signals (diagram credit: Cancer Research UK). Right) Diagram of the evolution of TKI treatment and resistance mechanisms that result in their failure (diagram credit: Murtuza et al. [101]).

1.1.4 Epidermal growth factor receptor TKIs

Development of tyrosine kinase inhibitors is ongoing due to their initial effect and subsequent failure due to drug resistance. Targetable oncogenic driver mutations were found to be present in 64% of patients who harbour tumours of the lung adenocarcinoma subtype and the use of genotype-directed therapies improved survival in these patients [57]. In Figure 1.1.1(right hand figure), we present a schematic of the development of EGFR TKI treatment and the evolution of resistance to these drugs. In what follows, we outline how such resistance emerges.

Gefitinib (tradename Iressa, AstraZeneca) is a first-generation TKI applied to EGFR mutant NSCLC. Gefitinib reversibly binds to the tyrosine kinase domain and inhibits the activation of the anti-apoptotic Ras signal transduction cascade. A landmark 2004 phase III clinical trial presented by Lynch et al. demonstrated that NSCLC patients with chemotherapy-refractory disease (i.e. these patients' tumours are no longer responsive to chemotherapy) and activating EGFR mutations, exon 19 deletions and exon

21 substitutions, had better rates of survival when treated with gefitinib than patients without EGFR mutations [86]. Gefitinib and erlotinib, another first-generation EGFR TKI are currently recommended as first-line treatment in NSCLC patients with EGFR mutations. However, response to these drugs is short-lived. Within a median time of 9-12 months, NSCLC patients acquire resistance to first-generation EGFR TKIs due to a secondary-point mutation T790M (threonine to methionine), which increases the affinity for adenosine triphosphate (ATP) to bind to the tyrosine kinase domain [153]. This mutation also affects the binding affinity of the drug [57, 93]. The T790M mutation is known to occur in ∼50-60% of patients [93].

Second-generation TKIs, afatinib and dacomitinib, irreversibly bind to the EGFR tyrosine kinase domain with greater affinity than first-generation TKIs. Dacomitinib was shown to improve progression-free survival (PFS) by 5.5 months compared to gefitinib and improves median overall survival by 7.3 months [112, 147]. When gefitinib or afatinib were applied as first-line therapies to treat treatment naïve EGFR mutant NSCLC patients, afatinib showed slightly better progression free survival than gefitinib (median of 11 months compared to 10.9 months, respectively) at the cost of higher toxicity [111]. Originally, second-generation TKIs were developed to combat T790M mutations. In vitro, afatinib and dacomitinib demonstrated activity against T790M+ mutant cells but at doses that result in toxicity effects in vivo [29]. Hence, second-generation TKIs have not improved the overall survival of T790M+ NSCLC patients and the T790M mutation is the leading cause of failure of first and second-generation TKIs. The pathway to T790M+ EGFR NSCLC is varied. Studies have shown T790M+ cells may be present prior to treatment and may also be aquired during TKI treatment [55]. Although T790M mutations are most common, first and second-generation TKI resistance can occur via alternative routes such as amplification of MET and mutations in the KRAS and BRAF genes (see Figure 1.1.1) [101].

A third-generation TKI, osimertinib (tradename Tagrisso, AstraZeneca) has been developed to target EGFR sensitising mutations (exon 19 deletions and L858R substitutions) and the secondary point T790M mutation. Osimertinib binds irreversibly to the C797 residue on the EGFR tyrosine kinase domain and overcomes the effect of the

T790M mutation that enhances the affinity for ATP to bind to the tyrosine kinase domain [57, 153]. Osimertinib was first FDA approved in 2015 when it was shown to be effective in patients with metastatic EGFR mutated NSCLC with the acquired T790M mutation after EGFR TKI therapy [68]. In clinical trials, osimertinib has demonstrated its superiority over chemotherapy. In a phase III clinical trial, the objective response rate (ORR) of osimertinib was 71% versus 31% in patients treated with combination chemotherapy involving pemetrexed and either carboplatin or cisplatin and PFS for osimertinib was more than double that for chemotherapy (10.1 versus 4.4 months) [98]. Due to osimertinib having higher specificity for EGFR exon 19 and T790M mutations, osimertinib is better tolerated than first-generation TKIs. In a phase III clinical trial, response to osimertinib was significantly longer than first-line gefitinib and erlotinib in treatment-naïve NSCLC patients presenting exon 19 deletions and exon 21 L858R substituitions, with a median progression-free survival of 18.9 compared to 10.2 months [129]. Based on these findings, in 2018, osimertinib was also approved as a first-line therapy for advanced EGFR-mutant NSCLC, regardless of whether T790M+ cells are present.

In recent times, resistance mechanisms to osimertinib have been identified. These include the C797S (cysteine to serine) point mutation in the tyrosine kinase domain. This mutation is the most common mutation in exon 20 and contributes to 10-26% of cases where resistance to second-line osimertinib therapy has occurred [98]. The C797S mutation causes the loss of the covalent bond between the mutant EGFR and osimertinib [82]. The C797S mutation is, however, sensitive to geftinib, making it a popular target for combination first-generation gefitinib and third-generation osimertinib therapy when the T790M mutation is also present. A current clinical trial combining gefitinib and osimertinib is ongoing [119], as discussed further below. Other point mutations resulting in the failure of osimertinib treatment are discussed in the review [82]. In addition to point mutations, bypass signalling mechanisms, such as the amplification of the MET gene leads to the over activation of the c-MET protein. The c-MET protein is a receptor tyrosine kinase on the cancer cell surface which is encoded for by the MET gene. Amplification of this protein drives resistance to osimertinib (see Figure 1.1.1) [82, 101].

Cells which harbour both T790M and C797S mutations are currently fully drug-resistant to three generations of TKI. We consider such cells in this thesis. Fourth-generation TKIs are being developed to combat the C797S mutation [143]. Further, combination therapies involving MET inhibitors have been applied [101, 141]. However, the failure of TKIs is common. Studies that uncover the mechanisms behind this resistance and that can help predict the evolution to treatment resistance in NSCLC are vital in order to determine how best to control/slow disease progression.

In this thesis, we consider cells with single genetic alterations that confer resistance to TKIs. In chapters 2 and 3, we describe the growth and response of heterogeneous populations of tumour cells to TKI treatment whereby these cells have single point mutations like those presented here. Although we focus on EGFR mutations, the parameterisation of the models developed in these chapters is representative of systems of NSCLC mutations that result in drug resistance. In chapters 4 and 5, we work with experimental data describing the growth and response of EGFR mutant NSCLC cell lines with exon 19 deletions, T790M+ and T790M+C797S mutations to treatment with gefitinib and osimertinib. We outline these studies in section 1.3.

1.1.5 Combination therapies applying TKIs

Combination treatments have been used to tackle TKI resistance in NSCLC. Preclinical studies of EGFR TKI treatment combined with radiotherapy have been considered. Further, preclinical and clinical trials applying EGFR TKIs and chemotherapy have taken place. In this thesis, we explore combination TKI therapy applying two TKIs. We refer the reader to the reviews given in [88, 149] for details on EGFR TKI combination therapy with chemotherapy, radiotherapy, immunotherapy and targeted therapies such as MET inhibitors.

Preclinical evidence has suggested that combining geftinib and osimertinib may delay resistance. This has motivated studies in patients to determine the feasibility of dual TKI therapy involving gefitinib and osimertinib. A phase I/II study involving 27 patients studied the feasibility and maximum tolerated dose (MTD) of combining gefitinib and osimertinib for advanced EGFR mutant treatment-naïve patients [119]. This feasibility study found that combination TKI treatment is tolerable as first-line EGFR

NSCLC treatment. Information on survival outcomes and the presence of acquired resistance mechanisms is pending. In chapters 2 and 3 we investigate how simultaneous TKI therapy impacts the way in which tumour composition evolves over time. The relevance of our model to the clinical trial presented here is extensive as the model distinguishes between fully drug-sensitive cells and those with mutations T790M, C797S and T790M+C797S which exhibit different levels of resistance to gefitinib and osimertinib.

An ongoing phase I clinical trial involving 48 patients aims to compare the overall survival of simultaneous versus alternating treatment with osimertinib and gefitinib. For one cohort, concurrent osimertinib (80mg daily) and gefitinib (250mg daily) are applied together. For another cohort, the two drugs are alternated every 4 weeks. Results from this clinical trial are pending [110].

1.2 Mathematical models of tumour growth, drug resistance and treatment response in cancer

The interplay between intrinsic cellular processes and the tumour's interaction with its microenvironment makes experimental studies of drug resistance expensive and time consuming. Using fewer resources, mathematical models have the potential to increase our understanding of these interactions and provide insight into the mechanisms that result in treatment failure. When coupled with pharmacological considerations, mathematical models can play a key part in the exploration of prospective novel treatments. This includes predicting enhanced dose regimens and scheduling protocols that prevent the emergence of drug resistance. Mathematical modelling can be used to study the effect of combination treatments as well as guide experimental design and offer quantitative predictions that can be tested in a non-costly and time efficient manner through the use of fewer experiments.

In order to develop relevant and realistic mathematical models, suitable parameterisation is necessary. To do so, collaboration with experimentalists and a good understanding of the biological mechanisms driving the system are required. Data from in vitro and in vivo experimental studies are useful for parametrising, calibrating and vali-

dating models. In reciprocation, mathematical models can reveal limitations in data and can then be used to propose new experiments that will increase our understanding of the mechanisms underlying a biological system. Such experiments may indicate whether an increased number of parameters within a model can be estimated, thereby increasing the utility of the model as a predictive tool.

A range of approaches can be used to model tumour growth, treatment response and resistance. These approaches include, but are not limited to, stochastic models, continuum models, agent-based models and hybrid modelling techniques. In this thesis, we focus on spatially-averaged stochastic and continuum models of tumour growth and treatment response (more details of which are given later in this section and in section 1.3).

Below, we introduce several mathematical models that are commonly used to describe tumour growth dynamics. These include multi-compartment models that have been used to study the effects of cell-cell interactions and the tumour microenvironment. We then focus on branching-process models that have shaped the way we model the emergence of drug resistance in tumours, due to mutations, and treatment response. We proceed to introduce models used in the pharmaceutical industry to compare drugs used to treat cancer and explain how these models have been used to uncover the impact of resistance on treatment response. Over time, the use of (clinical, real world) data in conjunction with mathematical models, to infer the mechanisms that drive the dynamics of a biological system, has become more widespread. For example, parameter estimation, global sensitivity analysis and parameter identifiability analysis can provide additional insight into factors which influence tumour cell dynamics, drug resistance and treatment response. We will use these methods in chapters 4 and 5 and review them in sections 1.2.5 and 1.2.6. Further information about mathematical models of tumour evolution, treatment scheduling and drug-effect modelling can be found in the following reviews [3, 6, 15, 38, 44, 80].

1.2.1 Simple tumour growth models

Phenomenological models describing how tumour volume or mass change over time often consist of a single ordinary differential equation (ODE). The simplest model assumes that

the tumour volume, $V(t)$, grows exponentially over time so that,

$$\frac{dV}{dt} = rV, \tag{1.2.1}$$

where the parameter r denotes the net growth rate of cells (birth minus death). Exponential growth captures the early stages of tumour growth. It does not, however, capture tumour dynamics at longer times in vitro and/or in vivo: in both cases, growth eventually saturates.

Mathematical models that capture the full dynamics of tumour growth assume the growth rates are sigmoidal functions of the tumour volume, $V(t)$ [44]. Consider, for example, the Gompertz growth model which can be written as

$$\frac{dV}{dt} = rV \log\left(\frac{K}{V}\right), \tag{1.2.2}$$

where r describes the net proliferative ability of the cells and K describes the equilibrium size or the carrying capacity of the tumour. The Gompertz model has been used to model tumour growth and can be readily modified to account for treatment [99, 105].

The logistic equation also exhibits sigmoidal growth and can be written as follows,

$$\frac{dV}{dt} = rV\left(1 - \frac{V}{K}\right), \tag{1.2.3}$$

where r denotes the net growth rate of cells (birth minus death) and K denotes the carrying capacity. The logistic equation is a mainstay in ecology and has also been used to model the growth of bacterial populations [44]. The growth curves for the Gompertz and logistic models are similar. Further, both provide phenomenological insight which can be used to predict tumour growth trajectories.

More complex models of tumour growth include the Richards growth equation model,

$$\frac{dV}{dt} = rV\left(1 - \left(\frac{V}{K}\right)^{\beta}\right), \tag{1.2.4}$$

where r and K are the same as in equation (1.2.3) and the parameter, $\beta \geq 0$ describes the strength of competition for resources; whereby, for lower values of β, the strength of competition is increased [117].

The generalised-logistic model is more complex (i.e. contains more parameters) than the Richards growth equation and is written as follows,

$$\frac{dV}{dt} = rV^{\alpha}\left(1 - \left(\frac{V}{K}\right)^{\beta}\right)^{\gamma}, \tag{1.2.5}$$

where the parameters α and γ are exponents for the exponential growth and growth limiting factor terms, respectively. The parameter, β, denotes the competition strength parameter as in equation (1.2.4). We refer to reader to the review in [136] which contains analysis of the models defined by equations (1.2.2)-(1.2.5).

The Bertalanffy equation can be written as,

$$\frac{dV}{dt} = \theta V^{2/3} - \eta V. \tag{1.2.6}$$

It assumes that growth occurs at a rate proportional to the surface area where θ denotes the rate of cell growth and η denotes the rate at which the tumour loses mass due to cell death [44, 78].

The choice of model used to describe tumour growth dynamics is strongly driven by the type of data available, for example, in vitro, mice or human data. The Gompertz model is routinely used to model in vivo tumour growth dynamics [79, 121]. In vitro tumour cell proliferation assays have been modelled using the logistic growth model [144]. Sarapata et al. fitted the exponential, Gompertz, logistic growth and the Bertalanffy models to a range of different mice and human data sets (bladder cancer, breast cancer, colon cancer, head and neck squamous cell carcinoma, hepatocellular carcinoma, lung cancer, melanoma, ovarian cancer, pancreatic cancer and renal cell carcinoma) [121]. Sarapata et al. found, when comparing the fits of each model to the different mice and human data sets, the logistic growth model gave a better fit to the liver, lung, breast and melanoma data than the Gompertz model. However, the Gompertz model gave a better fit to the ovarian and pancreatic data sets. In most cases, the Bertalanffy equation and the exponential model produced bad fits to the data relative to the Gompertz and logistic models. However, for colon cancer, the Bertalanffy equation gave a better fit [121].

In chapter 4 we examine the suitability of the above models, with the exception of the Bertalanffy model, to describe EGFR NSCLC cell lines that are grown in vitro to

100% confluence and whose growth is limited by spatial effects. We focus on the ability to infer system parameters as model complexity increases.

1.2.2 Multi-compartment models of heterogeneous tumours

Cancer is an evolutionary process [106]. Tumours can be viewed as complex ecosystems containing multiple cell populations, with different genetic, epigenetic and phenotypic characteristics (e.g. drug resistance) [94]. Simple extensions of the single ODE tumour growth models presented above can be used to describe heterogeneous populations: models of this type are often termed multi-compartment models [99, 108]. For example, a simple two-compartment ODE model, in which sensitive and resistant cells compete, can be used to study the growth and response to treatment of heterogeneous tumours. The concept of harnessing competition effects to control the emergence of drug resistance during treatment was proposed by Monro and Gaffney and others [41, 99]. They adapted a Gompertz model of tumour growth to account for two populations; chemo-sensitive and chemo-resistant cells. When considering two populations (sensitive and resistant cell lines, for example) in a Gompertz growth model, equation (1.2.2) can be modified to a system of ODEs of the form,

$$\frac{dS}{dt} = rS\log\left(\frac{K}{S+R}\right), \tag{1.2.7}$$

$$\frac{dR}{dt} = rR\log\left(\frac{K}{S+R}\right), \tag{1.2.8}$$

to incorporate the effects of competition between sensitive and resistant cell lines denoted by $S(t)$ and $R(t)$, respectively. Monro and Gaffney found that reducing the dose of administered chemotherapy could prolong survival by increasing competition between the drug-sensitive and drug-resistant cell populations. More recent mathematical models which consider the effects of competition on tumour cells have aided the design of clinical trials in which competition between cells is exploited to suppress the outgrowth of drug-resistant populations [155].

Through combining multi-compartment models with experimental data, it is possible to investigate the effect of interactions between different cell populations on the tumour's growth dynamics and its response to treatment [5, 72, 92, 108]. A recent study by

Paczkowski et al. used a Lotka-Volterra model to study interactions between radio-sensitive and radio-resistant prostate cells [108]. They fitted their model to tumour volume data from homogeneous and heterogeneous cell populations and were able to characterise the types of interactions between different cell populations and how these interactions influenced response to radiotherapy. Kaznatcheev et al. quantified the types of interactions between sensitive and resistant NSCLC cell lines to alectinib, an ALK tyrosine kinase inhibitor [72]. They fitted exponential growth models to dynamic data from different cell phenotypes to estimate their growth rates. These studies show how cell-cell interactions impact tumour heterogeneity. Other studies, reviewed in [133], have characterised interaction types and the influence of the microenvironment on tumour cells. The types of interactions between different cell types and their effects on tumour composition are examined in the modelling study of chapter 2 of this thesis.

Multi-compartment models can be used to investigate the effect that processes such as competition have on tumour response to treatment. Including additional biological detail in such models can cause problems with model calibration, due to lack of suitable data. When fitting models that distinguish two or more species to data from a single dependent variable, it is often difficult to accurately estimate all model parameters. For example, given dynamic data on total tumour cell numbers only, it can be difficult to accurately infer a tumour's internal cellular composition [36].

1.2.3 Stochastic models of drug resistance and treatment response

In chapter 3 of this thesis, we develop stochastic models of tumour growth to study the emergence of drug resistance and response to TKI treatment.

Many stochastic models describing the emergence of resistance have been proposed. The work of Luria and Delbrück in 1943 has influenced many studies of the accumulation of drug resistance in cancer. Motivated by in vitro experiments, they formulated an exponential branching process growth model to model the rate at which mutations, which confer resistance to a virus, occur during bacterial growth. This concept of resistance and its evolution also applies to tumour cells, and the distribution of mutant clones derived from the Luria-Delbrück model is widely used to estimate mutation rates in cancer

[38, 85, 157]. Their work underpins many ensuing models developed to understand how resistance evolves in cancer [38, 157].

The Luria-Delbrück model was extended by Goldie and Coldman to include cell death [27]. Using a stochastic birth-death process which accounts for mutations from a sensitive to a resistant cell type, they found that with a higher rate of cell death, the predicted size of the resistant cell population was much larger than it was in similar models which neglect cell death [27]. In more recent work, Iwasa et al. used a birth-death branching process to describe an exponentially growing tumour population, starting from a single sensitive cell [67]. They incorporated mutations into the model that give rise to a resistant cell with a fixed probability per cell division. They found that the proportion of resistant cells in the tumour at detection depended on the detection size and the probability of mutation, with larger detection sizes increasing the probability of resistance (and similarly for the probability of mutation). Concurring with the results of Goldie and Coldman, they found the occurrence of resistance to increase with an increased rate of apoptosis [67]. Later work by the same authors with Haeno extended Iwasa et al.'s work to include two mutations during expansion [53].

Several authors have considered the effects of competition on the emergence of drug resistance by employing a birth-death process with competition [38]. For example, Dewanji et al. considered mutant populations that grow via a birth-death process with Gompertz growth laws. The authors found that including cell death, as well as Gompertz growth of mutant cells in such models resulted in larger variation in the Luria-Delbrück distribution of mutant cells compared with earlier models [35]. Motivated by these studies, in chapter 3, we develop a stochastic branching process, with random intrinsic mutation events, and logistic growth. We use this model to investigate how intrinsic mutation rates (Darwinian mutations) and the growth rates of resistant cells affect the variability in tumour composition at diagnosis.

The development of models to investigate treatment scheduling and the emergence of drug resistance stems from the seminal work of Norton and Simon [105]. They introduced the concept of dose-densification to tackle *kinetic resistance* that describes resistance due to there being only a small fraction of tumour cells that are dividing at a given

time and are therefore sensitive to chemotherapy [105]. Dose-densification describes the application of higher doses of drug as the tumour size decreases. Another form of resistance to treatment is *molecular resistance*. Molecular resistance mechanisms include those that reduce or eliminate a cell's response to treatment, as exemplified by drug efflux pumps, genetic alterations of the drug binding targets, resulting in ineffective drug binding and the upregulation of alternative survival pathways [14, 62].

The seminal work of Goldie and Coldman involves stochastic models with random mutation events [27]. Goldie and Coldman concluded that, in order to minimise the probability of resistance, early and aggressive application of chemotherapy is necessary. Further, if toxicity effects prevent the simultaneous application of two drugs, the chance of a cure is increased by the rapid alternation of treatment, known today as the Goldie-Coldman hypothesis [26, 48, 49]. However, the Goldie-Coldman hypothesis failed in a clinical trial that compared sequential and alternating chemotherapy protocols in breast cancer patients [11]. It was found that the sequencing of chemotherapy gave better survival rates than when the drugs were applied in an alternating fashion. For sequential and alternating treatment, the relapse-free survival was 42% versus 28%, respectively and the total survival was 58% versus 44%, respectively. Day expanded on the Goldie-Coldman hypothesis to consider asymmetry of the two drugs [33]. Day presented the 'worst-drug rule' whereby, in order to kill resistant clones, when two non-cross resistant drugs are available that have different efficacies, the less effective drug should be applied first, followed by the better drug. When applied to NSCLC patients, the hypothesis failed due to the rapid deterioration of the patients following first-line treatment [52].

Nonetheless, Goldie and Coldman have shaped the way in which we model treatment and molecular resistance. More recent modelling approaches have been designed to investigate the effects of combination EGFR TKI therapy on NSCLC tumour cell dynamics. Chmielecki et al. derived a population growth model consisting of NSCLC sensitive and resistant cells [24]. They predicted that, combining high and low dose treatment in short bursts would enhance the killing of resistant cells. Results were promising in mouse models, but the effect was not reproduced in the clinic due to the sufficient peak serum concentrations of erlotinib not being reached as a result of toxicities [24, 151].

The failure of mathematically inspired, pre-clinical treatment scheduling studies in the clinic could be driven by variability in tumour-specific factors; a patient's tumour may contain cells with different growth rates and mutation propensities. Further, mutations are random events; they generate stochastic cell dynamics, which result in differences in compositions across tumours.

In chapter 2, we use a deterministic model to examine how changes in parameter values associated with mutation rates, growth rates and cell death rates influence treatment response. The work in chapter 2 may increase our understanding of the factors within a tumour that drive variation in response to treatment across patients. In chapter 3, we investigate the role of stochasticity on pre-existing resistance and treatment response using realistic tumour sizes and parameter values. Hence, in chapters 2 and 3, we determine factors that influence inter-patient heterogeneity (differences in patients' tumours which are of the same histological subtype). Further, we determine factors influencing intratumoural heterogeneity across tumours with the same assumed characteristics, in particular, we examine how such intratumoural heterogeneity changes when mutation rates and cell growth rates change.

1.2.4 In vitro models of TKI response

During preclinical drug-development, in vitro models and in vivo mouse model studies are often used to understand the pharmacokinetic profiles of drugs, their mechanisms of action and toxicity profiles, drug-drug interactions and dose regimens. For example, in vitro experiments are used to measure the dose-response of TKIs in NSCLC [59]. From these experiments, key features of the drug-effect on cells and characterisation of drug-sensitivity can be measured. In pharmacokinetic-pharmacodynamic (PK-PD) modelling, dose-response curves describing the response of cancer cells to different drug concentrations are often modelled by the Hill equation which allows comparison of pharmacological drug profiles [1, 51, 137]. In pharmacology, the Hill equation relates back to the binding of ligands to macromolecules on a cell surface and reflects the cell's response to the ligands. The non-linearity of the dynamics, in particular, the saturating behaviour of the model, reflects the occupancy of the macromolecules by the ligands. Further, the binding of a ligand to a macromolecule is often enhanced by the binding of

other ligands, called cooperative binding. The higher the degree of cooperative binding, the steeper the curve is that describes the relationship between ligand concentration and macromolecule occupancy.

The Hill equation consists of four parameters and can be written as follows,

$$h(c) = \frac{E_{\max} c^m}{IC50^m + c^m}. \tag{1.2.9}$$

Equation (1.2.9) describes the relationship between the drug concentration, c, and its effect on its target. In equation (1.2.9), the maximum efficacy, $E_{\max}$, represents the greatest attainable drug effect. The half-maximal drug concentration, $IC50$, is the drug concentration that achieves 50% of the maximum drug effect. It describes the potency of the drug; lower $IC50$ values correspond to greater drug potency. The gradient of the response, m, describes the cooperativity of the ligand binding ($m > 1$, positive cooperativity, $m < 1$, negative cooperativity) and summarises the change in cellular response per unit dose.

The parameters of the Hill equation are used to compare pharmacological drug profiles. Such parameters can be used to compare the sensitivity and resistance of cancer cells and determine the impact of different TKIs on cells with certain mutations. In vitro, $IC50$ values are widely used to guide TKI choices in the clinic [29, 51, 137]. For example, Cross et al. performed dose-response experiments on NSCLC cell lines to compare the $IC50$ values for osimertinib to those of earlier generation TKIs for several EGFR mutants [29]. Hirano et al. used in vitro experiments to predict the range of doses of first, second and third-generation TKIs that provide safe, effective therapy [59]. They did so by comparing the potency ($IC50$ values) of the TKIs to EGFR mutant NSCLC cells with different mutations. Further studies have proposed that the binding affinity of the drug, characterised by the parameter m in Equation (1.2.9), also provides a measure of drug resistance [137]. Vainstein et al. used the Hill equation to model the response of ABL kinase inhibitors in CML and demonstrated that smaller values of m correspond to resistance in BCR-ABL mutants. A review of the applications of the Hill equation in pharmacology is given by Goutelle et al. in [51].

The Hill equation expresses the way in which the drug concentration acts and estimates of the associated parameter values which can help characterise the drug-effect.

In chapter 5, we model the dose-response of first and third-generation TKIs on NSCLC cell lines with different EGFR mutations. We estimate Hill equation parameters and, in so doing, gain insight into their mechanisms of action for different cell lines.

1.2.5 Sensitivity analysis

Sensitivity analysis can be used to investigate the effect of model parameters on model behaviour. Local sensitivity analysis involves varying a single model parameter and holding the remaining parameters fixed. In doing so, the effect of varying a specific parameter on a particular model output can be quantified and compared with the effect of varying other model parameters to identify the parameter which has the greatest effect on model output. A shortcoming of local sensitivity analysis is that parameters are varied one-at-a-time and considered only at a single point in parameter space, with the impact of their variation on quantities of interest assumed to be linearly independent.

Global sensitivity analysis techniques have been developed to address these shortcomings. It involves simultaneous sampling of all model parameters and identifies the effect of single and combinations of model parameters on the model output [90]. The extended Fourier amplitude sensitivity test (eFAST) simultaneously searches the parameter space and, by assigning a frequency to each model parameter, measures the propagation of that frequency through the model so that the influence a parameter has on model output can be quantified [120]. eFAST extends FAST and measures the contribution of higher order terms and thus correlations with other parameters for a given parameter of interest [31, 90, 120]. In chapter 2, we apply eFAST to our model to determine factors influencing the proportions of resistant cells in a tumour over time and at diagnosis.

1.2.6 Parameter inference

When working with data, one should ensure all model parameters are structurally identifiable. Structural identifiability analysis is carried out prior to the consideration of data and relates to the postulated structure of the model alone [115]. A model is structurally identifiable when all model parameters can be uniquely determined from the output measurements of the experiment, whereby unique model parameters result in unique

observed model outputs [7]. Structural identifiability is a prerequisite to practical identifiability, parameter estimation and experimental design. Structural identifiability is well understood for linear ODEs, however, the tractability of the techniques used to assess structural identifiability for non-linear ODEs is a problem. When working with models with increasing complexity (i.e. the number of model parameters), different conclusions on structural identifiability can be drawn when using different techniques [22]. A comparison of methods used to perform structural identifiability analysis on non-linear ODE models can found in the reviews [22, 97].

In this thesis, we employ structural identifiability methods in chapters 4 and 5. In chapter 4, we use two different approaches to assess the structural identifiability of models that describe the in vitro growth of NSCLC cell lines; the generating series approach, implemented using GenSSI2 in MATLAB and, a differential algebra approach, implemented using DAISY [8, 23, 84]. We begin by using GenSSI2 due to its ease of use and the software provides output information on the global and local identifiability and the non-identifiability of model parameters. We find however that only local identifiability can be assessed when considering models with exponent parameters. In response to this, for a given model with exponent parameters, we introduce new variables and augment the state-space model to obtain a system of polynomials and rational functions. In this form, the model is in the appropriate form for applying a differential algebra approach to assess structural identifiability. We can then assess the global identifiability of the models in this form when using DAISY. In chapter 5, we use a generating series approach, implemented using GenSSI2 in MATLAB, to assess the structural identifiability of a model that describes the response of NSCLC cells to TKI treatment.

For all models that we consider in this thesis, parameters represent biological processes such as cell growth rates or the maximum rate of inhibition of a TKI on a NSCLC cell. For the development of mathematical models that describe the system of interest accurately, we must ensure that our models are parametrised appropriately. In chapters 2 and 3 of this thesis, we develop theoretical models of tumour growth and treatment response. We parametrise these models using estimates from specific NSCLC TKI studies and more general studies involving other cancer cells and drugs

[24, 29, 34, 48, 72, 73, 99, 131, 134, 137]. We note that model parameter values may vary across patients and cancer types which may result in different model predictions; this is addressed in chapter 2. In chapters 4 and 5, we parametrise our models using preclinical in vitro data sets.

A number of methods can be used to infer model parameters from data. A frequentist approach assumes that parameter of interest, θ is unknown, but that a unique parameter value exists and is fixed, and the only information that contributes to its estimated value is the data. A Bayesian approach assumes the parameter of interest, θ, is a random variable, described by a probability distribution with a mean and variance. When using a Bayesian approach, the experimenter defines an initial distribution, or prior, for the parameter, θ. The prior incorporates any initial belief about the parameter. It is updated in light of the data via a likelihood function using Baye's rule. The resulting probability distribution for the parameter of interest is defined as the posterior distribution. Baye's rule is written as;

$$p(\theta|y) \propto p(y|\theta)p(\theta), \tag{1.2.10}$$

$$posterior \propto likelihood \times prior,$$

where $p(\theta|y)$ denotes the posterior distribution for the parameter θ, conditioned on data y, $p(y|\theta)$ is the likelihood function and $p(\theta)$ is the prior. The posterior distribution presents information about the parameter, θ given the data, y. The likelihood function incorporates information from the data. Hence, in Bayesian inference, we obtain a posterior distribution on the parameter space that assigns a higher probability to areas of the parameter space that are in closer agreement to the observed data [58]. The structure of the posterior distribution can be used to assess the uncertainty in the parameter and its practical identifiability which we will discuss further in chapter 4 in which a Bayesian approach to parameter inference is applied [58, 116].

In order to generate posterior distributions of model parameters, sampling algorithms are used. Markov Chain Monte Carlo (MCMC) defines a class of algorithms for sampling from the parameter space. MCMC works by generating independent and identically distributed samples (iid). By generating these samples, MCMC produces a Markov chain consisting of samples of the posterior distribution whose limiting distribution is

the posterior distribution [58]. MCMC is commonly implemented using the Metropolis-Hastings algorithm [96]. In chapter 4, we use MCMC to quantify the uncertainty of our model, describing the growth of NSCLC cell lines when fitted to the data. We provide more details of the algorithm used to implement our MCMC in chapter 4.

Optimisation methods, such as non-linear least squares (NLLS) can be used to estimate model parameters via a frequentist approach and provide a measure of uncertainty about that estimate. NLLS is widely used in the pharmaceutical industry and results are obtained more quickly compared to Bayesian inference techniques such as Markov chain Monte Carlo (MCMC).

The ability to infer model parameters when fitting models to experimental data depends on the amount of available data and the level of noise within a data set. Practical identifiability describes the ability for parameters to be estimated from finite noisy data [115, 116, 125]. A model could be structurally identifiable but practically non-identifiable. Methods for assessing practical parameter identifiability include, but are not limited to, the profile likelihood method and the MCMC approach [58, 115, 116], both of which have been applied in a biological setting [13, 58, 75, 116, 125, 127]. In chapters 4 and 5, we employ the profile likelihood method to assess the practical identifiability of the unknown model parameters. In this method, likelihood-based confidence intervals of model parameters are calculated. We use these intervals to examine whether a parameter can be robustly inferred from the available data [115]. We provide more details of the profile likelihood method in chapter 4.

In chapter 4, we employ both the profile likelihood method and the MCMC approach, but focus primarily on the profile likelihood method to assess the practical identifiability of the model parameters. To tackle the issue of structural and practical non-identifiability, two methodologies can be used: (i) model reduction where reparameterisation of the model can take place in which identifiable groups of parameters are considered and/or, non-identifiable model parameters are fixed; (ii) further data that will allow for the dependent variables (e.g. live and dead tumour cell populations) of the system to be identified. For model reduction, structural identifiability techniques such as the Taylor series approach and the differential algebra approach can help deter-

mine identifiable groups of model parameters. However, it may not be useful to group model parameters to achieve identifiability, especially when the aim is to uniquely estimate parameters specific to a biological mechanism, for example, cell growth rates. An alternative is to fix non-identifiable parameters. By fixing non-identifiable parameters however, a bias on the remaining estimated parameters may occur as estimated values of these parameters could depend on the fixed value of the non-identifiable parameter. In chapters 4 and 5, we carry out model reduction to achieve practically identifiable models by fixing practically non-identifiable parameters. In chapter 5, we design experiments to obtain additional measurements that we propose will increase the identifiability of our model.

In-depth descriptions of the methods presented here are given in the methods sections and appendices of the relevant thesis chapters.

1.2.7 Summary

In section 1.2 we have reviewed mathematical models that describe tumour growth, the emergence of drug resistance and treatment response. Different modelling frameworks capture different aspects of tumour growth and can be used to determine the tumour-specific factors that influence such behaviours. Data-driven modelling techniques allow for the identification of biological parameters, and, associated mechanisms that drive treatment response and tumour growth. Modelling techniques can be used to suggest experiments to test hypotheses based on modelling observations. In conclusion, progress has been made in the field of mathematical modelling applied to drug resistance in cancer and treatment response but more work is required to develop models that can be used to make reliable clinical predictions.

1.3 Thesis outline, aims and objectives

In this thesis we present four studies that integrate mathematical modelling and scheduling design, in vitro and in vivo experimental data analysis, identifiability analysis and experimental planning. Chapters 2 and 3 are theoretical, covering both continuum and stochastic settings and are motivated by NSCLC specific parameters and treatment design in order to understand which factors influence NSCLC tumour cell heterogeneity

and response to TKI treatment. In chapters 4 and 5, we exploit available in vitro data to develop and validate mathematical models that describe the growth of different NSCLC cell lines and their response to different TKIs. We also use our models to identify limitations in the data and propose experiments that can be used for preclinical development of TKI treatment, and, to increase our understanding of their mechanisms of action on NSCLC cells.

The remainder of this thesis is organised as follows and in what follows, we summarise the aims and objectives for each chapter:

In chapter 2, we aim to determine what tumour-specific factors in a heterogeneous tumour contribute to the emergence of drug resistance and response to TKI treatment. In particular, what factors influence the level of resistance in a tumour at diagnosis and to what extent do these factors influence tumour response to first-line combination TKI treatment. To answer this question, we introduce a spatially averaged continuum model describing the growth of a NSCLC tumour and its response to combination TKI treatment where we consider two drugs. The tumour consists of four cell populations with varying levels of resistance to two TKIs; fully drug-sensitive, partially drug-resistant (cells are resistant to one drug out of the two) and fully drug-resistant. Factors included in the model are; mutation rates, cell growth rates and the types of interactions between cells (e.g. beneficial or detrimental). In order to determine which factors influence the prevalence of resistance in a tumour at diagnosis, we perform a global sensitivity analysis of our model using eFAST. Our analysis shows that the propensity for a tumour to mutate and the rate at which resistant cells divide have a strong influence on the proportion of resistant cells in a tumour over time and at diagnosis and, using our model, we predict vastly different proportions of pre-existing resistant cells when these factors are varied. We find changes in the mutation rate and the growth rates of resistant cells to greatly influence a tumour's response to combination TKI treatment. In more detail, we consider two treatment protocols; the first applies two TKIs simultaneously, at moderate doses until progression occurs. The second protocol applies two TKIs simultaneously, at high doses for a week, followed by a 3 week break, which is repeated over the course of a year. We demonstrate that tumours which present similar progressions can consist of

vastly different cellular compositions.

In chapter 3 we extend the model from chapter 2 and introduce stochasticity. We aim to determine whether the effect of stochasticity on the level of variation in cell growth dynamics, pre-existing resistance levels and treatment response across a set of stochastic simulations is more or less pronounced when different tumour-specific properties are assumed in the model. In particular, we ask, to what extent changes in the (i) propensity for cells to mutate and, (ii) the relative growth rate of resistant cells to sensitive cells influence the variation in untreated cellular dynamics and pre-existing resistant proportions at diagnosis. We also ask how different mutation rates contribute to the level of variation in treatment response across stochastic simulations. To answer these questions, we extend the model developed in chapter 2 to include stochasticity. We simulate our model using a τ-leap approximation method with adaptive step size calculation which allows tumours with large cell numbers (10^8 cells) to be considered. We show that mutation rates and the growth rates of resistant cells relative to sensitive cells drive variation in tumour composition at diagnosis across a set of stochastic realisations. In particular, we demonstrate that lower mutation rates and higher growth rates of resistant cells increase the deviation between simulations in a set of stochastic simulations, and, deviation between the stochastic models and the deterministic models that are presented in chapter 2. We show that lower mutation rates drive variation in tumour response to first and second-line TKI treatment and we observe increased deviation between simulations and the stochastic and deterministic models. We investigate the effect of the proportion of pre-existing resistant cells in a tumour on its response to first-line treatment and we find a noticeable decrease in treatment efficacy for tumours with large pre-existing resistant proportions. Further, we investigate the extent to which proportions of resistant cells at progression on first-line treatment influence a tumour's response to second-line treatment and find there to be a considerable decrease in the efficacy of second-line treatment when larger resistant proportions at progression on first-line treatment are predicted.

In the second part of the thesis, we move away from theoretical studies and turn to studies involving mathematical models coupled with preclinical data sets.

In chapter 4, we consider in vitro tumour growth data describing the growth of four NSCLC cell lines. We aim to determine whether, given the data available to us, we can determine a suitable model from a selection of models from which we can uniquely estimate the growth rates of the different NSCLC cell lines and, by so doing, characterise different NSCLC cell lines with respect to their growth rates. Upon selecting the appropriate model, we also aim to determine to what extent the parameters of the selected model can be estimated when fitted to the in vivo tumour growth data. To address these aims, we consider a selection of models with sigmoid growth dynamics that vary in complexity and we develop a model selection procedure so as to determine the appropriate model to describe the in vitro data. The model selection procedure combines information theory, used to assess the goodness-of-fit of the models to the data and, identifiability analysis which allows us to analyse whether, given the model and the data available to us, we are able to uniquely estimate parameters. We show that the most complex models yield the best information criteria/goodness-of-fit metrics yet suffer from practical identifiability issues. We select the model with the smallest number of practically non-identifiable model parameters and, by fixing these parameters, achieve practical identifiability. We perform MCMC to quantify the uncertainty in the model fit to the in vitro data. Our MCMC demonstrates practical identifiability of the model which corroborates results found using the profile likelihood method. We use our practically identifiable model to compare the growth properties of different NSCLC cell lines in vitro. We find that the practical identifiability of this model decreases when we apply it to coarse-grained, noisy data describing the in vivo growth dynamics of NSCLC cell lines.

In chapter 5 we extend the models from chapter 4 to include the effects of treatment applying a single TKI. We consider preclinical in vitro dose-response curves describing the response of four NSCLC cell lines in a single TKI where two different TKIs are considered. The four cell lines have varying levels of resistance to the two TKIs due to how many resistance-conferring mutations they have. We aim to determine whether, given the data available to us, we can uniquely estimate the drug-effect parameters for the different TKIs and the different cell lines. By so doing, we aim to determine how different

TKIs affect different NSCLC cell lines. In the case where unique estimates cannot be determined, we ask whether we can propose experiments whereby additional measurements are taken and whether, when fitting the models to the additional data, we can yield unique estimates of the model parameters. To answer our research questions, we develop a model that consists of live and dead cells. We assume the live cells are killed by the drug and move into the dead cell compartment. We assess structural identifiability of the model and we employ the profile likelihood method to assess whether the parameters can be uniquely determined when we fit the model to the in vitro dose-response curves. We find our model to be structurally identifiable. However, by performing practical identifiability analysis, we show that it is not possible to identify drug-effect mechanisms given the available data and we demonstrate how misleading predictions of tumour cell trajectories can be made when our model is practically non-identifiable (i.e. contains practically non-identifiable parameters). Through fixing the practically non-identifiable parameters, we achieve a practically identifiable model which we then fit to the data, which allows us to compare the sensitivity of different NSCLC cell lines to different TKIs. On a *post hoc* basis, we propose experiments that could be performed in a preclinical setting to determine how TKIs act on the NSCLC cells and we test the suitability of these experiments using synthetic data sets.

In the final thesis chapter, we summarise our findings, discuss their implications and outline possible directions for future work.

Chapter 2

A deterministic model of tumour growth and response to combination TKI treatment

2.1 Introduction

In this chapter, we develop a deterministic model describing the growth of a tumour consisting of four cell populations, with varying levels of resistance to two tyrosine kinase inhibitors (TKIs) (see chapter 1 for descriptions of TKI treatment). Our model accounts for mutations and the effects of competition and cellular interactions on cell growth.

We perform a global parameter sensitivity analysis of our model using the extended Fourier amplitude sensitivity test (eFAST) [120]. In this way, we identify those model parameters that have the greatest influence on the proportion of fully drug-resistant cells in an untreated tumour at, (i) diagnosis and, (ii) over time. We use results from our sensitivity analysis to investigate how proportions of fully drug-resistant cells at diagnosis change as we vary parameters that are specific to non-small cell lung cancer (NSCLC). We also use our model to examine the effects of cellular interactions on tumour cell dynamics. We then extend our model to study the effect of treatment with two TKIs. We investigate the effects of varying influential parameters (subject to eFAST results) on the outcomes of first-line treatment. These parameters include the probability of mutation, the growth rates of resistant cells and the effect of sensitive cells on the growth rates of resistant cells. We show that tumours with similar progression times can have vastly different levels of resistance. This work reveals how varying factors in the tumour, such as a tumour's propensity to mutate and the rate at which cells divide, may

drive patient heterogeneity at diagnosis and diverse responses to treatment on first-line TKIs.

In chapter 1, we discussed models developed by Dewanji et al. where competition and cell death is considered when modelling resistant tumour cell populations [35]. Motivated by these studies, in this chapter we consider the prospect of generalising such models to consider TKI treatments, where multiple inhibitors are available but all of which ultimately suffer from resistance. Generalising previous work to consider multiple treatment modalities that ultimately exhibit resistance represents the novel aspects of the modelling developed in this chapter, as does tailoring this framework to TKIs. In chapter 1, we also discussed models developed by Kaznatcheev et al. whereby population-dependent growth rates were considered when modelling an exponentially growing tumour population consisting of sensitive and resistant cells [72]. In this chapter, we consider population-dependent birth and death rates in our model which consists of four different cell populations as opposed to two as in [72].

The remainder of this chapter is structured as follows. In section 2.2, we develop our deterministic model of tumour growth. The methods used to analyse the model are described in section 2.3. In section 2.4, we begin by simulating our model and illustrating its behaviour and we then present results of a global sensitivity analysis performed on the model. We also present results on predicted proportions of fully drug-resistant cells at diagnosis whilst varying key model parameters, which include, proportions of cells that mutate per cell division and cell growth rates. In section 2.5, we extend our model to account for treatment with TKIs. We present results for first-line treatment outcomes in section 2.6. The chapter ends in section 2.7, where we discuss our findings.

2.2 Untreated tumour growth model

2.2.1 Model development

We develop a well-mixed model to describe the growth of a tumour consisting of four cell populations with varying levels of resistance to two TKIs. The four cell populations are as follows: $S(t)$ denotes the number of cells sensitive to drugs one and two; $R_1(t)$ denotes the number of cells resistant to drug one and sensitive to drug two; $R_2(t)$ denotes

the number of cells resistant to drug two and sensitive to drug one; $R_{12}(t)$ denotes the number of fully drug-resistant cells (see Figure 2.2.1 for model schematic). We assume single mutations confer resistance to drugs one or two, and cells with both mutations (R_{12} cells) are resistant to both drugs. We begin by introducing a system of ordinary differential equations (ODEs) describing the growth of an untreated tumour.

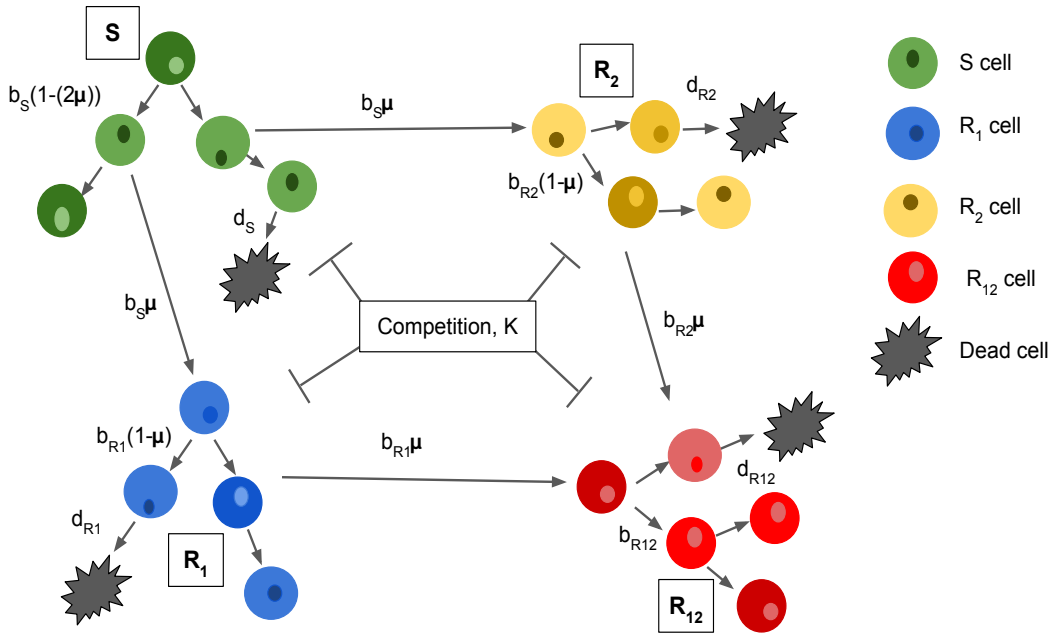


Figure 2.2.1: Schematic of our mathematical model of tumour growth in the absence of treatment. Four cell populations are considered. Sensitive cells (green cells), are denoted by $S(t)$; cells that are sensitive to drug two and resistant to drug one (blue cells) are denoted by $R_1(t)$; cells that are sensitive to drug one and resistant to drug two (yellow cells) are denoted by $R_2(t)$; fully drug-resistant cells (red cells) are denoted by $R_{12}(t)$. The parameter μ denotes the proportion of cells that mutate per cell division from (S to R_1), (R_2 to R_{12}), (S to R_2) and (R_1 to R_{12}). S cells divide to produce S cells, ($S \to S + S$), R_1 cells ($S \to S + R_1$) and R_2 cells ($S \to S + R_2$) at rates, $b_S(1 - 2\mu)$, $b_S\mu$ and $b_S\mu$, respectively. S cells die at rate, d_S and the net growth rate of S cells is $r_S = b_S - d_S$. Similarly, R_1 cells divide to produce R_1 cells ($R_1 \to R_1 + R_1$) and R_{12} cells ($R_1 \to R_1 + R_{12}$) at rates, $b_{R_1}(1 - \mu)$ and $b_{R_1}\mu$, respectively. R_1 cells die at rate, d_{R_1} and have a net growth rate of $r_{R_1} = b_{R_1} - d_{R_1}$. Similarly for R_2 cells. R_{12} cells divide to produce R_{12} cells at rate, $b_{R_{12}}$ and die at rate, $d_{R_{12}}$ and have a net growth rate, $r_{R_{12}} = b_{R_{12}} - d_{R_{12}}$.

Governing equations

The dimensional equations governing the growth of the four cell populations are as

follows:

$$\frac{dS}{dt} = \underbrace{r_S S\Big(1 - \frac{S + R_1 + R_2 + R_{12}}{K}\Big)}_{\text{logistic growth with competition for resources/space}} - \overbrace{\mu b_S S}^{\text{mutations to } R_1} - \underbrace{\mu b_S S}_{\text{mutations to } R_2}, \tag{2.2.1}$$

$$\frac{dR_1}{dt} = r_{R_1} R_1\Big(1 - \frac{S + R_1 + R_2 + R_{12}}{K}\Big) + \mu b_S S - \overbrace{\mu b_{R_1} R_1}^{\text{mutations to } R_{12}}, \tag{2.2.2}$$

$$\frac{dR_2}{dt} = r_{R_2} R_2\Big(1 - \frac{S + R_1 + R_2 + R_{12}}{K}\Big) + \mu b_S S - \overbrace{\mu b_{R_2} R_2}^{\text{mutations to } R_{12}}, \tag{2.2.3}$$

$$\frac{dR_{12}}{dt} = r_{R_{12}} R_{12}\Big(1 - \frac{S + R_1 + R_2 + R_{12}}{K}\Big) + \mu b_{R_1} R_1 + \mu b_{R_2} R_2. \tag{2.2.4}$$

The parameter, r_S (units day^{-1}), in equation (2.2.1) denotes the net growth rate of S cells and is the difference between the birth rate of S cells, denoted by b_S (units day^{-1}), and the death rate of S cells, denoted by d_S (units day^{-1}). Hence, the net growth rate of S cells, r_S, is written as,

$$r_S = b_S - d_S. \tag{2.2.5}$$

We assume that the rate of cell death of S cells, d_S, is proportional to the birth rate, b_S, so that, $d_S = b_S/\sigma$, where $\sigma \geq 1$ is a dimensionless parameter. Thus, the net growth rate of the tumour is determined by the value of σ such that, as $\sigma \to 1$, $d_S \to b_S$ and the net growth rate in equation (2.2.5) decreases. The assumption that birth and death are linked is motivated by studies in which the rates of cell cycle progression and programmed cell death have been shown to be coupled by shared genes such as the cyclin D1 gene, a regulator of the G1 phase of the cell cycle and shared factors such as the Cdk kinase, which activates pro-mitotic and apoptotic pathways [74, 113].

We assume that the birth rate, b_S, is population-dependent, whereby the birth rate depends on the proportions of the different cell types in the tumour. We provide more details of population-dependent birth rates in section 2.2.2 of this chapter. By assuming the birth rate is population dependent, the death and net growth rates are also population dependent.

For R_1, R_2 and R_{12} cells, birth rates, denoted by b_{R_1}, b_{R_2} and $b_{R_{12}}$, respectively, death rates, denoted by d_{R_1}, d_{R_2} and $d_{R_{12}}$, respectively, and net growth rates, denoted by r_{R_1}, r_{R_2} and $r_{R_{12}}$, respectively, follow in exactly the same way.

Mutations are assumed to take place such that a fixed proportion of cells mutate per cell division and we assume mutations are intrinsic whereby no external factors influence the propensity for a cell to mutate, for example, the tumour microenvironment. The dimensionless parameter μ denotes the proportion of cells that mutate per cell division from (S to R_1), (R_2 to R_{12}), (S to R_2) and (R_1 to R_{12}). Thus, an S cell divides, with mutation, to produce an S cell and an R_1 (or R_2) cell at rate, $b_S\mu$ (or rate $b_S\mu$). Similarly, an R_1 cell divides, with mutation, to produce an R_1 cell and an R_{12} cell at rate, $b_{R_1}\mu$ and an R_2 cell divides, with mutation, to produce an R_2 and R_{12} cell at rate, $b_{R_2}\mu$. We assume that resistance is not reversible.

Cells are assumed to compete with each other for resources or space. The carrying capacity, K (in units of number of cells), describes the maximum number of tumour cells that can be supported by the available space or resources. In the absence of other cell types, we assume that a cell population undergoes logistic growth, to the carrying capacity; for simplicity, the carrying capacity is assumed to be the same for all cell populations.

2.2.2 Population-dependent birth rates

While we could assume that the birth and death rates in equations (2.2.1)-(2.2.4) are constants, in what follows, we consider a more general class of models in which birth rates, b_S, b_{R_1}, b_{R_2} and $b_{R_{12}}$ depend on the tumour composition. This assumption is motivated by a recent game-theory study presented by Kaznatcheev et al. in which in vitro experiments were performed on NSCLC parental cell lines and cell lines resistant to alectinib, a TKI [72]. They demonstrated that the growth rate of a cell type depends on the abundance of other cell types in the tumour population. They used payoff matrices to present their findings, to illustrate the types of cell-cell interactions that can arise, and, the effects that these interactions can have on cell growth rates. For example, beneficial interactions increase the growth rate or detrimental interactions decrease the growth rate. We generalise this approach in our model.

We first present a simple example taken from [72] whereby two cell lines are considered: parental cells, denoted by P, and resistant cells, denoted by R. Let b_P and b_R denote the birth rates of parental and resistant cells, respectively, and, let $p_P = P/(R+P)$

and $p_R = R/(P+R)$ denote the proportions of parental and resistant cells at time, t, respectively. In order to define the functional forms for b_P and b_R, we first introduce the payoff matrix, $\mathbf{M}$ where parameters, α, β, γ and δ are positive constants,

$$\mathbf{M} = \begin{array}{c|cc} & \text{Parental (P)} & \text{Resistant (R)} \\ \hline \text{Parental (P)} & \alpha & \alpha\beta \\ \text{Resistant(R)} & \alpha\delta\gamma & \alpha\delta \end{array}. \tag{2.2.6}$$

In $\mathbf{M}$, the parameter α denotes the intrinsic rate of cell division of parental cells (units day^{-1}) in a monotypic population of parental cells. The dimensionless parameter, β denotes the relative increase ($\beta > 1$), or decrease ($\beta < 1$), in the rate of cell division of a parental cell when it interacts with resistant cells. We suppose that, as the tumour population becomes more resistant (i.e. $p_P \to 0$, $p_R \to 1$), the cell division rate of a parental cell will approach $\alpha\beta$ (units day^{-1}). The dimensionless parameter, δ denotes the extent to which the division rate of resistant cells, in a monotypic population (of resistant cells), is greater than ($\delta > 1$), or lower than ($\delta < 1$), the division rate of the parental cells. Hence, $\alpha\delta$ (units day^{-1}) is the intrinsic rate of cell division of a resistant cell in a monotypic population of resistant cells. The dimensionless parameter, γ describes the relative increase ($\gamma > 1$), or decrease ($\gamma < 1$), in the cell division rate of resistant cells when they interact with parental cells, compared to their division rate when there are no parental cells ($\alpha\delta$). Therefore, as the proportion of parental cells increases (i.e. $p_P \to 1$, $p_R \to 0$), the cell division rate of a resistant cell approaches $\alpha\delta\gamma$ (units day^{-1}).

We assume further that the cell division rates, b_P and b_R of parental and resistant cells are linearly dependent on the proportion of the cells of a particular type at a given time. Thus, we write:

$$b_P = \alpha\left(p_P + \beta p_R\right), \tag{2.2.7}$$

$$b_R = \alpha\delta\left(\gamma p_P + p_R\right). \tag{2.2.8}$$

We note that, if we were to fix parameters, $\beta = 1$ and $\gamma = 1$ in our model, interactions between different cell types would not increase or decrease the birth rates, and, therefore, the birth rates would have to be fixed with values $b_P = \alpha$ and $b_R = \alpha\delta$.

We extend the example presented above to describe interactions between the four cell types in our model. We view S cells in our model as the parental cells (P), and, we assume they have the payoff matrix entries, α and $\alpha\beta$ in (2.2.6), associated to them. We assume further that R_1, R_2 and R_{12} cells have the same payoff matrix entries, $\alpha\delta\gamma$ and $\alpha\delta$ as the resistant cells, R in (2.2.6). Under these assumptions, interactions between the four cell types can be summarised via the following payoff matrix, $\mathbf{A}$:

$$\mathbf{A} = \begin{array}{c|cccc} & S & R_1 & R_2 & R_{12} \\ \hline S & \alpha & \alpha\beta & \alpha\beta & \alpha\beta \\ R_1 & \alpha\delta\gamma & \alpha\delta & \alpha\delta & \alpha\delta \\ R_2 & \alpha\delta\gamma & \alpha\delta & \alpha\delta & \alpha\delta \\ R_{12} & \alpha\delta\gamma & \alpha\delta & \alpha\delta & \alpha\delta \end{array}, \tag{2.2.9}$$

and birth rates can be described by the following system of equations:

$$\begin{bmatrix} b_S \\ b_{R_1} \\ b_{R_2} \\ b_{R_{12}} \end{bmatrix} = \mathbf{A} \begin{bmatrix} p_S \\ p_{R_1} \\ p_{R_2} \\ R_{12} \end{bmatrix}, \tag{2.2.10}$$

where p_S denotes the proportion of S cells in the tumour

$$p_S = \frac{S(t)}{S(t) + R_1(t) + R_2(t) + R_{12}(t)}, \tag{2.2.11}$$

with similar expressions for and p_{R_1}, p_{R_2} and $p_{R_{12}}$. Thus, the birth and death rates of the sensitive cells (S) are,

$$b_S = \alpha[p_S + \beta(p_{R_1} + p_{R_2} + p_{R_{12}})], \tag{2.2.12}$$

$$d_S = \frac{\alpha}{\sigma}[p_S + \beta(p_{R_1} + p_{R_2} + p_{R_{12}})]. \tag{2.2.13}$$

The rates of birth and death for the other cell populations take the form:

$$b_{R_1} = b_{R_2} = b_{R_{12}} = \alpha\delta[\gamma p_S + (p_{R_1} + p_{R_2} + p_{R_{12}})], \tag{2.2.14}$$

$$d_{R_1} = d_{R_2} = d_{R_{12}} = \frac{b_{R_1}}{\sigma}. \tag{2.2.15}$$

2.2.3 Initial conditions

We assume that initially the tumour comprises 100 sensitive cells, with no R_1, R_2 or R_{12} cells, so that,

$$S(0) = 100, \qquad R_1(0) = R_2(0) = R_{12}(0) = 0. \tag{2.2.16}$$

To summarise, our model comprises a system of ODEs defined by equations (2.2.1)-(2.2.4) where the birth and death rates are defined by equations (2.2.12)-(2.2.15) and the initial conditions prescribed in equation (2.2.16).

2.2.4 Parameters

A summary of model parameters and estimates of their values is given in Table 2.2.1. Previous estimates of NSCLC growth rates in monotypic populations range between

Param.	Description	Default value	Units	Uniform pdf ranges	Ref.
μ	Proportion of cells that mutate per cell division	10^{-6}	dim.	$[10^{-9}, 10^{-4}]$	[24, 48, 99]
σ	Birth to death rate ratio, $\frac{b_X}{d_X}$ for $X \in \{S, R_1, R_2, R_{12}\}$	1.05	dim.	[1.02, 1.1]	[73, 131]
K	Carrying capacity	5×10^8	cells	$[1.3\times10^8, 1.3\times 10^{12}]$	[34]
α	Intrinsic rate of cell division of S cells	0.156	day^{-1}	[0.11, 0.27]	[72]
β	Effect of R_1, R_2 or R_{12} cells on the growth rate of S cells	1	dim.	[0.5, 1.5]	[72]
γ	Effect of S cells on the growth rates of R_1, R_2 and R_{12} cells	1	dim.	[0.5, 1.5]	[72]
δ	Relative intrinsic rate of cell division of resistant cells to S cells	1	dim.	[0.5, 1.5]	[72]

Table 2.2.1: Default model parameter values, descriptions, units and their respective uniform distribution ranges used in eFAST analysis. We include references for their source.

0.1 and 0.24 (day^{-1}) [72]. Based on [72], we suppose $\alpha \in [0.11, 0.27]$ (day^{-1}) and fix $\alpha = 0.156$ (day^{-1}) as the default value.

We consider values of (dimensionless) parameters β, γ and δ in [0.5, 1, 1.5] to investigate the effects of the division rates of resistant cells that are higher ($\delta > 1$), lower

($\delta < 1$), and equal ($\delta = 1$) to those of the sensitive cells, and, the beneficial ($\beta > 1$ and/or $\gamma > 1$) and detrimental effects ($\beta < 1$ and/or $\gamma < 1$) of cellular interactions.

A wide range of probabilities of mutations is reported in the literature. Estimated orders of $O(10^{-7}-10^{-4})$ for the probability of mutation per cell cycle have been presented in the experimental literature [48]. Using mutation probabilities in the range of (10^{-8} to 10^{-4}), Chmielecki et al. estimated the number of NSCLC resistant cells to gefitinib and erlotinib in a population of 3 million cells [24]. In the literature probabilities of mutations are reported as mutations are assumed to be random, spontaneous events that may occur in a given cell division. In this chapter, we employ deterministic models and therefore describe μ as the proportion of cells that mutate per cell division where we consider values of μ in the range 10^{-9} to 10^{-4}. Following Monro and Gaffney, we employ a default value of $\mu = 10^{-6}$ [99]. In chapter 3, we consider stochasticity whereby mutations are random events that can occur with a fixed probability per cell division. Hence, in chapter 3, we describe μ as a fixed probability of mutation per cell division.

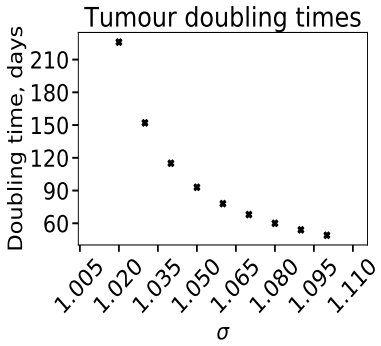


Figure 2.2.2: Tumour doubling time (days) for a range of σ values. For a given value of σ, we simulate equations (2.2.1)-(2.2.4) where the birth and death rates are defined by equations (2.2.12)-(2.2.15). The doubling time is the time taken for the total tumour population to increase from (0.5×10^8) to 10^8 cells. Parameters: all parameters are fixed at the default values in Table 2.2.1, except for the carrying capacity, $K = 1.3 \times 10^{12}$ and σ, which is varied between 1.02 and 1.1.

In our model, σ describes the birth rate divided by the death rate. We employ a range of values of σ in $[1.02, 1.1]$. We motivate our choice of σ values as follows. The *cell loss factor* defines the proportion of cells undergoing mitosis to cells dying in a tumour population. A higher cell loss factor corresponds to a greater number of cells undergoing cell death [131]. Cell loss factors have been recorded to vary between 0.79 and 0.95

across adenocarcinoma pulmonary tumours [73]. By considering the range $[1.02, 1.1]$, we assume cell death rates range between 0.91 to 0.98 of cell division rates.

The volume doubling times (VDT) of NSCLC tumours with the exon 19 EGFR (epidermal growth factor receptor) mutation, range from approximately 91 to > 1000 days [156]. Nakamura et al. estimated a median value of 188 days across 111 NSCLC patients while Zhang et al. estimated a median value of 214 days across 88 patients [102, 156]. To demonstrate the suitability of our choice of values of σ, such that tumours grow with doubling times which match the times presented in the literature, we simulate our model for a range of values of σ. For a given σ value, the tumour doubling time is the time taken for the total tumour population to double from 0.5×10^8 cells to 10^8 cells. In Figure 2.2.2, we show how the doubling times change as σ increases. These doubling times are in good agreement with the above experimental estimates. We employ a default value of $\sigma = 1.05$ which corresponds to a death rate that is approximately 95% of the birth rate, and, a tumour doubling time of 93 days via our model with the default parameters given in Table 2.2.1.

We estimate the carrying capacity as follows. A detectable tumour is 1cm^3 (1g in weight) and contains approximately 10^8 cells [34]. We consider a lower bound of $K = 1.3 \times 10^8$ cells. In a study presented by Norton, he fitted a Gompertz model to breast cancer tumour growth data, presented by Bloom et al. [10, 104]. In so doing, he estimated a lethal tumour population number of 10^{12} cells. Therefore, we consider an upper bound of 1.3×10^{12} cells which corresponds to a volume of close to $1000cm^3$ [134]. We employ the default value of $K = 5 \times 10^8$ cells.

2.3 Methods

We implement our model in Python 3.7.6. We integrate the governing ODEs defined by equations (2.2.1)-(2.2.4) using the solve_ivp function from the Python library, SciPy 1.4.1. This function numerically integrates a system of ODEs with defined initial conditions. We use the Python libraries Matplotib 3.1.3 and Seaborn 0.11.1 to produce plots and combine them using Inkscape.

2.3.1 Global sensitivity analysis

In order to understand the impact of varying model parameters on the level of resistance in an untreated tumour, we perform variance-based global sensitivity analysis using the extended Fourier amplitude sensitivity test (eFAST). The eFAST algorithm was developed by Saltelli et al. and is based on the FAST method produced by Cukier et al. [31, 120]. eFAST quantifies the variation in model output and determines what fraction of the output variance is caused by variation in each input parameter. When performing eFAST on our model, for each model parameter, it yields two sensitivity index values; S_i, the first-order sensitivity index, represents the proportion of the output variance due to variation in a single model parameter; S_{T_i}, the total-order sensitivity index, takes into account the terms in the model in which model parameters interact, and, defines the proportion of the output variance when combinations of model parameters are considered. Hence, S_{T_i} contains all the contributions at first and higher orders of a given parameter to the model output of interest. eFAST differs from its prequel, FAST, in that a total-order sensitivity index is given.

When using eFAST on our model, we consider two outputs of interest; the first is the proportion of fully drug-resistant cells, $p_{R_{12}}$, at diagnosis, when the total tumour population reaches 10^8 cells [34]; the second is the proportion of fully drug-resistant cells in the tumour over time. For the latter, we calculate the first and total-order sensitivity indices of the model parameters at fixed time points, for a range of selected time points within the period of 0 to 15 years.

We refer the reader to the appendix for an in-depth description of eFAST. We also include results of convergence analysis that we performed using methods proposed by Marino et al. [90].

2.4 Results of untreated tumour growth

In this section, we present typical solutions of the untreated tumour growth model. We also present results of our global parameter sensitivity analysis. We show how the proportion of fully drug-resistant cells, at diagnosis and over time, changes as we vary parameters, identified by eFAST, as influential. We then investigate the role of

different types of cellular interactions on the cellular composition of a tumour over time by simulating our model for varying values of parameters, β, γ.

2.4.1 Typical model simulations

Figure 2.4.1 presents typical simulations of the untreated tumour growth model. In Figure 2.4.1(a), the default parameter values are employed (see Table 2.2.1). In Figures 2.4.1(b)-(f), we vary one model parameter from the set [β, δ, γ, μ, σ]. We plot the S, R_1 and R_{12} cell populations only and neglect solutions of the R_2 populations. This is because our model is symmetric in that we assume the R_1 and R_2 populations have the same growth, death and mutation rates and thus, are equal, since the initial conditions are also the same. Relaxing this constraint would be an interesting direction for future work but is not considered in this thesis.

In Figure 2.4.1(a), S cells grow to carrying capacity ($K = 5 \times 10^8$ cells). As the S population increases, mutational flux into the R_1 compartment takes place and we observe an increase in this population at approximately $t = 1$ years. At later times ($t = 4$ years), the R_{12} population increases due to mutations into this compartment.

In Figure 2.4.1(b), we do not observe markedly different dynamics as β varies. This is because the R_1 population is approximately 3 orders of magnitude smaller than the S cells. For example, at time $t = 10$ years, the S population is made up of approximately 5×10^8 cells and the R_1 population consists of approximately of 5×10^5 cells. Consequently, the effect of the resistant cells on the growth rates of S cells is low.

Figures 2.4.1(c)-(e) show that, as values of δ, γ and μ increase, the resistant cell populations (R_1 and R_{12}) increase as expected as these cells have higher growth rates and there is a larger mutational flux into these compartments.

Slower growth of all cell populations is observed for decreasing values of σ in Figure 2.4.1(f), as lower values of σ correspond to higher cell death rates.

2.4.2 Sensitivity analysis

In order to determine those model parameters that have the greatest influence on a tumour's evolution to resistance, we perform global sensitivity analysis on the untreated model using eFAST (see Appendix for details on eFAST). The outputs of interest are

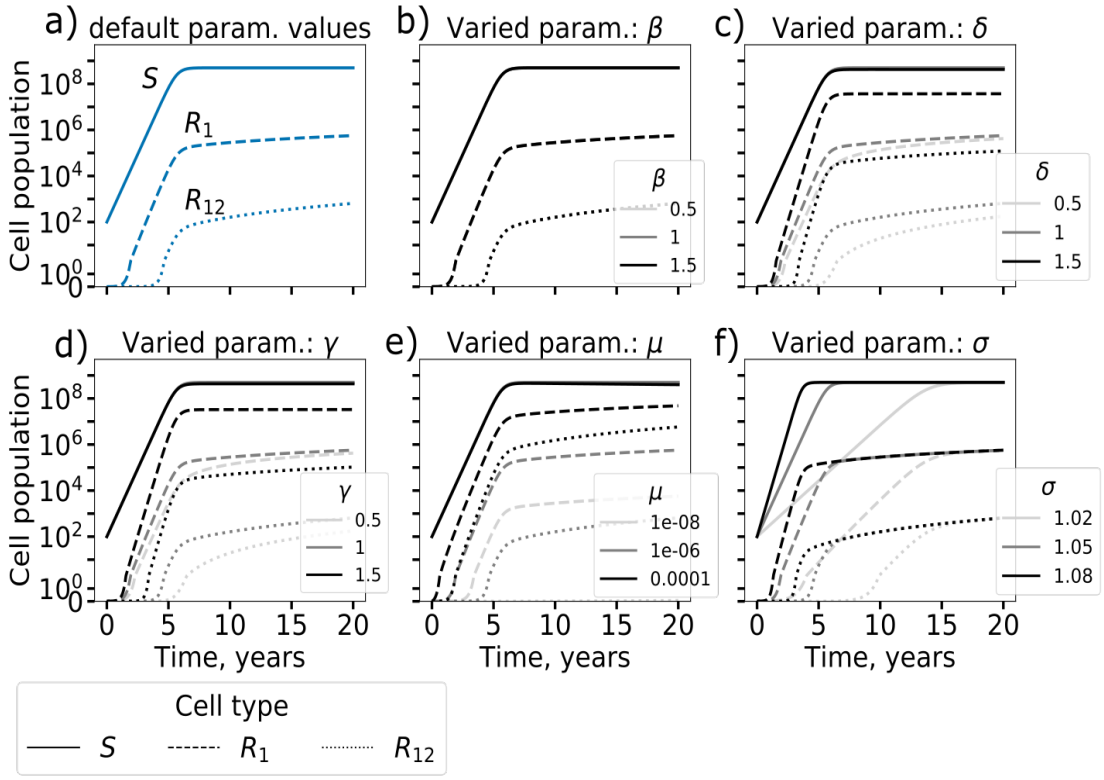


Figure 2.4.1: Typical model solutions showing how cell numbers change over time for an untreated tumour. Key: line styles represent cell types; $S(t)$= block lines; $R_1(t)$=dashed lines; $R_{12}(t)$=dotted lines. Colours represent parameter values used where black, dark grey and light grey represent high, mid and low values of the varied parameter, respectively. Subplots: (a) Model solutions with default model parameters given in Table 2.2.1. (b) Model solutions with β values in [0.5, 1, 1.5] where light grey lines represent $\beta = 0.5$; dark grey; $\beta = 1$ and black; $\beta = 1.5$. The same colour scheme holds for the following plots and their respective parameter values; (c) δ values in [0.5, 1, 1.5]; (d) γ values in [0.5, 1, 1.5]; (e) μ values in $[10^{-8}, 10^{-6}, 10^{-4}]$; (f) σ values in [1.02, 1.05, 1.08]. In a given subplot, for non-varied parameters, default parameter values given in Table 2.2.1 are used.

$p_{R_{12}}$, the proportion of the population that is fully drug-resistant (see equation 2.2.11) at, (i) diagnosis (i.e. when the total tumour population reaches 10^8 cells) and, (ii) over time. We present results in Figure 2.4.2. Figures 2.4.2(a) and (b) present the total and first-order sensitivity indices for the model parameters, over time, respectively and show the parameters which have the greatest influence on $p_{R_{12}}(t)$ values over time. Figure 2.4.2(c) presents the first and total-order sensitivity indices for the model parameters, including the dummy parameter, showing the extent to which the parameters influence the values of $p_{R_{12}}(t)$ at diagnosis, when the tumour reaches a total tumour population

of 10^8 cells. In Figure 2.4.2(c), for a given model parameter, the height of the black bar indicates the value of the first-order sensitivity index and the height of the grey bar indicates the value of the total-order sensitivity index. The error bars represent two standard deviations about the mean of the sensitivity index (10 indices in total for a given model parameter as 10 resamples are employed when simulating eFAST which is discussed in the appendix of this chapter). Parameters with sensitivity indices that are significantly larger than that of the dummy parameter (p-value < 0.01) are signified with an asterisk.

The results presented in Figure 2.4.2 reveal that changing the parameter μ, denoting the proportion of cells that mutate per cell division, has the largest influence on the proportion of fully drug-resistant cells over time (Figure 2.4.2(a)-(b)) and at diagnosis (Figure 2.4.2(c)). Figure 2.4.2(c) shows further that, at diagnosis, δ, the relative growth rate of resistant cells to sensitive cells, is the second most influential parameter followed by, γ, the effect of the sensitive cells on the growth of the resistant cells, and then σ, the birth to death rate ratio (see Figure 2.2.2 which shows that smaller values of σ correspond to slower rates of tumour growth).

Figure 2.4.2(a) shows that, during the initial stages of tumour growth (for $0 < t <$ 3 years) the growth rate of the sensitive cells, α, and the birth to death rate ratio, σ, have large total order sensitivity indices. During these early times, the sensitive cell population increases, increasing the flux of the mutant cells into the R_1 and R_2 compartments and, hence, into the fully drug-resistant cell compartment. Therefore, the rate, α, at which the sensitive population grows, has a significant effect on the size of the fully drug-resistant population during these early stages of tumour growth. For $t > 6$ years, lower sensitivity to α is observed and the birth to death rate ratio, σ is no longer significant (black line is not plotted for $t > 6$ years). We attribute this change to the fact that, at long times, there is a larger R_{12} population whose growth is dominated by the values of parameters, δ and γ which are shown to be the second and third most influential parameters at these times, respectively. Sensitivity of the R_{12} population over time to the parameter, μ remains high and constant for all time which is unsurprising as this parameter determines the rate at which cells become doubly resistant. We note

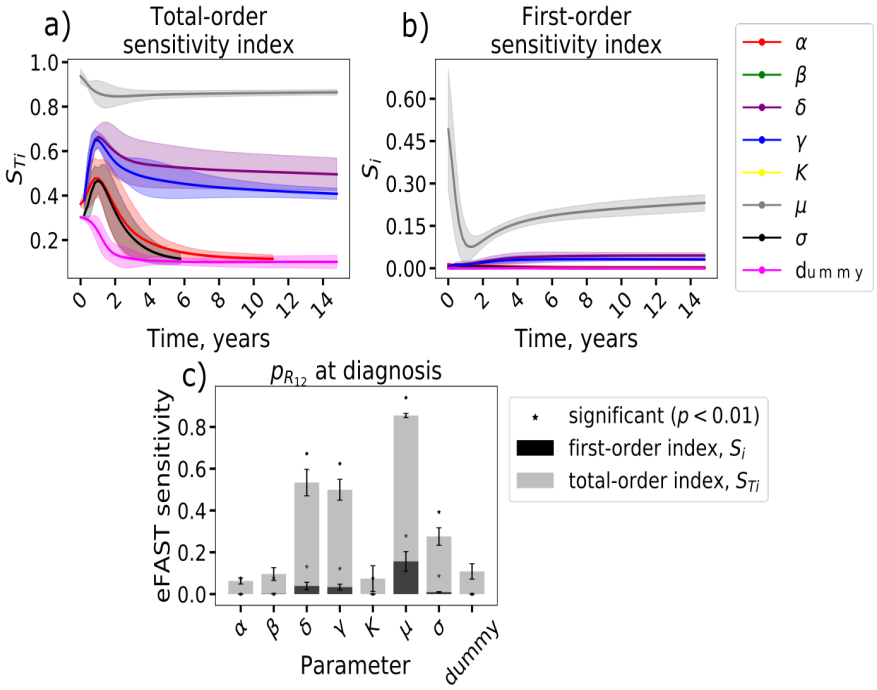


Figure 2.4.2: Results from global sensitivity analysis, using eFAST, applied to the untreated tumour growth model. We implement eFAST with the number of samples, $N_S = 5000$ and the number of resamples, $N_R = 10$ (see appendix for details of eFAST). (a) Total-order sensitivity index, S_{Ti} for model parameters, over time. (b) First-order sensitivity index, S_i for model parameters, over time. In plots (a) and (b), lines represent mean sensitivity index of the sample ($N_R = 10$ samples in total) and the shaded regions represent two standard deviations about the mean sensitivity index for a given parameter. The legend corresponding to plots (a) and (b) is given to the right of subplot (b). In plots (a) and (b), when a parameter is no longer significant, the line corresponding to its sensitivity index is terminated. (c) First and total-order sensitivity indices for model parameters at diagnosis (10^8 cells). An asterisk, * signifies significance with a p-value < 0.01 compared to the distributions of first and total-order sensitivity indices corresponding to the dummy parameter (distributions are made up of $N_R = 10$ values in total). Error bars represent two standard deviations about the mean of the sensitivity indices. Model parameters, summarised in Table 2.2.1, are sampled from a uniform distribution with ranges also given in Table 2.2.1.

that parameters K and β are not significant and sensitivity indices associated to these parameters are therefore not plotted in Figures 2.4.2(a) and (b). The same can be said for Figure 2.4.2(c), demonstrating that the varying values of K and β do not influence the proportion of R_{12} cells at diagnosis.

The global sensitivity analysis has revealed the parameter μ, denoting the proportion of cells that mutate per cell division, is the most influential parameter that drives drug resistance, followed by the growth rate of resistant cells relative to sensitive cells, δ, the effect of S cells on the growth rates of resistant cells, γ, and then the birth to death rate ratio, σ. Next, we use eFAST results to predict the proportion of fully drug-resistant cells at diagnosis as these key model parameters are varied.

2.4.3 Quantifying pre-existing resistance

We use our model to investigate how $p_{R_{12}}$ changes over time and at diagnosis as key model parameters values are varied. In this way, we aim to quantify the extent to which these parameters drive patient heterogeneity at diagnosis. The results are summarised in Figure 2.4.3. Diagonal entries of Figure 2.4.3 show changes in $p_{R_{12}}$ over a 10-year period when a single parameter (μ, δ, γ and σ) is varied within their bounds given in Table 2.2.1 (all other model parameter values are fixed at their default values, also given in Table 2.2.1).

The heatmaps in Figure 2.4.3 show how the value of $p_{R_{12}}$ at diagnosis changes when two parameters are varied. Blank spaces represent cases where there are effectively no R_{12} cells (i.e. $p_{R_{12}} < 10^{-8}$).

We observe that changes in the parameter μ, denoting the proportion of cells that mutate per cell division, produce the largest variation in $p_{R_{12}}$ values, with low values of μ ($\mu = 10^{-4}$) producing tumours with a large proportion of R_{12} cells at $t = 10$ years. These results are consistent with our global sensitivity analysis, where μ had the largest first-order sensitivity index (see in Figure 2.4.2). Similarly, variability in the cellular dynamics as a single parameter is varied decreases through the hierarchy of influential parameters (μ being the most influential and σ being the least influential). For resistant cells whose growth rates are higher than sensitive cells ($\delta > 1$), $p_{R_{12}} > 0$ whereas, for lower growth rates of these cells ($\delta < 1$), $p_{R_{12}} = 0$. Similarly, larger values

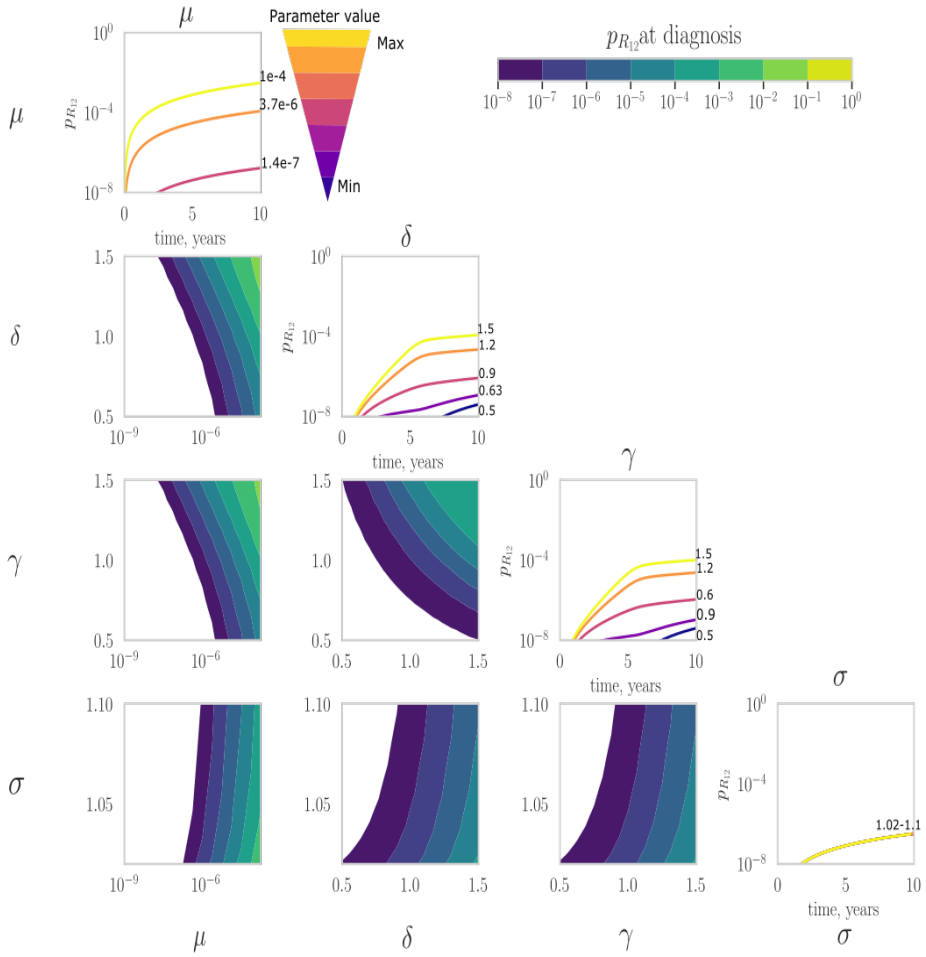


Figure 2.4.3: Predicted proportion of fully drug-resistant cells, $p_{R_{12}}$ at diagnosis (heatmap plots) and over time (lineplots): Heatmap plots: tumours are grown until the total tumour population reaches 10^8 cells (diagnosis) at which time, $p_{R_{12}}$ is measured. Dark blue and yellow heatmap colours represent low and high values of $p_{R_{12}}$, respectively, in which two key parameters from μ, δ, γ and σ, are varied simultaneously in a pairwise manner. White represents cell proportions less than 10^{-8}. The horizontal colour bar presents the range of $p_{R_{12}}$ values $[10^{-8}, 1]$. Line plots: describe $p_{R_{12}}$ over time in which one key parameter is varied in a single plot. The ranges of the parameters, μ, δ, γ and σ are given in Table 2.2.1. Yellow and blue lines in the line plots correspond to the maximum and minimum parameter values in the predefined range (see Table 2.2.1), respectively. When parameters are not varied, they take the default values given in Table 2.2.1.

of $p_{R_{12}}$ are observed when there is a mutualistic relationship between the S and the R_{12} cells ($\gamma > 1$) rather than a competitive relationship ($\gamma < 1$). The variability in p_{R12} is markedly reduced when σ varies in the range $[1.02, 1.1]$ compared to other influential

parameters.

The heatmaps in Figure 2.4.3 show that $p_{R_{12}}$ at diagnosis increases by 7 orders of magnitude as the proportion of cells that mutate per cell division, μ and the growth rates of resistant cells increase. For $\mu = 10^{-4}$ and $\delta = 1.5$, the model predicts that the tumour contains between 1% and 10% fully drug-resistant cells. Similar results are observed when the the proportion of cells that mutate per cell division and the positive effect of S cells on the growth rate of the R_{12} cells (γ) increase. We observe higher proportions of fully drug-resistant cells for lower σ values (corresponding to higher death rates). We attribute this to longer times to detection for lower σ values allowing for more time for mutations to occur.

In summary, we have used the model to study the impact of varying key parameters which affect the proportion of the tumour that is resistant. We have shown how the level of resistance in a tumour at diagnosis changes as the characteristics of the tumour cells vary, such as their propensity to mutate. The reported range of mutation probabilities and growth rates of resistant cells in the literature (the most influential parameters in our study), indicate that variation in these parameters could be the main factor affecting resistance levels at diagnosis [24, 48, 72, 99]. Furthermore, the significance of γ suggests that further experiments are needed to assess the effects of cell-cell interactions on tumour cell dynamics, as in the study by Kaznatcheev et al. [72].

Next, we use our model to investigate how different types of cellular interactions influence the way in which tumour composition evolves over time.

2.4.4 Effect of cellular interactions on tumour composition over time

The eFAST results presented in section 2.4.2, revealed γ to be the third most influential parameter on $p_{R_{12}}$ values in an untreated tumour, at diagnosis, and over time. These results show that the effect of interactions between S cells and resistant cells are significant. Here, we investigate how the proportions of different cell types in a tumour change over time as the cellular interaction parameters, β and γ vary.

We simulate our model for 10 years and show how cell proportions change over time. We present results in Figure 2.4.4. We consider resistant cells with growth rates that are higher ($\delta > 1$), lower ($\delta < 1$), and equal ($\delta = 1$) to the sensitive cells. We do not

plot proportions of R_2 cells over time as the dynamics of these cells are the same as for the R_1 cells. We employ a large value for K ($K = 1.3 \times 10^{12}$ cells) so that the effects of competition are negligible. We also employ a high value of $\mu = 10^{-4}$, so that resistant populations reach sizes for which their interactions with the sensitive cells are not negligible as a larger proportion of cells are mutating per cell division.

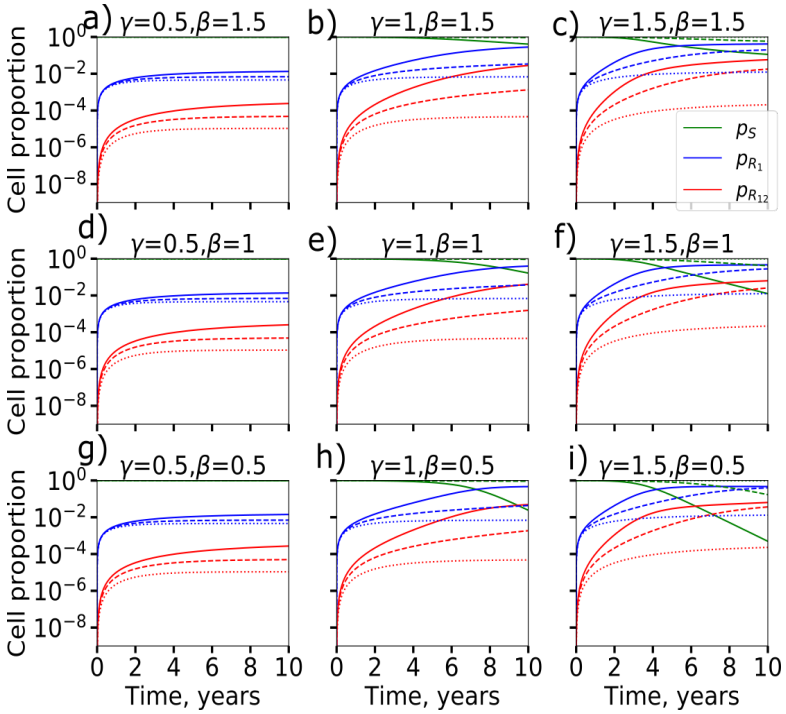


Figure 2.4.4: Series of plots showing how the proportion of tumour cells changes over time as the parameters, β, γ and δ vary. Key: colours represent the proportions of cell types: p_S=green; p_{R_1}=blue; $p_{R_{12}}$=red. In each subplot we consider three values of δ: dotted lines = $\delta = 0.5$; dashed lines = $\delta = 1$; block lines = $\delta = 1.5$. Columns correspond to different values of $\gamma \in [0.5, 1, 1.5]$. Rows correspond to different values of $\beta \in [0.5, 1, 1.5]$. All tumours are grown for a period of 10 years. The legend is given in the top right figure. Parameters: $\alpha = 0.11(\text{day}^{-1})$, $\mu = 10^{-4}$, $K = 1.3 \times 10^{12}$ cells and $\sigma = 1.03$.

The results presented in Figure 2.4.4 show that, for all cases, the proportion of R_1 and R_{12} cells increase as δ increases. This is because the resistant cells have increased growth rates for increasing values of δ. Similarly, $p_{R_{12}}$ increases as γ increases (left to right columns correspond to low to high values of γ, respectively). This is because the benefit to resistant cells from sensitive cells increases for increasing values of γ. These results are consistent with our eFAST analysis, which found γ and δ to be the most influential parameters after μ. For $\gamma = 1.5$ and $\beta = 0.5$ in Figure 2.4.4(i), when resistant

cells have a higher growth rate relative to the sensitive cells ($\delta = 1.5$, block lines), these cells become the dominant tumour population. When $\gamma \geq 1$, for decreasing values of β, which describes the case where resistant cells have an increasingly negative effect on the growth rate of the sensitive cells, larger resistant proportions are observed and smaller proportions of S cells are observed.

These results show how different types of cellular interactions affect the tumour dynamics and highlight the importance of accounting for cellular interactions when modelling cell growth rates in vitro. When we use findings in [72] in our model, the tumour contains large proportions of fully drug-resistant cells ($p_{R_{12}} > 10^{-2}$) at 10 years (see Figure 2.4.4(c)). This could have a significant effect on treatment outcomes. We note that the properties of cell lines may change when they interact and they may be patient specific. While we have focussed on two cell lines here, the model could also be used to investigate the interactions with different cell lines and tumour types.

2.5 Treated tumour growth model

2.5.1 Treatment model development

In this section, we extend the model defined by equations (2.2.1)-(2.2.4) to account for treatment with two TKIs. We continue to assume population-dependent birth, death and net growth rates (see section 2.2.2 for details of population-dependent rates). Specific combination treatment protocols of interest are introduced after we develop the treatment model.

We use the model to investigate the effect of varying key model parameters (μ, δ, γ and σ) on treatment outcomes and tumour composition during treatment. We remind the reader of the parameter μ that denotes the proportion of cells that mutate per cell division. The parameter δ denotes the growth rate of resistant cells relative to sensitive cells. The parameter γ denotes the effect of the sensitive cells on the growth rate of the resistant cells and the parameter σ denotes the birth to death rate ratio.

In what follows, we present the system of equations that describes the tumour cell dynamics when treatment is applied. Following on from this, we provide an explanation of the treatment model terms.

Governing equations with treatment

The dimensional equations describing tumour growth and the response to treatment are as follows:

$$\frac{dS}{dt} = r_S S\Big(1-\frac{S+R_1+R_2+R_{12}}{K}\Big) - \overbrace{\mu b_S S}^{\text{mutations to } R_1} - \underbrace{\mu b_S S}_{\text{mutations to } R_2} - \overbrace{\frac{b_S(\theta_1+\theta_2)}{\sigma}S}^{\text{death by drugs one and two}}, \tag{2.5.1}$$

$$\frac{dR_1}{dt} = r_{R_1} R_1\Big(1-\frac{S+R_1+R_2+R_{12}}{K}\Big) + \mu b_S S - \underbrace{\mu b_{R_1} R_1}_{\text{mutations to } R_{12}} - \overbrace{\frac{b_{R_1}\theta_2}{\sigma}R_1}^{\text{death due to drug two}}, \tag{2.5.2}$$

$$\frac{dR_2}{dt} = r_{R_2} R_2\Big(1-\frac{S+R_1+R_2+R_{12}}{K}\Big) + \mu b_S S - \underbrace{\mu b_{R_2} R_2}_{\text{mutation to } R_{12}} - \overbrace{\frac{b_{R_2}\theta_1}{\sigma}R_2}^{\text{death due to drug one}}, \tag{2.5.3}$$

$$\frac{dR_{12}}{dt} = r_{R_{12}} R_{12}\Big(1-\frac{S+R_1+R_2+R_{12}}{K}\Big) + \mu b_{R_1} R_1 + \mu b_{R_2} R_2. \tag{2.5.4}$$

Before we introduce the terms in the treatment model defined by equations (2.5.1)-(2.5.4) that represent the effects of treatment, we recap on the terms we have previously introduced in the untreated model (equations (2.2.1)-(2.2.4)) which are present in the treatment model.

In equation (2.5.1), the parameter b_S denotes the population-dependent birth rate of the S cells such that

$$b_S = \alpha[p_S + \beta(p_{R_1} + p_{R_2} + p_{R_{12}})], \tag{2.5.5}$$

where p_S, p_{R_1}, p_{R_2} and $p_{R_{12}}$ are the proportions of the S, R_1, R_2 and R_{12} cells in the tumour at a given time and β denotes the effect of the resistant cells on the growth rate of the S cells (see section 2.2.2 for details on population-dependent birth and death rates). Similarly, in equations (2.5.2)-(2.5.4), b_{R_1}, b_{R_2} and $b_{R_{12}}$ denote the population-dependent birth rates for the R_1, R_2 and R_{12} cells, respectively. The rate of cell death, d_S is population-dependent as d_S is modelled as a proportion of the population-dependent birth rate such that $d_S = b_S/\sigma$ (similarly for d_{R_1}, d_{R_2} and $d_{R_{12}}$ and the R_1, R_2 and R_{12} cells, respectively). In equation (2.5.1), the parameter r_S is population-dependent

and denotes the net growth rate of the S cells, i.e. the birth rate minus the death rate whereby,

$$b_S - d_S = (1 - 1/\sigma)b_S$$

(similarly for r_{R_1}, r_{R_2}, $r_{R_{12}}$ for the R_1, R_2 and R_{12} cells, respectively).

We now introduce the terms in equations (2.5.1)-(2.5.3) that represent death of the drug sensitive cells due to drugs one and two. Let the constant, dimensionless parameters, θ_1 and θ_2 denote the effects of drugs one and two, respectively. We assume that drugs one and two act on drug sensitive cells by (linearly) increasing their rates of cell death (recall the cell death rate is modelled as a proportion of the birth rate, $d_X = b_X/\sigma$, for $X \in \{S, R_1, R_2, R_{12}\}$). For example, in equation (2.5.1), the rate, $b_S(\theta_1 + \theta_2)/\sigma$, describes the effect of drugs one and two on the fully drug sensitive cells, S. We motivate this by the fact that first and third-generation TKIs, gefitinib and osimertinib, have been shown to induce apoptosis in NSCLC cells [19, 20, 124].

We now summarise the death terms due to drugs one and two. In equation (2.5.1), $b_S(\theta_1 + \theta_2)/\sigma$ denotes the rate of death of S cells due to drugs one and two. Since R_1 cells are resistant to drug one, we assume that only drug two affects their death rate, giving rise to an additional cell death term of the form $b_{R_1}\theta_2/\sigma$ in equation (2.5.2). Similarly, since R_2 cells are resistant to drug two, only drug one affects their rate of cell death at rate $b_{R_2}\theta_1/\sigma$ (see equation (2.5.3)). Since R_{12} cells are fully drug-resistant, they are not affected by either drug. In the absence of drug ($\theta_1 = 0$, $\theta_2 = 0$), the treatment model defined in equations (2.5.1)-(2.5.4) is equivalent to the model defined in equations (2.2.1)-(2.2.4). For the remainder of this chapter we refer to the parameters θ_1 and θ_2 as the amount of drugs one and two, respectively.

We close the governing equations by importing the initial conditions prescribed in equation (2.2.16). From tumour initiation, tumours are grown untreated ($\theta_1 = 0$, $\theta_2 = 0$ in equations (2.5.1) - (2.5.4)) until the total tumour population reaches diagnosis (10^8 cells), at which time, treatment is introduced. By allowing the tumour to grow untreated prior to diagnosis, the tumour may consist of drug-resistant cells at diagnosis as a result of mutations taking place into the fully drug-resistant cell compartment. The growth of resistant cell lines in the absence of treatment has been demonstrated in a study

presented by Kaznatcheev et al. in which in vitro co-culture experiments of sensitive and resistant cell lines were carried out [72]. In this study, they found TKI resistant cell lines grew during the course of the experiment when interacting with sensitive cell lines. The resistant cells were shown to have a higher growth rate when grown in co-culture with sensitive cells as opposed to when the cells were grown in monoculture experiments. Interestingly, the growth of the resistant cells was accelerated when in drug. In future work, it would be of interest to investigate acquired resistance mechanisms which may lead to increased mutational flux to drug-resistant phenotypes when treatment is applied. Furthermore, upon the introduction of treatment, the rate of which resistant cells divide as well as the rate at which cells mutate may change. We do not consider said variations of the model in this thesis and these investigations are subject to future work. We do however investigate the effect of pre-existing resistance on the efficacy of first-line combination TKI treatment in chapter 3.

2.5.2 Treatment parameters

Yates et al. presented an experimental study in which, increasing the dose of the AKT inhibitor, AZD5363 caused a switch from cytostatic (slowing/halting of growth of sensitive cells) to cytotoxic effects (killing large numbers of cells) [150]. Motivated by this study ([150]), we model cytostatic levels of drug by employing small values of θ_1 and θ_2 and we model cytotoxic levels of drug by employing high values of θ_1 and θ_2 and we assume the same drug-effect mechanism in increasing the death rate for both types (cytotoxic and cytostatic) of drugs [150]. See Table 2.5.1 for values of θ_1 and θ_2 associated to cytostatic and cytotoxic levels of drugs.

2.5.3 Combination treatment protocols

We consider two treatment protocols. The goal of each treatment protocol is, (i) to maximise the time to progression (the tumour is said to progress when its size is 120% of the size at diagnosis) and, (ii) to minimise the size of the fully drug-resistant population at progression.

In practice, a diverse range of two-drug treatment protocols may be applied. We refer the reader to chapter 1 for a discussion of combination TKI treatments currently

being used in the clinic. Here, we focus on the protocols summarised in Table 2.5.1 and illustrated in Figure 2.5.1.

In Table 2.5.1, under Protocol A, the simultaneous application of two drugs at continuous doses describes the case in which the drug-effect parameters, θ_1 and θ_2 are held constant and fixed at values greater than zero from the start of treatment (diagnosis) until tumour progression. Hence, on a daily basis, the values of the parameters, θ_1 and θ_2 do not vary. Under Protocol B, the drug-effect parameters, θ_1 and θ_2 are held constant and fixed at values greater than zero over the course of a week and are then reduced to zero for a period of 3 weeks. This on-off cycle is repeated over the course of a year and we describe this protocol as one in which drugs are pulsed.

Under Protocol A, cytostatic drug levels are applied whereby low values of θ_1 and θ_2 are employed in the model (see Table 2.5.1 for cytostatic values of θ_1 and θ_2). Cytostatic levels of drugs result in the growth of the S population to halt since, lower values of θ_1 and θ_2 result in a death rate that is increased to a rate close to that of the cell birth rate. Conversely, under Protocol B, cytotoxic levels are used whereby high values of θ_1 and θ_2 are applied (see Table 2.5.1 for cytotoxic values of θ_1 and θ_2). Under Protocol B, cytotoxic drug levels result in a large decrease in the drug sensitive cell populations (S, R_1 and R_2) as the drugs increase the cell death rates to rates that are larger than the cell birth rates.

Before presenting typical simulation results, we pause to motivate our choice of protocols.

A - Simultaneous cytostatic TKI treatment. The protocol is motivated by a clinical trial in which first and third-generation TKIs, gefitinib and osimertinib are taken daily in combination [119]. In this trial, the maximum tolerated dose was 80mg osimertinib and 250mg gefitinib, daily [119]. During protocol A in our study, equal amounts of cytostatic levels of TKIs one and two are applied simultaneously (see Table 2.5.1). This halts the growth of S cells and slows the growth of R_1 and R_2 cells (see Figure 2.5.1a).

B - Simultaneous pulsed cytotoxic TKI treatment. Protocol B (see Table 2.5.1) corresponds to the simultaneous application of cytotoxic levels of TKIs one and two for a period of one week, followed by a three week break from treatment. Treatment then

Protocol	Description	Drug amount	Predicted response	Ref.
A	Both drugs applied simultaneously at moderate, continuous doses until progression occurs	Cytostatic levels, $\theta_1 = 0.02$ and $\theta_2 = 0.02$. The parameters θ_1 and θ_2 are held fixed at these values until tumour progression occurs.	*Cytostatic.* Growth of S cells halts. Growth of R_1 and R_2 cells is slowed (see Figure 2.5.1a).	[119]
B	Both drugs applied simultaneously and at continuous high doses for one week, followed by a three week drug-holiday, over the course of one year	Cytotoxic levels, $\theta_1 = 0.6$ and $\theta_2 = 0.6$. The parameters θ_1 and θ_2 are held fixed at these values for one week, followed by three weeks with $\theta_1 = 0$ and $\theta_2 = 0$. The cycle is repeated for the duration of a year.	*Cytotoxic.* Significantly reduces the drug-sensitive cell population sizes (large decrease in the S, R_1 and R_2 population sizes) (see Figure 2.5.1b).	[151]

Table 2.5.1: Summary of treatment protocols applied at diagnosis when the total tumour population reaches 10^8 cells. Protocol descriptions, amounts of drug applied in a given protocol, response, units and references are given.

resumes for another week, and so on, over the course of a year. This protocol results in a significant reduction in the proportion of drug sensitive cells (see Figure 2.5.1b). High-dose, pulsed exposure to TKIs has been applied in EGFR mutant lung cancer patients using erlotinib, a second-generation EGFR TKI treatment [151]. In this trial, erlotinib 1200mg days 1-2 was applied followed by erlotinib 50mg erlotinib days 3-7 weekly.

We use our treatment model defined by equations (2.5.1)-(2.5.4) to simulate tumour growth and apply a single protocol (A or B, see Table 2.5.1) when the tumour reaches diagnosis (10^8 cells). We record the time and proportion of R_{12} cells when this happens.

2.5.4 Typical simulations of treatment protocols

In Figures 2.5.1(a) and (b), we present cellular dynamics of tumours undergoing protocols A and B as first-line treatment, respectively. In Figure 2.5.1(a), protocol A halts the

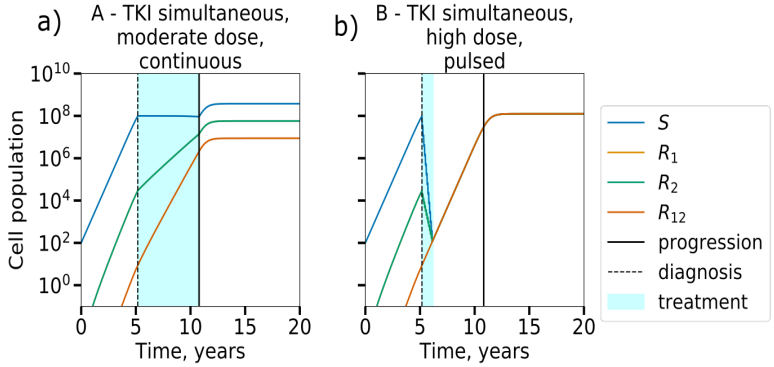


Figure 2.5.1: Typical simulations of tumour cell dynamics with first-line treatment applied at diagnosis. (a) Protocol A, (b) Protocol B. Key: S cells are presented by blue lines; R_1; green, R_2; yellow and R_{12}; orange. We note that R_1 and R_2 cells have the same dynamics, hence we observe only the solutions corresponding to R_1 cells (green). The cyan shaded regions indicate when treatment is applied. Black dashed lines represent the time of diagnosis, black solid line represents time of progression on or after (for protocol B only) first-line treatment. Parameters: the default parameter values given in Table 2.2.1 are used. Descriptions of protocols are given in Table 2.5.1.

growth of the S cells and slows the growth of the R_1 and R_2 cells. Conversely, under protocol B (see Figure 2.5.1(b)), there is a significant decrease in tumour size. For protocols A and B, we observe similar progression at approximately 5 years. Under protocol B, at progression, the size of the fully drug-resistant population (dark orange) is the same as that of the other cell types. Conversely, under protocol A, the fully drug-resistant cells are the smallest cell population at progression. We note that in Figure

2.5.1(a), treatment is applied for a period of up to 5 years which is unlikely to happen in a clinical setting. However, in the absence of suitable data with which to accurately parametrise the model, we perform simulations to assess its qualitative behaviour as different tumour-specific properties are varied. The work presented in this chapter paves the way for future investigations and may also be relevant for specific leukaemias where life long TKI treatment can be recommended with ten year overall survival in excess of 80% [60]. In particular, if time-course data were to become available, we could use our model to interpret it and, following validation of the model, predict responses to different treatment protocols.

2.6 Results of first-line treatment response

In this section, we use our model to predict treatment response to protocols presented in Table 2.5.1. We simulate treatment for a range of key parameter values (as identified using eFAST in subsection 2.4.2). In doing so, we investigate how different tumours respond to treatment with TKIs.

2.6.1 Effect of key parameter values on first-line treatment response

First, we use our model to investigate how key tumour-specific properties influence resistance and affect treatment outcomes (notably progression times). This analysis will help us understand what tumour characteristics drive variability in treatment responses in the clinic.

For a given set of key parameter values, we simulate two tumours and apply protocol A to one and protocol B to the other. We measure, (i) the time at which the tumour progresses, and, (ii) the proportion of fully drug resistant cells in the tumour at progression. In Figures 2.6.1 and 2.6.2, we present heatmap plots showing times, from the start of treatment, at which tumours progress, and, $p_{R_{12}}$ values at progression, respectively, as pairs of model parameters (μ, δ, γ and/or σ) are varied. In both plots, heatmaps below the diagonal describe the case where tumours are treated with protocol A at diagnosis and heatmaps above the diagonal describe the case where protocol B is applied at diagnosis. In both plots, diagonal lineplots describe the value of $p_{R_{12}}$ over

time as a single parameter is varied whereby different colours indicate different values of the parameter (yellow lines corresponds to high values and blue lines corresponds to low values) and different line styles indicate which protocol is being applied (block lines indicate protocol A and dashed lines indicate protocol B). In heatmaps presented in Figure 2.6.1, darker and lighter regions indicate quicker and slower times to progression, respectively. In heatmaps presented in Figure 2.6.2, darker and lighter regions indicate lower and higher proportions of R_{12} cells at progression, respectively. White areas in Figure 2.6.2 indicate that no R_{12} cells are present at progression, $p_{R_{12}} < 8.3 \times 10^{-9}$ $(1/(1.2 \times 10^8))$.

Referring to the heatmap plots in Figure 2.6.1 describing progression times, we observe that, under protocol A (simultaneous TKI treatment and continuous, moderate dose, heatmaps below the diagonal), changes in μ (proportion of cells that mutate per cell division), δ (relative growth rate of resistant cells to sensitive cells) and γ (effect of S cells on the growth rates of resistant cells) drive considerable variation in progression times, with times ranging from approximately 25 years to less than 3 years. Large values of μ, δ and γ result in rapid progression, with progression times of less than 3 years corresponding to higher values of these parameters.

In Figure 2.6.2, that presents $p_{R_{12}}$ values at progression, under protocol A (below the diagonal), for changes in μ, δ and γ, we observe a small range of $p_{R_{12}}$ values, with $10^{-4} \leq p_{R_{12}} \leq 10^{-1}$. We conclude that tumours with varying progression times may have similar proportions of fully drug resistant cells at progression. The poor progression times observed for large values of μ, δ and γ, are due to a large number of R_1 and R_2 cells present at diagnosis. The partial resistance of these cells means that the tumour burden is not controlled, resulting in quicker times to progression and higher proportions of fully drug resistant cells at progression.

Analogously, for tumours undergoing protocol B (high dose pulsed TKI treatment), progression times range between 0 and 16 years (see heatmaps above the diagonal in Figure 2.6.1). In particular, for a wide range of values of μ, δ and γ, similar progression times are observed and diverse values of $p_{R_{12}}$ are recorded, with fully drug resistant tumours ($p_{R_{12}} \geq 10^{-1}$) occurring at progression for high μ, δ and γ values.

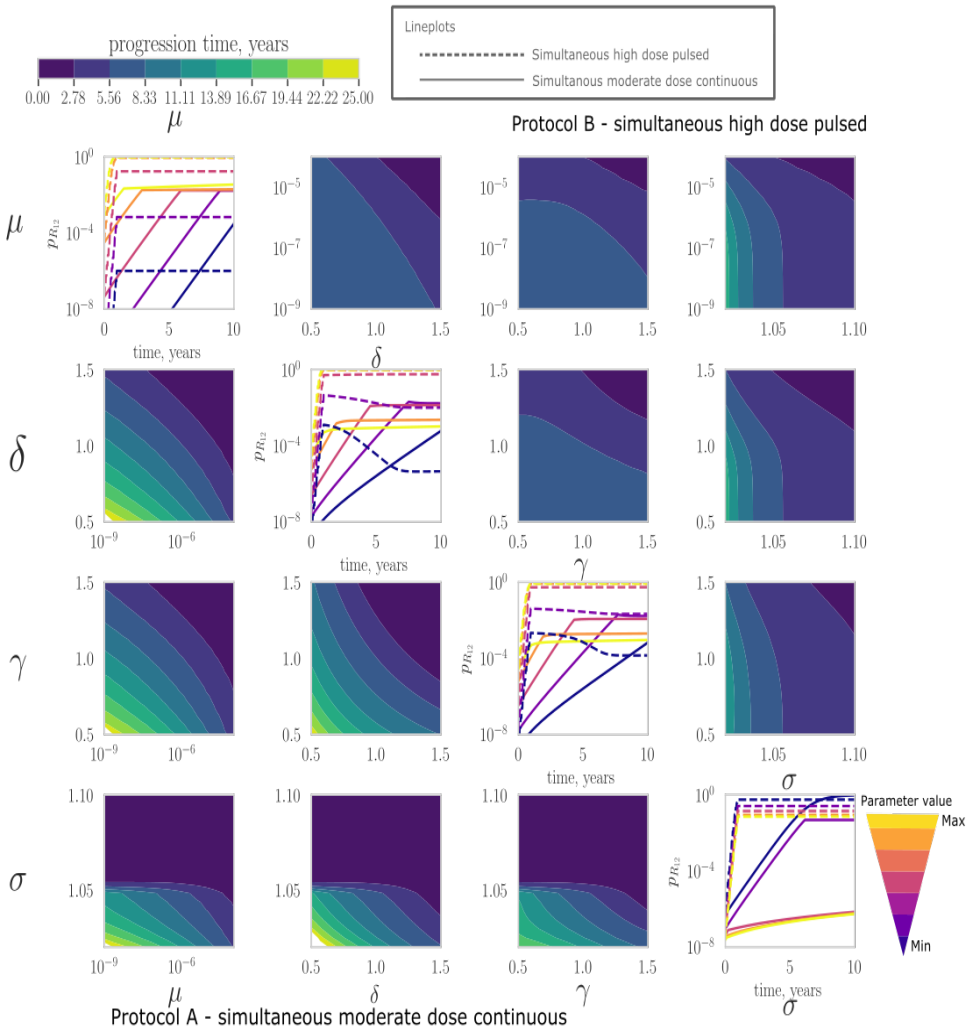


Figure 2.6.1: Predicted time to progression (20% increase in total tumour population size from diagnosis), in years from the start of treatment (heatmap plots) and $p_{R_{12}}$ values over time (diagonal lineplots). Heatmaps: Heatmaps represent progression times as two key parameters in (μ, δ, γ, σ) are varied simultaneously. Below the diagonal presents results from tumours undergoing protocol A. Above the diagonal presents results from tumours undergoing protocol B. The colourbar presents a range of progression times, $[0, 25]$ years. Lineplots: $p_{R_{12}}$ over time in which one key parameter is varied in a single plot. Dashed lines indicate tumours undergoing high dose, pulsed treatment (Protocol B) and block lines indicate tumours undergoing continuous, moderate treatment (Protocol A). Ranges of key parameter values are given in Table 2.2.1 with yellow lines representing the maximum parameter value and blue, the minimum. When parameters are not varied, parameters take the default values given in Tables 2.2.1.

In Figure 2.6.1, we observe earlier progression times corresponding to large values of $\sigma > 1.06$ (lower cell death rates), for both protocols. In particular, for $\sigma > 1.06$, progression times are < 5.6 years for protocols A and B. Not only do lower values of

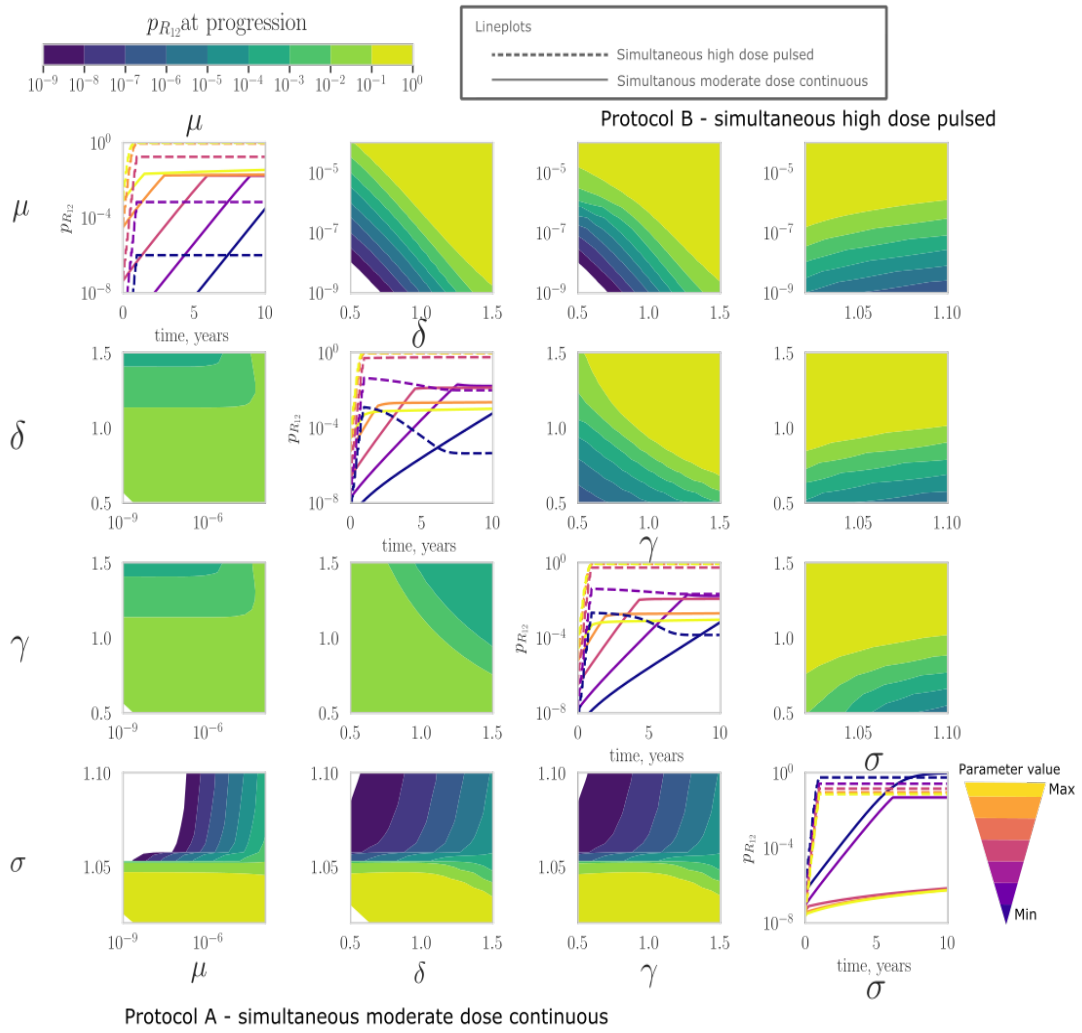


Figure 2.6.2: Predicted $p_{R_{12}}$ at progression (heatmap plots) and $p_{R_{12}}$ over time (lineplots). Heatmaps: Heatmap plots represent $p_{R_{12}}$ values at progression as two key parameters in the set (μ, δ, γ, σ) are varied simultaneously. Heatmaps below the diagonal present results from tumours undergoing protocol A. Heatmaps above the diagonal present results from tumours undergoing protocol B. The colourbar presents a range $p_{R_{12}}$ values at progression of $[10^{-8}, 1]$. White regions represent $p_{R_{12}} < 8.3 \times 10^{-9}$ $(1/(1.2 \times 10^8))$, when no R_{12} cells are present. Lineplots: $p_{R_{12}}$ over time in which one key parameter is varied in a single plot. Dashed lines indicate tumours undergoing high dose, pulsed TKI treatment (protocol B) and block lines represents tumours undergoing continuous moderate TKI treatment (protocol A). Ranges of key parameter values are given in Table 2.2.1 with yellow lines representing the maximum parameter value and blue, the minimum. When parameters are not varied they take the default values given in Table 2.2.1.

$\sigma < 1.06$ slow tumour growth, they also lead to a higher rate of cell kill of drug-sensitive cell populations. Hence, our model predicts that the drug will have a greater effect on a cell with a higher death rate.

In Figure 2.6.2, for tumours undergoing protocol A, with values of $\sigma < 1.05$, tumours are effectively fully drug-resistant at progression ($p_{R_{12}} \geq 10^{-1}$) and progression times span a range of 25 years (see Figure 2.6.1). These values of σ ($\sigma < 1.05$), correspond to higher death rates. Therefore, for these values, a stronger effect due to the drug is experienced by the cells as it linearly increase the death rate. Hence, a larger proportion of drug-sensitive cells will be removed at these σ values which could result in tumours with later progression times, yet with higher proportions of resistance at progression. Conversely, for $\sigma > 1.05$, a diverse range of $p_{R_{12}}$ values is recorded (see Figure 2.6.2) as a result of changes in μ, δ and γ values. For tumours undergoing protocol B, changes in σ do not greatly affect the $p_{R_{12}}$ values (see above the diagonal in Figure 2.6.2) and we observe later progression times for lower values of σ.

Diagonal entries in Figures 2.6.2 and 2.6.1 are the same across the two plots and show how the proportion of fully drug-resistant cells changes over time in the tumour while one key parameter is varied. The proportion of R_{12} cells increases in tumours undergoing high dose pulsed treatment (dashed lines) for all parameters as drug-sensitive cells are killed by the drug. For a given treatment, we observe the largest variation in $p_{R_{12}}$ for changes in μ, followed by δ, γ and σ.

In summary, our model predicts a diverse range of treatment outcomes when tumours with different properties are treated with different TKIs. These results indicate possible drivers of heterogeneity in the clinic and unveil the effect a particular treatment protocol has on the prevalence of resistance in a tumour at progression. In particular, we have found that two tumours with similar progression times may have markedly different levels of resistance, which could drive diverse responses to second-line treatment. We investigate the impact of resistance at progression on the efficacy of second-line treatment in the next chapter. We note that, for a given set of tumour-characteristics, neither protocol was able to simultaneously delay times to progression whilst also reducing the proportion of R_{12} cells at progression.

2.7 Discussion

Clinical results have demonstrated the failure of first and second-line TKI treatment in the clinic. Ongoing clinical studies are investigating the role of combination TKI treatment on tackling drug resistance [119]. In this chapter, we have developed an ODE model of tumour growth consisting of four cell populations which range from fully drug-sensitive to fully drug resistant to two TKIs. By incorporating factors in our model that are known to influence tumour heterogeneity and drug resistance, we have attempted to understand how certain tumour-specific factors affect pre-existing resistance and a tumour's response to combination TKI treatment.

Performing a global parameter sensitivity analysis on the model, we showed that the proportion of cells that mutate per cell division, μ, the relative growth rate of resistant cells to sensitive cells, δ and effect sensitive cells have on the growth rates of resistant cells via cell-cell interactions, γ have the greatest influence on the number of drug resistant cells in an untreated tumour. We simulated our model for a range of values of μ, δ and γ and found that the proportion of fully drug resistant cells increases at diagnosis, with magnitudes spanning six orders where μ had the greatest effect followed by δ and γ.

Interestingly, based on our global sensitivity analysis results, we find the birth to death rate ratio, σ to be the fourth most influential parameter affecting pre-existing resistance levels at diagnosis, with lower values of σ (which correspond to higher cell death rates) resulting in larger resistant proportions. We conclude that this is because it takes longer for the tumour to be diagnosed which results in longer periods of time in which fully drug resistant cells can emerge. These results corroborate those presented in [67, 53]. Models developed in [67, 53] have predicted an increase in resistant populations in a tumour of a certain size when higher death rates are assumed.

A method presented by Steel et al. has been used to estimate cell turnover rates in vivo. Steel et al. introduced the *cell loss factor* (as discussed in section 2.3). Aforementioned, cell loss factors have been shown to vary between 0.79 and 0.95 across adenocarcinoma pulmonary tumours [73]. In our model, we have considered values of σ which correspond to cell death rates up to 98% that of birth rates. At such values, we have demonstrated the effects of higher cell death rates on resistance levels. In light of our

findings, we suggest that more attention be focussed on the impact of cell turnover on resistance levels.

A key contribution of this chapter is the necessity to consider not only mutation rates and fitness levels of resistant cells, when investigating the emergence of drug resistance in a tumour, but also the effects of cellular interactions and how they affect proliferation rates of cells and tumour composition over time. Using our model, we demonstrated noticeable differences in tumour composition, driven by types of interactions between sensitive and resistant cell types.

The results and model parameterisations used in this chapter give rise to qualitative outcomes. In practice, Chmielecki et al. demonstrated variation in TKI treatment response due to variation in the abundance of T790M positive EGFR-mutant NSCLC cells in the tumour population [24]. Chmielecki et al. presented in vitro studies describing the response of EGFR-mutant tumours to erlotinib, a second-generation TKI. They considered a range of tumours consisting of parental drug-sensitive NSCLC cells and T790M cells at different proportions. They found that the larger the proportion of T790M cells in the tumour population, the weaker the effect of the drug on the tumour. In this chapter, although our model contains four cell populations, our results present the same qualitative outcomes as the in vitro studies presented by Chmielecki et al.. We have shown that an increase in the proportion of cells that mutate per cell division and in the cell division rate of resistant cells results in weaker effects of the drug as there is a larger proportion of resistant cells in the tumour.

We extended our model to consider the effects of treatment and we focussed on treatment protocols involving two TKIs in combination. A question that was of particular interest was, to what extent do changes in the key parameters drive diverse responses to treatment. We found that, in a tumour undergoing continuous TKI treatment at moderate dose, changes in the mutating proportion of cells per cell division and growth rates of resistant cells drive extensive variation in progression times (0 to 20+ years), with less variability in fully drug-resistant levels at progression. For tumours undergoing high dose, pulsed combination TKI treatment, however, small variation in resistance levels were observed, yet larger variations in progression times were predicted. An increased

level of resistance will have considerable impact on the efficacy of second-line treatment. Hence, this result highlights the need for careful consideration when selecting second-line treatments for patients whose tumours progress at similar times. Based on these results, we would suggest careful examination of tumour composition at progression, as drug resistance levels at this stage will greatly influence the efficacy of second-line treatment. While biopsies can be taken, they provide a spatially and temporally limited snapshot of the tumour composition and the process of collecting such samples is invasive. Liquid biopsies however, detect and characterise circulating free DNA (cfDNA) in the blood by only requiring a blood sample and, have also been established to detect the emergence of drug resistance, including NSCLCs [30, 128].

In this chapter we considered two TKI combination treatment protocols. However our model can be applied to any tumour type and treatment regime with a different parametrisation of the model. In particular, the model lends itself to studies of a range of treatment protocols. This is highly relevant as a phase II clinical trial applying alternating osimertinib and gefitinib treatment is ongoing [119]. In order to improve the relevance of our model to a clinical relevance, we implemented a number of realistic and well founded modelling assumptions, with practical tumour sizes at diagnosis and treatment protocols motivated by recent NSCLC clinical trials [34, 119, 151]. By making these assumptions, our model can be used to generate hypotheses about the evolution of resistance, and how to combat it, in a clinical setting.

In this chapter we do not consider pharmacokinetic models when considering treatment. Due to the complexity of the model, including pharmacokinetic models may limit our understanding of the model behaviour. If, in future work, we were to simplify the model, the inclusion of pharmacokinetic models would be interesting.

In this chapter, we have not presented steady state analysis. Such analysis could help us to determine constraints on model parameters. We attempted steady state analysis using Mathematica during the early phases of model development. However, the steady state solutions were intractable due to the complexity of the model and due to the inclusion of population-dependent growth rates. Hence, we have gained insights using numerical investigations rather than analytical. If we were to assume fixed growth rates

rather than population-dependent growth rates, steady state solutions may be tractable and this is subject to future work.

A possible extension of this work is to develop an agent-based modelling (ABM) framework to investigate the effects of space and stochasticity on tumour cell growth. To carry out this research, we would first simplify the model. For example, we might focus initially on a tumour consisting of S, R_1 and R_{12} cells and a single TKI. In our well-mixed model presented in this chapter, growth rates depend on the proportions of other cell types. In a spatially resolved model however, the growth of each cell is likely to be more heavily influenced by its neighbouring cells than those that are further away. Furthermore, the stochastic nature of the mutations may also shape the tumour composition; we will address this in the next chapter.

One limitation of our model is the assumption that cell-cell interactions directly influencing the birth and death rates of the cells when competition between cells is assumed. Under our assumption of logistic growth, for a given cell line, the growth limiting term in our model, $(-r\frac{S+R_1+R_2+R_{12}}{K})$ is proportional to the population dependent growth rate, r of that cell. If we suppose r is increased due to the presence of another cell type, at large cell populations, the strength of competition on these cells will increase. A simplification of the model could take place in which exponential growth is assumed. This simplified model could be applied to in vitro, co-culture tumour growth experiments, such as those in [72], and can be used to quantify the effects of cell-cell interactions on cell growth rates.

To summarise, in this chapter, we developed an ODE model to identify how different tumour characteristics influence pre-existing resistance and responses to TKI treatment. We identified key drivers of patient heterogeneity and investigated their contribution to diverse treatment outcomes seen in the clinic.

Extinction of cells and tumour eradication during treatment are two phenomena which can occur in tumours. Their investigation requires the consideration of stochastic events at small cell numbers, conditions under which an ODE may not be valid. In the next chapter, we will address this shortcoming by introducing a stochastic framework in which random fluctuations can lead to stochastic tumour cell dynamics. We will use

the same modelling assumptions presented here, in order to identify conditions under which stochastic effects impact our conclusions reached in this chapter. In particular, we will explore whether, for a given parameter regimen, stochastic models are better suited to modelling the evolution to drug resistance than deterministic models of the type presented in this chapter.

Appendix

2.A Extended Fourier Amplitude Sensitivity Test (eFAST)

In this section, we describe eFAST and demonstrate its implementation on a simplified version of our untreated tumour growth model defined in equations (2.2.1)-(2.2.4) [120]. We then present results of convergence testing, carried out to determine the appropriate number of samples, N_S and resamples N_R when performing eFAST on the untreated tumour growth model. We carry out convergence testing to ensure reliability in our eFAST results.

eFAST partitions the variance in the model output by determining what fraction of the variance is due to variation in each input parameter. eFAST gives the first and total-order sensitivity indices, S_i and S_{T_i}, respectively, for each model parameter [120].

Description of eFAST using a simplified model example

To demonstrate how eFAST works, we implement the method on a simplified version of the full model. In the simplified model, we neglect population-dependent birth rates and assume cell growth rates are fixed. The simplified model describes the growth of a tumour in the absence of drug, consisting of S and R_1 cells only. S cells mutate to produce R_1 cells and we assume cell populations compete for resources with a carrying capacity, K. Using the same notation as in the main text, the dimensional equations of our simplified model are as follows:

$$\frac{dS}{dt} = \alpha\left(1-\frac{1}{\sigma}\right)S\left(1-\frac{S+R}{K}\right) - \mu\alpha S, \tag{2.A.1}$$

$$\frac{dR}{dt} = \alpha\delta\left(1-\frac{1}{\sigma}\right)R\left(1-\frac{S+R}{K}\right) + \mu\alpha S. \tag{2.A.2}$$

Parameter	Description	Units	Uniform pdf range
α	Rate of cell division of S cells	days^{-1}	[0.11, 0.27]
δ	Relative growth rate of R cells to S cells	dim.	[0.5, 1.5]
σ	Birth to death rate ratio	dim.	[1.02, 1.1]
K	Carrying capacity	cells	$[1.3 \times 10^8, 1.3 \times 10^{12}]$
μ	Proportion of mutating cells per cell division	dim.	$[10^{-9}, 10^{-4}]$
dummy	dummy parameter which has no effect on the model dynamics	dim.	$[0, 10]$

Table 2.A.1: Table of parameters and ranges for simple model parameters defined in equations (2.A.1)-(2.A.2). We use the same parameter notation as that presented in the main text. We also use the same parameter ranges. References for the sources of the parameter ranges can be found in Table 2.2.1.

We assume initially that the tumour contains only S cells, with $S(0) = 100$ and $R_1(0) = 0$. Model parameters for the simplified model are summarised in Table 2.A.1, together with the range of parameter values employed in our eFAST analysis with uniform sampling (these ranges are the same ranges as those used in the main text). Our model output of interest is the proportion of R_1 cells at 5 years, where $p_{R_1} = R_1/(S+R_1)$.

EFAST implementation on simple model

1. *Define parameter ranges* For each model parameter, we define a range of values the parameter can take (see Table 2.A.1). To test whether the influence of a particular parameter on the model output is significant, we introduce a dummy parameter (see Table 2.A.1) which has no effect on model output [90].

2. *Assign frequencies to parameters* Every model parameter is assigned an integer frequency ω_i. Using the protocol proposed by Saltelli et al. [120].

3. *Obtain samples of parameter values* For a given parameter, i with an assigned frequency w_i, a sinusoidal function (a search curve), $x = g(s)$ for $s = 1, ...N_S$ is used, where

$$x(s) = \frac{1}{2} + \frac{1}{\pi} \sin^{-1}(\sin(\omega_i s + \phi_i)), \tag{2.A.3}$$

where N_S denotes the number of samples and $0 \leq \phi_i \leq 2\pi$ is a random phase shift uniformly sampled in the range $[0, 2\pi]$. Hence, $x(s)$ is a vector of length N_S whose values range between 0 and 1. The parameter value is calculated as the inverse of

the uniform cumulative distribution function at x_l in $x(s)$ for $l = 1, ..., N_S$, whose range is predefined (see Table 2.A.1). Hence, we obtain a set of parameter values, $p(s)$ of length N_S whose values range between the upper and lower bounds, defined in Table 2.A.1 where,

$$p(s) = p^{min} + (p^{max} - p^{min})x(s). \tag{2.A.4}$$

The uniform distribution is appropriate when there is a large documented range of possible parameter values (for example, mutation rates) and we have little *a priori* knowledge of the model parameters [90].

In global sensitivity analysis all model parameters are varied. Hence, the sinusoidal functions for each of the input parameters, when combined, produce a multidimensional search curve, resulting in many different parameter combinations being employed in the model. As we are considering parameter values that span across magnitudes, we consider a uniform distribution of the log (base 10) of the parameter value to ensure sampling is not densely clustered around certain values. For example, the mutation probability has a range of $[10^{-9}, 10^{-4}]$ (see Table 2.A.1). We sample from the range [-9, -4] and apply the sampled log (base 10) values back into the mutation probability when simulating the model. We do the same for the carrying capacity, K.

4. *Resampling* We implement a random resampling scheme (number of resamples, N_R) for each search curve by introducing a vector consisting of random phase shifts, ϕ_i (see equation (2.A.3)), each corresponding to a single model parameter, with a unique vector employed for each resample ($h = 1, ..., N_R$), in the sinusoidal function (search curve). For each parameter, we obtain N_R distinct search curves each with N_S samples.

5. *Generating model outputs from our parameter samples* For each parameter, we have N_R sets of N_S samples, giving a total of (number of parameters)$\times N_R \times N_S$ sets of model parameters. We simulate our model for each sample and obtain a total of (number of parameters)$\times N_R \times N_S$ model outputs.

6. *Performing the eFAST algorithm* For each resample, h for $h = 1, ...N_R$, we perform the eFAST algorithm to obtain first and total-order sensitivity indices for model parameters.

7. *Total variance in model output* For a given resample, h, $h = 1, ...N_R$, the total variance in model output is written as,

$$s^2 = \sum_{j=1}^{N} \frac{(y_j - \bar{y})^2}{N-1}, \tag{2.A.5}$$

where N denotes the total sample size (total number of model runs in the resample, (number of model parameters)$\times N_S$), y_j describes the model output of the j^{th} sample and $\bar{y}$ describes the sample mean.

8. *Calculating first-order sensitivity index, S_i* The first-order sensitivity index, S_i of parameter i is determined by holding all other model parameter values fixed and measuring the variation in the model output when the single parameter i is varied. The variation in model output due to variation in parameter i, is defined as follows:

$$s_i^2 = 2(A_i^2 + B_i^2), \tag{2.A.6}$$

where,

$$A_i = \frac{1}{\pi} \int_{-\pi}^{\pi} f(u) \sin(w_i u) du, \tag{2.A.7}$$

and,

$$B_i = \frac{1}{\pi} \int_{-\pi}^{\pi} f(u) \cos(w_i u) du. \tag{2.A.8}$$

A_i and B_i in equations (2.A.7)-(2.A.8), are the Fourier coefficients at frequency w_i and f denotes the model output. The first-order sensitivity index is,

$$S_i = s_i^2 / s^2, \tag{2.A.9}$$

that is the fraction of model output variance produced by variation in the input parameter where s^2 describes the total output variance defined by equation (2.A.5).

9. *Total-order sensitivity index, S_{Ti}* The total-order sensitivity index, S_{Ti} for parameter i is the remaining variance in model output when the first-order sensitivity

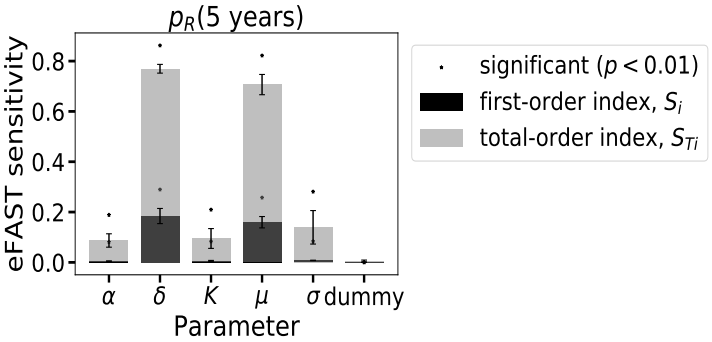


Figure 2.A.1: First and total-order sensitivity indices of simplified model parameters for $p_{R1}(1865)$. * presents significance at p-value < 0.01 when compared to the dummy parameter. Grey represents S_{Ti}, black represents S_i. $N_S = 4000$, $N_R = 5$.

indices associated to all other model parameters is removed. S_{Ti} is defined as follows

$$S_{Ti} = 1 - S_{Ci} \tag{2.A.10}$$

where S_{Ci} is the combined first-order sensitivity indices of the complimentary set of parameters, which is calculated using equation (2.A.9), with frequencies associated to the parameters in the complimentary set.

10. This process is repeated for each model parameter and N_R resamples. Therefore, we obtain N_R values of S_i and S_{Ti} for each model parameter, including the dummy parameter.

11. *Significance testing* To test whether the influence of a particular parameter on the model output is significant, we use a 2-sample t-test to determine whether the distribution of sensitivity indices (N_R in total) for a given parameter i is significantly larger than the distribution of indices associated with the dummy parameter at a significance level of 0.01 [90]. By doing so, we assume the distributions of N_R sensitivity indices are normally distributed.

We present eFAST results showing sensitivity indices for the simplified model with $N_S = 4000$ and $N_R = 5$, in Figure 2.A.1. The model output of interest is the proportion of resistant cells 5 years after tumour initiation. The results show that the relative fitness of resistant cells, δ, followed by the mutation probability, μ, have the largest effect on the

proportion of R_1 cells, 5 years after tumour initiation. In the model, since μ and δ are present in terms with other parameters, the total-order sensitivity index is larger than the first-order sensitivity index as it accounts for higher order parameter interactions.

2.B Convergence testing on eFAST

By performing eFAST on the untreated tumour growth model presented defined in equations (2.2.1)-(2.2.4), we determine which model parameters (see Table 2.2.1 for a summary of model parameters) influence the proportion of fully drug-resistant cells, $p_{R_{12}}$ at diagnosis and over time.

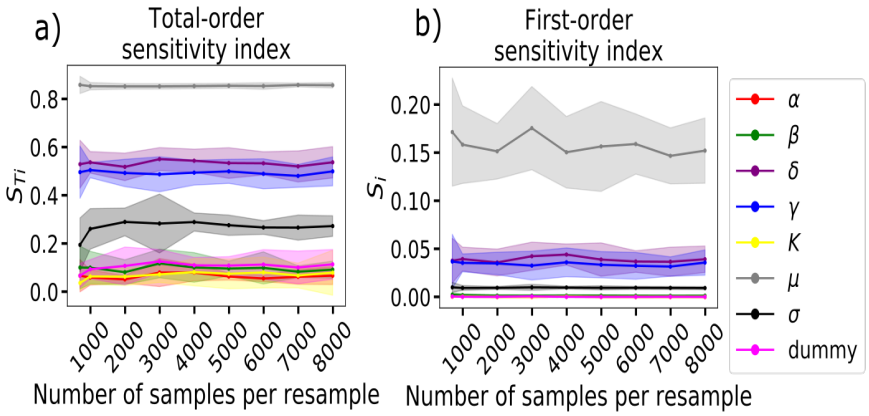


Figure 2.B.1: Results of global sensitivity analysis applied to untreated tumour growth model in the absence of drug: a) and b) First and total-order sensitivity indices for model parameters at diagnosis for increasing sample sizes, N_S with $N_R = 10$ (number of resamples), respectively. Block lines represent the sample mean and shaded regions represent two standard deviations about the mean sensitivity index for a given parameter. Parameters are varied uniformly from a uniform distribution whose ranges are given in Table 2.2.1.

We performed eFAST on our full model for a range of values of N_S values in order to determine the appropriate number of samples needed for reliable results. We present the results of our convergence testing in Figure 2.B.1.

Figures 2.B.1(a) and (b) show how the total and first-order sensitivity indices for each model parameter changes as we increase the numbers of samples, when $N_R = 10$. Based on these results, we conclude that $N_S = 5000$ is a suitable number of samples as we observe no change in the hierarchy of the top four most influential parameters for $N_S > 5000$. Furthermore, the spread (described by the shaded regions that describe $\pm$

2 standard deviations about the mean index) decreases slightly, however the change in spread about the mean value is small for increasing N_S for all parameters.

Saltelli et al. proposed a suitable number of resamples whereby $16 < \omega_{max}/N_R < 64$ ensures sampling of parameters over each curve is not too sparse ($\omega_{max}/N_R = 16$) or too dense ($\omega_{max}/N_R = 64$) where ω_{max} is the largest frequency among the set of frequencies assigned to model parameters [120]. For $N_S = 5000$ and $N_R = 10$, the maximum frequency assigned to a parameter $\omega_{max} = 624$ for the number of model parameters in our full model, hence $\omega_{max}/N_R = 62.4$. We conclude therefore that sampling of the model parameters is suitably spaced along the search curve.

Chapter 3

A stochastic model of tumour growth and response to combination TKI treatment

3.1 Introduction

In the previous chapter, we developed a spatially averaged model of tumour growth and we investigated the effects of tumour-specific factors on the proportions of resistant cells at diagnosis and during treatment. We found changes in the tumour's propensity to mutate and, in the growth rates of resistant cells to drive variation in pre-existing resistance levels and treatment response. In this chapter we reformulate the ODE as a stochastic birth-death process. We were motivated to develop the stochastic model, presented in this chapter due to the following reasons. Random mutation events lead to tumour cell populations with low copy numbers. At low numbers, these cells are subject to stochastic fluctuations which may lead to cell extinction and tumour removal, driven by treatment. Such phenomena can not be generated in a deterministic setting. In particular, two realisations of a stochastic model, initiated from the same set of model parameter values, may yield divergent growth dynamics and cellular composition. Hence, stochastic fluctuations at low cell numbers are important as they may amplify variability in resistance levels at diagnosis and subsequent responses to treatment.

In chapter 2, we simulated tumours with different characteristics by employing different values of mutation probabilities and growth rates of resistant cells in the deterministic model. In so doing, we investigated how changes in the mutation probabilities and growth rates of cells affected patient heterogeneity at diagnosis, and during treatment.

In this chapter, we examine heterogeneity across tumours with the same characteristics (i.e. mutation probabilities and growth rates of resistant cells), when stochastic effects are considered. We perform multiple realisations of the stochastic model and quantify the variation in, cellular dynamics in an untreated tumour, pre-existing resistance levels at diagnosis and, response to treatment within the set of stochastic realisations, where each repeat is initiated with the same parameter values. In so doing, we compare the level of variation within a set of stochastic simulations, initiated from the same set of parameter values, to the level of variation in a different set of stochastic simulations, initiated from a different set of model parameters, where a different probability of mutation and growth rate is assumed.

We aim to answer the following research questions;

i In an untreated tumour, how does variation (within a set of stochastic repeats) in cellular dynamics over time, and in levels of pre-existing resistance at diagnosis, differ when different values of the mutation probability and the relative growth rates of the resistant cells (to sensitive cells), are employed in the model?

ii How does variation (within a set of stochastic repeats) in treatment response differ for different mutation probabilities?

iii In what regions of the parameter space do outcomes from the stochastic model deviate from those of the deterministic model, developed in chapter 2?

By addressing questions (i)-(iii), we aim to shed light on the pros and cons of using stochastic and deterministic modelling approaches to study the emergence of drug resistance and to predict responses to treatment. We aim also to highlight regions of the parameter space in which stochastic effects could have a significant effect on model predictions and when they can be neglected.

As discussed in chapter 2, the models developed in this chapter are motivated by studies that consider competition and cell death when modelling drug resistance [35]. In chapter 1, we discussed models presented by Haeno et al. and Iwasa et al. who modelled exponentially growing populations of sensitive and resistant tumour cells using

birth-death branching processes with a fixed probability of mutation per cell division [53, 67]. In this chapter, we generalise such models to consider four cell populations whereby stochastic birth, death and mutation events can take place and we consider TKI treatments and these features represent the novel aspects of the modelling developed in this chapter.

The remainder of this chapter is structured as follows. In section 3.2, we develop our stochastic model. The methods that we use to analyse the model are outlined in section 3.3. In section 3.4, we present simulation results and quantify the variation in model outputs as the mutation probabilities and resistant cell growth rates vary. The chapter ends in section 3.6, where we discuss our findings.

3.2 Stochastic model

We develop a stochastic analogue of the deterministic model presented in chapter 2 to describe the growth of a tumour and its response to treatment. We carry over the same modelling assumptions and parameter notation as in chapter 2. As such, we assume that the tumour consists of four cell populations which differ in their resistance to two tyrosine kinase inhibitors (TKIs). The number of cells at time, t, that are sensitive to both drugs is denoted by $S(t)$. The number of cells that are resistant to drug one and sensitive to drug two is denoted by $R_1(t)$. The number of cells that are resistant to drug two and sensitive to drug one is denoted by $R_2(t)$, and the number of cells resistant to both drugs is denoted by $R_{12}(t)$. We impose the same initial conditions as in chapter 2, assuming that, at $t = 0$, the tumour comprises 100 sensitive cells (S cells), and no resistant cells.

Before we describe a typical stochastic realisation, we present propensities associated with the cellular reactions (birth, death and mutation) that can take place on a given time step. We summarise the parameter values that we use for the model simulations, and outline the algorithm used to simulate our stochastic process. We note that, in what follows, we use "reactions" and "events" interchangeably when referring to cellular reactions.

3.2.1 Event propensities

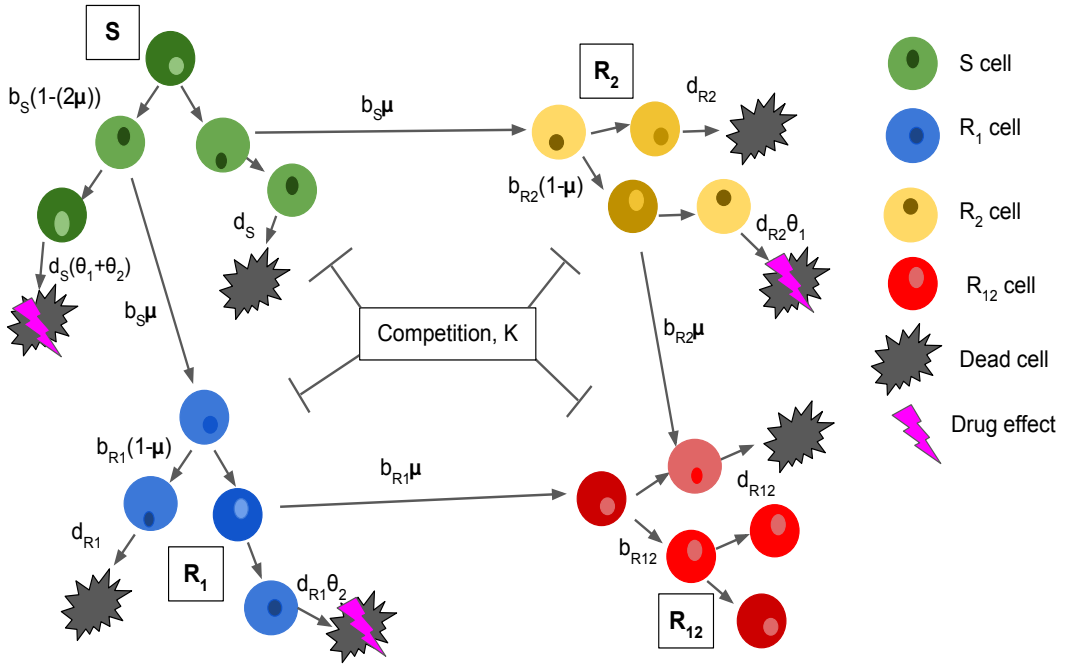


Figure 3.2.1: Schematic of our stochastic model of tumour growth and treatment response. Four cell populations are considered. Sensitive cells (green cells), are denoted by $S(t)$; cells that are sensitive to drug two and resistant to drug one (blue cells) are denoted by $R_1(t)$; cells that are sensitive to drug one and resistant to drug two (yellow cells) are denoted by $R_2(t)$; fully drug-resistant cells (red cells) are denoted by $R_{12}(t)$. The parameter μ denotes the intrinsic probability of mutation per cell division from (S to R_1), (S to R_2), (R_1 to R_{12}) and from (R_2 to R_{12}). S cells divide to produce S cells, ($S \to S + S$), R_1 cells ($S \to S + R_1$) and R_2 cells ($S \to S + R_2$) at rates, $b_S(1 - 2\mu)$, $b_S\mu$ and $b_S\mu$, respectively. S cells die at rate, d_S and the net growth rate of S cells is $r_S = b_S - d_S$. Similarly, R_1 cells divide to produce R_1 cells ($R_1 \to R_1 + R_1$) and R_{12} cells ($R_1 \to R_1 + R_{12}$) at rates, $b_{R_1}(1 - \mu)$ and $b_{R_1}\mu$, respectively. R_1 cells die at rate, d_{R_1} and have a net growth rate of $r_{R_1} = b_{R_1} - d_{R_1}$. Similarly for R_2 cells. R_{12} cells divide to produce R_{12} cells at rate, $b_{R_{12}}$ and die at rate, $d_{R_{12}}$ and have a net growth rate, $r_{R_{12}} = b_{R_{12}} - d_{R_{12}}$. S cells are affected by drugs one and two at rate, $d_S(\theta_1 + \theta_2)$. R_1 cells are affected by drug two at rate, $d_{R_1}\theta_2$ and R_2 cells are affected by drug two at rate, $d_{R_2}\theta_1$. Propensity functions associated to cellular reactions presented in this schematic are summarised in Table 3.2.1 along with death rates associated to cellular competition.

Figure 3.2.1 presents a model schematic describing the cellular reactions (birth, death and mutation) that can take place on a given time step of duration, τ. For completeness, we summarise these events here:

i cell division, without mutation,

ii cell division, with mutation (for S, R_1, R_2 cells only as R_{12} cells do not mutate),

iii cell death

The propensities that we associate with these events are summarised in Table 3.2.1 and illustrated in Figure 3.2.1. In Table 3.2.1, we denote, by Λ_{n+1}, the state of the stochastic process at the end of the event, given that $\Lambda_n = [S, R_1, R_2, R_{12}]$.

Cellular reaction	Λ_{n+1}	**Propensity function**
Birth without mutation		
$S \to S + S$	$[S+1, R_1, R_2, R_{12}]$	$\overbrace{(1-2\mu)b_S S}^{\text{birth without mutation}}$
$R_1 \to R_1 + R_1$	$[S, R_1+1, R_2, R_{12}]$	$(1-\mu)b_{R_1}R_1$
$R_2 \to R_2 + R_2$	$[S, R_1, R_2+1, R_{12}]$	$(1-\mu)b_{R_2}R_2$
$R_{12} \to R_{12} + R_{12}$	$[S, R_1, R_2, R_{12}+1]$	$b_{R_{12}}R_{12}$
Birth with mutation		
$S \to S + R_1$	$[S, R_1+1, R_2, R_{12}]$	$\overbrace{\mu b_S S}^{\text{birth with mutation}}$
$S \to S + R_2$	$[S, R_1, R_2+1, R_{12}]$	$\mu b_S S$
$R_1 \to R_1 + R_{12}$	$[S, R_1, R_2, R_{12}+1]$	$\mu b_{R_1}R_1$
$R_2 \to R_2 + R_{12}$	$[S, R_1, R_2, R_{12}+1]$	$\mu b_{R_2}R_2$
Death		
$S \to \emptyset$	$[S-1, R_1, R_2, R_{12}]$	$\overbrace{d_S(\theta_1+\theta_2)S}^{\text{death due to drug}} + \underbrace{d_S S}_{\text{natural cell death}} + \overbrace{\frac{r_S}{K}(S(S-1))}^{\text{intraspecific competition}} + \underbrace{r_S\frac{S}{K}\Big(R_1+R_2+R_{12}\Big)}_{\text{interspecific competition}}$
$R_1 \to \emptyset$	$[S, R_1-1, R_2, R_{12}]$	$d_{R_1}R_1(1+\theta_2) + \frac{r_{R_1}}{K}(R_1(R_1-1)) + r_{R_1}\frac{R_1}{K}\Big(S+R_2+R_{12}\Big)$
$R_2 \to \emptyset$	$[S, R_1, R_2-1, R_{12}]$	$d_{R_2}R_2(1+\theta_1) + \frac{r_{R_2}}{K}(R_2(R_2-1)) + r_{R_2}\frac{R_2}{K}\Big(S+R_1+R_{12}\Big)$
$R_{12} \to \emptyset$	$[S, R_1, R_2, R_{12}-1]$	$d_{R_{12}}R_{12} + \frac{r_{R_{12}}}{K}(R_{12}(R_{12}-1)) + r_{R_{12}}\frac{R_{12}}{K}\Big(S+R_1+R_2\Big)$

Table 3.2.1: Propensity functions for the cellular reactions associated with the four cell types. These reactions include birth without mutation, birth with mutation, and death due to natural cell death, drug exposure and interspecific and intraspecific competition. Λ_{n+1} denotes the state of the stochastic process when a particular event occurs, given $\Lambda_n = [S, R_1, R_2, R_{12}]$.

For all cell lines, the propensity function associated with cell death incorporates natural cell death, death due to drug, and death due to competition within (intraspecific) and between (interspecific) cell compartments. For S cells, we include a factor, $(S-1)$, in the intraspecific competition term, as the cell numbers are discrete, and, when $S=1$, no interspecific competition will occur. The same argument applies to R_1, R_2 and R_{12}.

3.2.2 Parameters

In chapter 2, μ denotes the proportion of cells that mutate per cell division. In this chapter, as we consider the effects of stochasticity, therefore, μ denotes the fixed probability of mutation per cell division. We parametrise the model using the same ranges for the mutation probability, μ, and the relative growth rate of the resistant cells to sensitive cells, δ, as in chapter 2. All other parameters are fixed at the default values used in chapter 2. A summary of model parameters and values is given in Table 3.2.2.

In chapter 2, we assumed population-dependent rates of cell division and cell death. For simplicity, in this chapter we assume instead that the rates of cell division and cell death are fixed throughout a given stochastic realisation of the model (see Table 3.2.2) due to the following reasons. No two realisations of the stochastic process will produce exactly the same growth dynamics and tumour composition as time evolves. When population-dependent birth and death rates are implemented, these rates will also be subject to random fluctuations in the tumour composition. This randomness will, in turn, influence the variability in tumour compositions across the different realisations. In order to examine the effect of varying the mutation probabilities and relative growth rates of the resistant cells to the sensitive cells, we restrict attention to the fixed values of the cell division and death rates.

To summarise, our stochastic modelling framework comprises of the system of cellular reactions summarised in Table 3.2.1 with parameter values given in Table 3.2.2, and initial conditions prescribed in chapter 2.

3.2.3 Stochastic process simulation

We simulate the stochastic process using the τ-leap method with adaptive τ selection, a method presented by Cao et al. [17]. Details of the algorithm are given in the section

Param.	Description	Default value	Parameter range	Units
μ	Intrinsic probability of mutation per cell division	10^{-6}	$[10^{-8}, 10^{-6}, 10^{-4}]$	dim.
δ	Relative growth rate of resistant cells to sensitive cells	1	$[0.5, 1, 1.5]$	dim.
b_S	Rate of cell division of S cells	0.156	fixed	day^{-1}
b_{R_1}	Rate of cell division of R_1 cells	$b_S\delta$	see δ	day^{-1}
b_{R_2}	Rate of cell division of R_2 cells	$b_S\delta$	see δ	day^{-1}
$b_{R_{12}}$	Rate of cell division of R_{12} cells	$b_S\delta$	see δ	day^{-1}
σ	Birth to death rate ratio, $\frac{b_X}{d_X}$ for $X \in \{S, R_1, R_2, R_{12}\}$	1.05	fixed	dim.
d_S	Rate of natural cell death of S cells	b_S/σ	fixed	day^{-1}
d_{R_1}	Rate of natural cell death of R_1 cells	b_{R_1}/σ	see δ	day^{-1}
d_{R_2}	Rate of natural cell death of R_2 cells	b_{R_2}/σ	see δ	day^{-1}
$d_{R_{12}}$	Rate of natural cell death of R_{12} cells	$b_{R_{12}}/\sigma$	see δ	day^{-1}
r_S	Net growth rate of S cells (cell division rate minus natural cell death rate)	$b_S(1-1/\sigma)$	fixed	day^{-1}
r_{R_1}	Net growth rate of R_1 cells	$b_{R_1}(1-1/\sigma)$	see δ	day^{-1}
r_{R_2}	Net growth rate of R_2 cells	$b_{R_2}(1-1/\sigma)$	see δ	day^{-1}
$r_{R_{12}}$	Net growth rate of R_{12} cells	$b_{R_{12}}(1-1/\sigma)$	see δ	day^{-1}
K	Carrying capacity	5×10^8	fixed	cells
θ_1	Effect of drug one for cytostatic in protocol A	0.02	fixed	dim.
θ_1	Effect of drug one for cytotoxic in protocol B	0.6	fixed	dim.
θ_2	Effect of drug two for cytostatic in protocol A	0.02	fixed	dim.
θ_2	Effect of drug two for cytotoxic in protocol B	0.6	fixed	dim.

Table 3.2.2: Default parameter values, descriptions and their respective units (for references motivating these parameter values, see chapter 2). All model parameters are fixed except for the mutation probability, μ and, δ, the relative growth rate of the resistant cells to the growth rate of the sensitive cells. Dim. denotes dimensionless.

3.3 and are summarised below.

On a given time step, τ, the number of events (births, deaths and mutations) that occur within each cell population is Poisson distributed with a mean, equal to the product of the event propensity (given in Table 3.2.1) and the time step, τ. At the end of this time step, the time is updated and the state vector (Λ_{n+1}) is updated based on the net effect of all birth, death and mutation events that have taken place within a given cell population. Event propensities are also updated to account for changes in cell numbers and, a new time step, τ is calculated, whose value is bounded such that no appreciable change in values of propensity functions will take place within the next time step. For a given realisation of our model, we simulate our model from the tumour's initiation ($S(0) = 100$ and $R_1(0) = R_2(0) = R_{12}(0) = 0$ (cells)) until it reaches a detectable size of 10^8 cells. At this time, the tumour either, continues to grow untreated, or, treatment is applied. All simulations end 25 years after the tumour has reached a detectable size (diagnosis). Treatment for 25 years is not realistic in a clinical setting as adverse toxicity events can occur when treating patients with TKIs and further, the patient will not live this long. However, we are interested in the qualitative behaviour of the solutions of the stochastic model. In particular, we are interested in how tumour cell dynamics and composition differ when different tumour-specific properties, coupled with stochasticity are considered. Analysis of the stochastic model can pave the way for future investigations. If time-course data of tumour composition over time were to become available, we could attempt to validate the stochastic model and consider realistic timescales in line with the time-course data.

We apply the same treatment protocols A and B to the tumours, as in chapter 2.

- Protocol A involves the simultaneous and continuous application of two TKIs, at moderate doses. Protocol A is applied until progression occurs. Progression is when the total tumour population size reaches 120% that of the total tumour population at diagnosis (i.e. when the total tumour population at progression equals 1.2×10^8 cells). Simultaneous and continuous drug application describes the case in which the drug parameters, θ_1 and θ_2 are both constant and fixed at values greater than zero ($\theta_1 = 0.02$ and $\theta_2 = 0.02$) for the period of time in which

the tumour has not progressed. Hence, on a daily basis, the amount of the drugs do not vary from the fixed values, $\theta_1 = 0.02$ and $\theta_2 = 0.02$.

- Protocol B involves the simultaneous and continuous application of two TKIs at high doses for a period of 1 week followed by a 3-week break; this cycle is repeated over the course of a year. Analogous to Protocol A, when drugs are applied, on a daily basis, drug parameters, θ_1 and θ_2 are constant and fixed at values greater than zero ($\theta_1 = 0.6$ and $\theta_2 = 0.6$) and are fixed at values $\theta_1 = 0$ and $\theta_2 = 0$ when the drugs are not applied.

For computer simulations in which treatment is applied at diagnosis, we employ protocol A as a first-line treatment until progression occurs. We then switch to protocol B, as a second-line treatment. The drug effect parameters, θ_1 and θ_2 for drugs one and two, respectively, are held fixed at values which depend on whether protocol A or B is being administered (see Table 3.2.2).

For a given realisation of our model, we record the following times and the values of the proportions of R_{12} cells in the tumour, denoted by $p_{R_{12}}$, at those times:

- *Diagnosis* ($S + R_1 + R_2 + R_{12} = 10^8$ cells).

- *Progression on first-line treatment (protocol A)*, when total tumour population, at progression, equals $S+R_1+R_2+R_{12} = 1.2\times10^8$ cells). We also record progression times and $p_{R_{12}}$ values at progression, in untreated tumours.

- *Progression on second-line treatment (protocol B)*, when the tumour population returns to its size at the start of second line treatment ($S+R_1+R_2+R_{12} = 1.2\times10^8$ cells).

In order to increase the clinical relevance of our studies, we focus on tumour sizes that are observed in the clinic (10^8 cells at diagnosis [34]) and employ mutation probabilities (as low as $\mu = 10^{-8}$) that are taken from the literature (see Table 3.2.2). Within these tumours with low mutation probabilities, the number of R_{12} cells may be very small and, therefore, subject to random fluctuations. These may drive variation in tumour cell dynamics across model simulations.

Deterministic model simulation

We simulate the deterministic model, presented in chapter 2, in addition to simulating the stochastic model. We refer the reader back to chapter 2 for details of the deterministic model. When simulating the deterministic and stochastic models, so as to compare the tumour dynamics across the two frameworks, we employ the same parameter values and we fix birth and death rates in the deterministic model, thus neglecting population dependent birth-rates. We also record the key stages of tumour growth and treatment response (i.e. $p_{R_{12}}$ values at diagnosis and at progression on first and second line treatment and progression times) for the deterministic model simulations (see above).

3.3 Methods

We refer the reader to section 2.3 in chapter 2 of the thesis for details on both, how we simulate the deterministic model and how we plot the results.

3.3.1 Stochastic simulation algorithm

Before we present the algorithm used in this chapter to simulate our stochastic model, we outline Gillespie's stochastic simulation algorithm (SSA) and the τ-leaping method, from which our algorithm is derived [46, 47]. The SSA is an exact numerical method for simulating a well-mixed system of reactants whereby a single reaction takes place in a given time step. The SSA is therefore computationally expensive for large systems (e.g. tumour cell populations made up of 10^8 cells) as the calculated value of the time step, τ, is proportional to $(1/a_0(c))$ where $a_0(c)$ is the sum of all the propensity functions.

The τ-leap method is an approximation of the SSA which allows for larger systems to be simulated as multiple reactions can take place in a given time step. The τ-leap method results in quicker computation times, at the cost of accuracy. The number of events that occur during time step, τ is approximated by a Poisson random variable, with mean $a_j(x)\tau$, where $a_j(x)$ describes the propensity function associated with the reaction R_j.

We will now go into more detail about the τ-leaping method. The method assumes the following *Leap Condition: For the current state, Λ_n, τ must be small enough that*

the change in the state during times, t and $t + \tau$, will be small such that, no propensity function will suffer an appreciable change in its value [47]. By this it is meant that during the entire interval $[t, t + \tau)$, for a given reaction, R_j, its propensity function, $a_j(x)$, will remain essentially constant [47], that is, for any infinitesimal increment, dt, such that $t + \mathrm{dt} \in [t, t + \tau)$, $a_j(x)\tau$ will give the probability of the reaction, R_j firing, regardless of what the other reactions are doing [47].

A number of variants of the τ-leap method have been developed; the simplest assumes a fixed value of τ throughout [17]. A shortcoming in using a fixed value of τ is that, reactants at low copy numbers will have small values of propensity functions associated to them. Conversely, reactants with large copy numbers will have relatively large values of propensity functions. Hence, a large, predefined fixed value of τ may incur a large error, relative to the SSA, in the updated number of reactants with low copy numbers on a given time step. With this large change, the Leap Condition may therefore, not be met. A small fixed value of τ however, will increase the accuracy of the τ-leap method relative to the SSA yet, at the cost of longer computational times.

Tumours modelled in this chapter contain multiple subpopulations of cells with different sizes. To address this, we simulate our model using a τ-leap approximation method, with adaptive τ calculation, proposed by Cao et al. [17]. In this algorithm, at the end of each time step, τ is calculated such that relative changes in the propensity functions are uniformly bounded by a predefined tolerance parameter, ϵ ($0 < \epsilon \ll 1$). This ensures that the Leap Condition is met for each time step. The algorithm also prevents cell populations from becoming negative (this may occur if more deaths fire in a given time step than cells in the population) by employing a positive integer control parameter, n_c. Any reaction channel (for example, cell death of S cells) that is within n_c firings of exhausting the cell type is classified as a critical reaction. This τ-leaping procedure calculates values of τ such that, no more than one firing of a critical reaction will take place during the leap. Essentially, the algorithm simulates the critical reactions using the SSA and the non-critical reactions using the Poisson tau-leaping method. If a large n_c value is assumed, so that all reactions are critical, the modified Poisson τ-leaping procedure is effectively the SSA.

We summarise the steps of the algorithm here and we refer the reader to [17] for a more detailed description of the algorithm. Suppose we are at the n^{th} time step and the state vector of the stochastic process is $\Lambda_n = [S, R_1, R_2, R_{12}]$.

1. In state, Λ_n at time t_n, identify the critical reactions (critical reactions are defined above). This involves determining the maximum number of times, L_j, the reaction, R_j can fire before exhausting all its reactants. Any reaction is deemed critical if $L_j < n_c$. Go to step 2.

2. Compute the *candidate* time leap value, $\tau^{'}$ where, if all reactions are critical reactions, take $\tau^{'} = \infty$ and go to step 3. Otherwise, compute a value of $\tau^{'}$ that is based on only the non-critical reactions (shown in step 2 in [17]) and go to step 3. The value of $\tau^{'}$ is calculated such that the relative change in all the propensity functions is bounded by the tolerance parameter, ϵ. The value of $\tau^{'}$ is the largest permissible time step for the non-critical reactions.

3. If $\tau^{'}$ is less than $\frac{1}{10(a_0(x))}$ where $a_0(x)$ is the sum all the propensity functions, execute 100 steps of the Gillespie algorithm and return to step 1. Else, go to step 4.

4. Compute $a_0^c(x)$, the sum of the propensity functions of all the critical reactions. Generate a *second candidate* $\tau^{''}$ as a sample of the exponential random variable with a mean $1/a_0^c(x)$. $\tau^{''}$ is an estimate of the time to the next reaction.

5. Take the actual time step, τ to be the smaller of $\tau^{'}$ and $\tau^{''}$. Set the number of firings for each reaction accordingly:

   If $\tau^{'} < \tau^{''}$, take $\tau = \tau^{'}$. No critical reactions will fire in this time step and the number of firings of the non-critical reactions is sampled from the Poisson distribution with mean $a_j(x)\tau$, where j is the index for a non-critical reaction, R_j.

   If $\tau^{''} \leq \tau^{'}$, take $\tau = \tau^{''}$. Randomly select the index of a critical reaction, j_c, by sampling from a distribution of indices of the critical reactions with point probabilities $a_j/a_0^c(x)$. Allow only the reaction with index j_c to fire once and ensure

all other critical reactions do not fire. For all non-critical reactions, generate the number of firings as a sample of the Poisson distribution with mean $a_j(x)\tau$.

6. Update the number of reactants. If there is a negative reactant number, reduce τ' by half and return to step 3. Otherwise complete the leap by updating $\tau \leftarrow t + \tau$ and return to step 1. Else, stop.

In the appendix, we present results of convergence analysis that we performed to determine the appropriate values of the tolerance parameter, ϵ for use in the algorithm such that the τ-leap method is accurate to the SSA.

3.4 Results of untreated tumour growth with stochasticity

In this subsection, we address research question (i), given in the introduction of this chapter; in an untreated tumour, how does variation (within a set of stochastic repeats) in cellular dynamics over time and in levels of pre-existing resistance at diagnosis, differ when different values of the mutation probability, μ, and the relative growth rates of the resistant cells (to sensitive cells), δ, are employed in the model? For each set of μ and δ values, we simulate 7000 tumours, in the absence of drugs. Each stochastic process continues for a period of 25 years from the time of diagnosis.

3.4.1 Typical stochastic simulations of untreated tumour growth

To illustrate the behaviour of the stochastic model, in Figure 3.4.1, we plot 10 realisations of the stochastic model, describing untreated tumour cell populations over time, for a range of μ and δ values. We also include the solutions of the deterministic model (see chapter 2) in which we employ the same parameter values as in the stochastic model.

For solutions of the ODE model (thick coloured lines with black borders) presented in Figure 3.4.1, S cells (blue lines) grow initially followed by the growth of R_1 (orange lines) and R_2 (green lines) cell populations due to mutations taking place into these cell compartments from the S compartment. Mutations take place from the R_1 and R_2 cell populations to the R_{12} (pink lines) compartment and we observe an increase in the R_{12} population. Solutions of cell populations simulated using the stochastic model are shown as thin coloured lines in Figure 3.4.1. For the stochastic and ODE model simulations, we

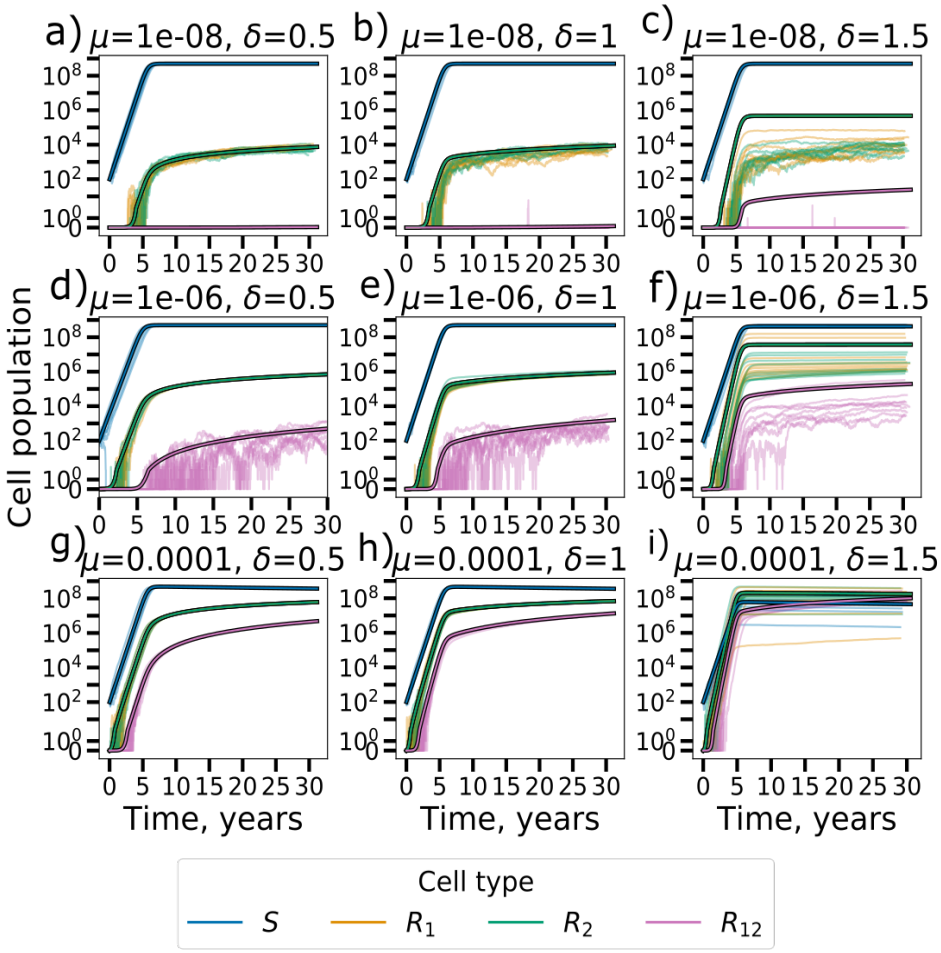


Figure 3.4.1: Typical stochastic simulations (10 realisations) and the deterministic model solution describing untreated tumour cell dynamics for a range of values of μ and δ. Thick lines with black borders represent the solution of the ODE. Thin lines represent stochastic realisations. The y-scale is plotted using the symmetric logarithmic scale which is logarithmic for cell populations with numbers greater than 1 and linear for cell populations less than 1. Hence, populations with no cells are plotted. Key: blue represents S cells, orange; R_1 cells, green; R_2 cells and pink; R_{12} cells. Parameters: all model parameters are fixed at the default values stated in Table 3.2.2, with the exception of μ and δ; their values are indicated on each subplot.

observe qualitatively similar solutions of cell population over time whereby, for $\mu > 10^{-8}$, we observe an increase in the S population followed by an increase in the R_1 and R_2 populations and, finally, an increase in the R_{12}. At long times, the S population is the largest followed by the R_1 and R_2 populations, and finally, the R_{12} population.

In Figure 3.4.1(f), where $\mu = 10^{-6}$ and $\delta = 1.5$, we observe large variation across stochastic solutions. We observe large variations in R_1 (orange thin lines) and R_2 (green thin lines) cell populations over time, across the 10 stochastic realisations and we observe

large variation in these cell numbers about the ODE solutions (thick lines with black borders). At time $t = 5$ years, the variation in the number of R_1 cells across the 10 realisations spans the range from approximately 10^5 to 10^8 cells. For $\delta \leq 1$ however (Figures 3.4.1(d)-(e)), the variation in the number of R_1 and R_2 cells is less pronounced. Taken together, these results suggest that higher growth rates of resistant cells give rise to larger variation in cellular dynamics across stochastic realisations.

In Figure 3.4.1(d), for $\mu = 10^{-6}$ and $\delta = 0.5$, solutions of the stochastic model show the S cells to become extinct within a year of tumour initiation, for one realisation. Further, in Figures 3.4.1(d)-(e), where $\mu = 10^{-6}$ and $\delta \leq 1$, fluctuations in R_{12} numbers that are simulated from the stochastic model persist for $t < 20$ years across the 10 realisations due to random mutations and extinction events.

In Figures 3.4.1(c) and (f), where $\delta = 1.5$ and $\mu \leq 10^{-6}$, the ODE model (thick lines with black borders) predicts higher numbers of resistant cell populations (the R_1, R_2 and R_{12} cell populations) than the majority of solutions generated using the stochastic model (thin coloured lines). For example, in Figure 3.4.1(f), at $t = 10$ years, the ODE model predicts an R_{12} population (pink thick line with black border) of 10^5 cells, whereas the stochastic realisations of the R_{12} population (pink thin lines) generated by the stochastic model present cell population sizes between 10 and 10^4 cells.

In Figure 3.4.1, we observe, as values of μ and δ vary, large differences in the size of the cell populations, generated from the ODE and stochastic models. In particular, differences in the size of the R_{12} population vary considerably for $\mu = 10^{-8}$ and $\mu = 10^{-6}$. We attribute this to the size of the R_1 and R_2 populations whereby larger populations (as a result of higher mutation rates) will result in an increased mutational flux into the R_{12} compartment. Based on these results, we can speculate that there is a critical value of μ that will result in fully drug-resistant cells being present in the tumour and we speculate that this critical value will depend on the carrying capacity. For example, in Figure 3.4.1, we assume a carrying capacity of $K = 5 \times 10^8$ cells. For $\mu = 10^{-8}$, when the S cells have reached sizes close to carrying capacity, a higher rate of mutational flux into the R_1 and R_2 compartments (and hence, into the R_{12} compartment) will occur. Conversely, if we were to assume a carrying capacity of $K = 10^6$ cells and a probability

of mutation, $\mu = 10^{-8}$, when S is large and close to carrying capacity, we would expect far fewer R_1 and R_2 cells and possibly no R_{12} cells. Hence, we speculate that a critical value of μ, from which R_{12} cells are present in the tumour, will greatly depend on the probability of mutation and the maximum size the tumour can reach.

For changes in δ, in Figure 3.4.1, for the ODE and stochastic model solutions we observe similar cellular dynamics for $\delta = 0.5$ and $\delta = 1$ compared to when $\delta = 1.5$. For both ODE and stochastic models, larger values of R_{12} cells are present when $\delta = 1.5$ compared to when $\delta = 0.5$ and $\delta = 1$. Based on these results, we could speculate that $\delta = 1$ is a critical value whereby we observe a noticeable increase in the R_{12} population for $\delta > 1$ relative to when $\delta \leq 1$. However, with respect to the number of R_{12} cells, changes in the value of μ has a more substantial impact on the response of the R_{12} cells whereby we observe larger changes in the size of the R_{12} population as μ varies relative to when δ is varied.

3.4.2 Quantifying variability in tumour cell dynamics

For a given set of values of μ and δ, we quantify the variability in the tumour dynamics by calculating,

$$v_u = 100 \times \frac{\text{standard deviation}}{\text{median}}, \tag{3.4.1}$$

across a set of 7000 stochastic simulations, at every 20 days. The results are presented in Figure 3.4.2. We do not plot values at which the median cell number is equal to zero.

In Figure 3.4.2, for $\mu \leq 10^{-6}$ (top and middle row in Figure 3.4.2), the spread in S, R_1 and R_2 cells is calculated for different values of δ (growth rate of resistant cells to sensitive cells). We observe that, at time $t = 10$ years, for low mutation probabilities ($\mu = 10^{-8}$), variation in numbers of R_1 and R_2 cells span 3 orders of magnitude: for $\delta = 0.5$, $v_u \approx 10$ whereas, for $\delta = 1.5$, $v_u \approx 10^4$. As μ increases, variation in R_1, R_2 and R_{12} decreases. We attribute this to an increase in the mutational flux for higher values of μ. This in turn, enables the resistant cells to increase to sizes beyond which they would be affected by random fluctuations.

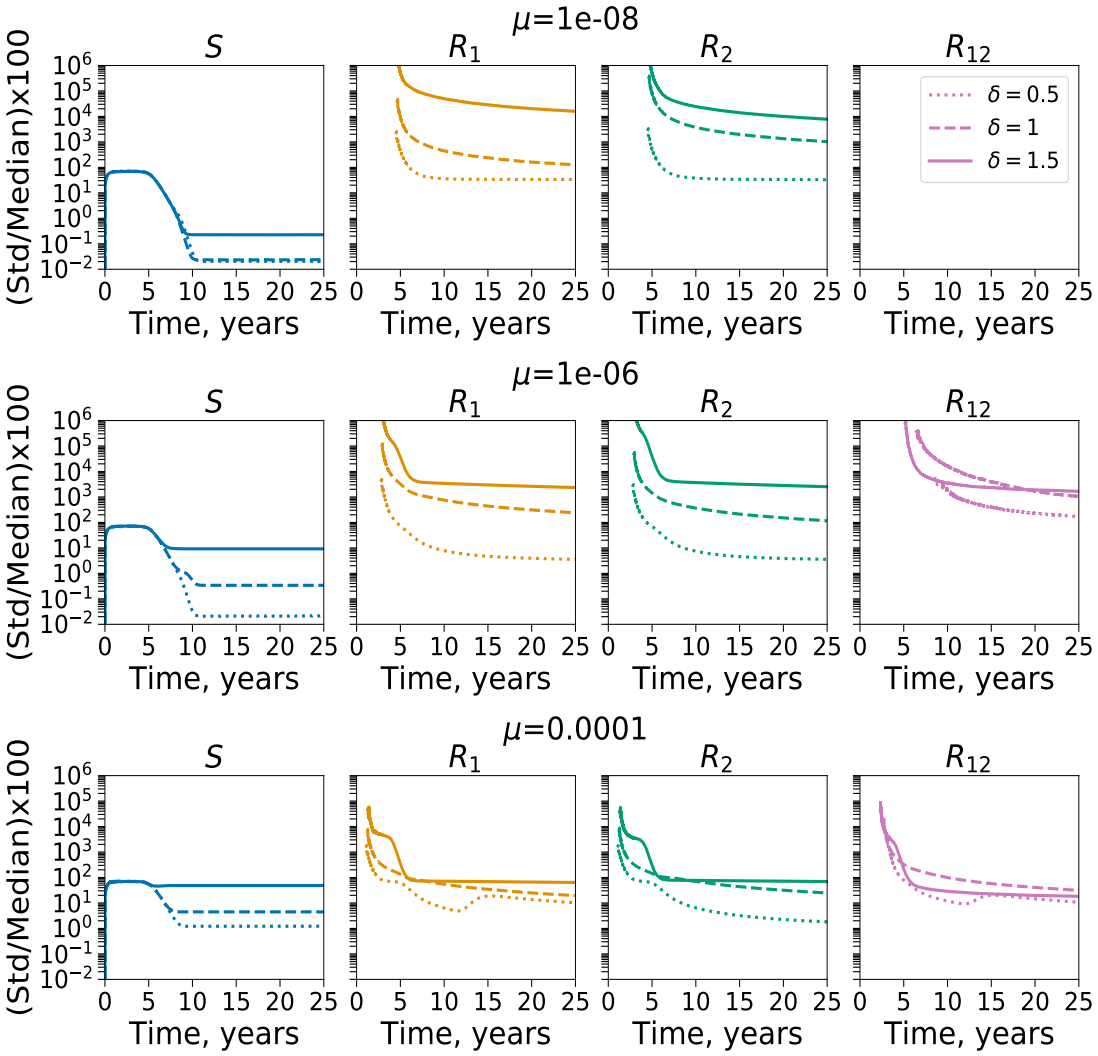


Figure 3.4.2: The standard deviation about the median (×100) of 7000 stochastic simulations plotted as a percentage for values of μ and δ. Each row represents a different value of μ ($\mu = 10^{-8}$ in top row). For each subplot, values of $\delta = 0.5$, $\delta = 1$ and $\delta = 1.5$ are presented as dotted, dashed and block lines, respectively. Parameters: for non-varied model parameters, default values are given in Table 3.2.2 are employed. Ranges of values of μ and δ are also given in Table 3.2.2. For $\mu = 10^{-8}$ and cell line R_{12}, no values are plotted as the median remained zero for all time.

Our results suggest that, when mutations are sporadic (low values of μ) and resistant cells have growth rates larger than sensitive cells (high values of δ), resistant cells are more likely to increase to where they are less sensitive to random fluctuations. Conversely, resistant cells with smaller growth rates are less likely to reach such sizes. Recalling research question (i), we conclude that large variation in untreated cell dynamics is observed for low values of μ and large values of δ. We deduce further that, in these

parameter regimes, the stochastic simulations will differ markedly from the solution of the deterministic model.

3.4.3 Quantifying variability in resistance at diagnosis

We continue to address research questions (i) and (iii) and, using our stochastic model, examine how changes in values of μ and δ impact variability in the proportion of fully drug-resistant cells in a tumour at diagnosis. For the stochastic and ODE models, for range of values of δ and μ, we simulate tumours until diagnosis (10^8 cells). At diagnosis, we record the mean, median, standard deviation/median of the proportion of R_{12} cells (denoted by $p_{R_{12}}$) across those of the 7000 stochastic realisations that have reached diagnosis (we note that less than 1% of tumours became extinct prior to diagnosis). We calculate $p_{R_{12}}$ values, predicted by the ODE model, for the same range of values of μ and δ, by simulating the ODE model until the total tumour population reaches 10^8 cells. We also estimate the probability of R_{12} cells being present at diagnosis, by calculating the proportion of stochastic realisations which contain at least one R_{12} cell. Our results are presented in Figure 3.4.4. We also plot histograms of $p_{R_{12}}$ values at diagnosis for μ in $[10^{-8}, 10^{-6}, 10^{-4}]$ and $\delta = 1.5$, in Figure 3.4.3.

We regard the median as a more informative measure than the mean, when interpreting values of $p_{R_{12}}$ due to the following reasons. If a large number of realisations have no R_{12} cells at diagnosis (see Figure 3.4.3(a)), then the mean value of $p_{R_{12}}$ will be skewed by a small selection of realisations with high values of $p_{R_{12}}$, relative to the remaining realisations. Figure 3.4.3 illustrates the skewed distributions of values of $p_{R_{12}}$ at diagnosis.

The results presented in Figure 3.4.4(a) show that at least one R_{12} cell will be present at diagnosis when $\mu = 10^{-4}$, for all values of δ. Conversely, for $\mu = 10^{-8}$ and $\delta < 1.5$, no stochastic simulations contain R_{12} cells. In Figure 3.4.4(b), for $\mu = 10^{-6}$ and $\delta = 1.5$, the ODE model predicts values of $p_{R_{12}} = 2.8 \times 10^{-5}$ and yet, only 56% of tumours generated using the stochastic model contain R_{12} cells (Figure 3.4.4(a)). Whether or not fully drug-resistant cells are present could have a significant effect on a tumour's response to treatment. Recalling research question (iii), we conclude that using an ODE model to predict such levels of resistance can be misleading. Especially when the

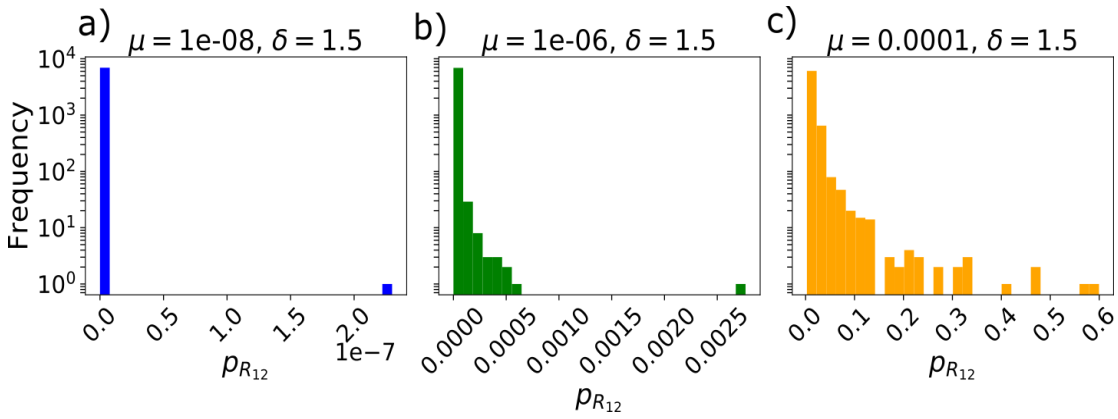


Figure 3.4.3: Histograms of values of $p_{R_{12}}$ at diagnosis for values of μ in $[10^{-8}, 10^{-6}, 10^{-4}]$ and $\delta = 1.5$.

relative growth rates of the drug-resistant populations are higher than the drug-sensitive populations.

In Figure 3.4.4(c), the median values of $p_{R_{12}}$ are presented for a range of values of μ and δ. We observe that, the median value of $p_{R_{12}}$ increases as the relative growth rate of the resistant cells (δ) and the mutation probability (μ) increases, attaining its maximum value of $p_{R_{12}} = 0.0123$ when $\mu = 10^{-4}$ and $\delta = 1.5$. These findings corroborate those presented in chapter two (see Figure 2.4.3).

In Figure 3.4.4(d), for high mutation probabilities ($\mu = 10^{-4}$), the spread (standard deviation/median) in $p_{R_{12}}$ values about the median increases as δ decreases, changing by an order of magnitude as δ decreases from $\delta = 1.5$ to $\delta = 0.5$. We observe considerable spread for $\mu = 10^{-6}$ and $\delta = 1.5$ with a value of 3928. We speculate however that, values of (standard deviation/median) are heavily skewed when the median is small and there exist realisations with large $p_{R_{12}}$ (see Figure 3.4.3(a)). We address this by calculating the quartile coefficient of variation, $QVC = (Q3 - Q1)/(Q3 + Q1)$ where $Q1$ and $Q3$ denote the lower and upper quartiles of a given distribution of $p_{R_{12}}$ values (see Figure 3.4.4(f)). We employ QVC as it is not influenced by values of $p_{R_{12}}$ that lie beyond the interquartile range. In Figure 3.4.4(f), we observe large dispersion values of ($QVC = 0.058$) for $\delta = 1.5$ and lower dispersion values of ($QVC = 0.036$) for $\delta = 1$ and ($QVC = 0.035$) for $\delta = 0.5$. Conversely, we observe increasing values of standard deviation/median for decreasing values of δ and $\mu = 10^{-4}$ (see Figure 3.4.4(d)) because it uses $p_{R_{12}}$ values that lie outside the interquartile range. For all parameter values, such

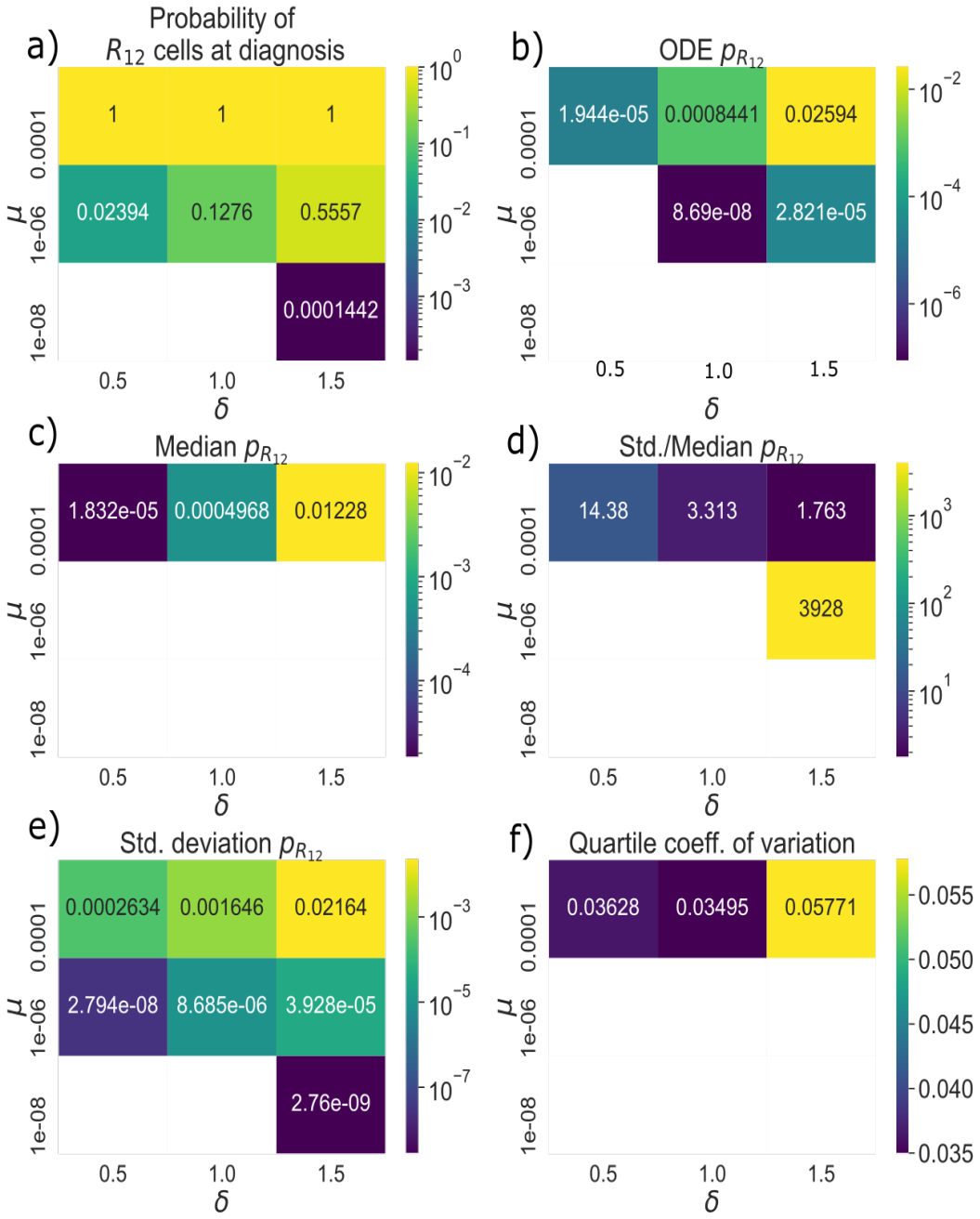


Figure 3.4.4: (a) Probability of R_{12} cells present at diagnosis across simulations in which diagnosis occurs for a range of μ and δ values. (b) Predicted $p_{R_{12}}$ values generated using the ODE model. (e)-(f) Statistics of the proportions of pre-existing fully drug-resistant cells ($p_{R_{12}}$) for a set of stochastic realisations; (c) median and (d) standard deviation/median, (e) standard deviation, (f) Quartile coefficient of variation, $QVC = (Q3-Q1)/(Q3+Q1)$ where $Q1$ and $Q3$ denote the upper and lower quartiles of a given distribution of $p_{R_{12}}$ values. Blank squares correspond to probabilities equal to zero and means and median values equal to zero. In (a), white squares indicate cases where the probability is equal to zero. In (b), white squares indicate cases where the value of the ODE is less than 10^{-8} (i.e. no R_{12} cells present). In (c)-(d), white squares indicate cases where the median is equal to zero.

deviations presented in Figure 3.4.4(d) and (f) are large enough to highlight the ODE model is not appropriate when predicting resistance proportions within a tumour.

To summarise, through examination of the model, we are able to answer research question (i); we have found that variability in $p_{R_{12}}$ values at diagnosis increases as δ increases and μ decreases.

In response to research question (iii) (what regions of the parameter space do the stochastics deviate from the ODE solution), we have demonstrated that an ODE system may not always be appropriate for predicting pre-existing resistance proportions; in certain regions of parameter space, it always predicts resistance, whereas, when stochastic effects are considered, only a selection of tumours may have doubly resistant populations. At this stage, we can only hypothesise that the ODE is not a suitable choice of model relative to the stochastic model as predictions made by our models are not validated. To know which model is appropriate would require validation of both the ODE and stochastic models. Such validation would require high-resolution data describing the number of cells in populations within a tumour.

Our results illustrate how tumour-specific properties, when coupled with stochastic effects, can have profound effects on a tumour's composition and it's growth dynamics. In the next section, we investigate how the mutation probability, μ affects the variability in treatment outcomes.

3.5 Results of treatment response with stochasticity

In this section, we address research question (ii): how does variation (within a set of stochastic repeats) in treatment response differ for different mutation probabilities? We generate 2623 stochastic realisations for each value of μ in $[10^{-8}, 10^{-6}, 10^{-4}]$. We fix $\delta = 1$. For simplicity, we focus on varying μ only, because values of the probability of mutation per cell division, reported in the literature, vary greatly [99, 48]. For each stochastic realisation, when the tumour reaches diagnosis (10^8 cells), we apply first-line treatment via protocol A. If progression occurs during first-line treatment, then we switch to protocol B. Treatment is applied for a period of 25 years from the time of diagnosis.

3.5.1 Typical stochastic simulations of treated tumour cell dynamics

For each value of μ, we randomly select 10 realisations to illustrate the range of dynamics that occur during treatment when stochastic effects are considered. These realisations are presented in Figure 3.5.1. The solution of the deterministic model is also included, to facilitate comparison of the stochastic and deterministic models.

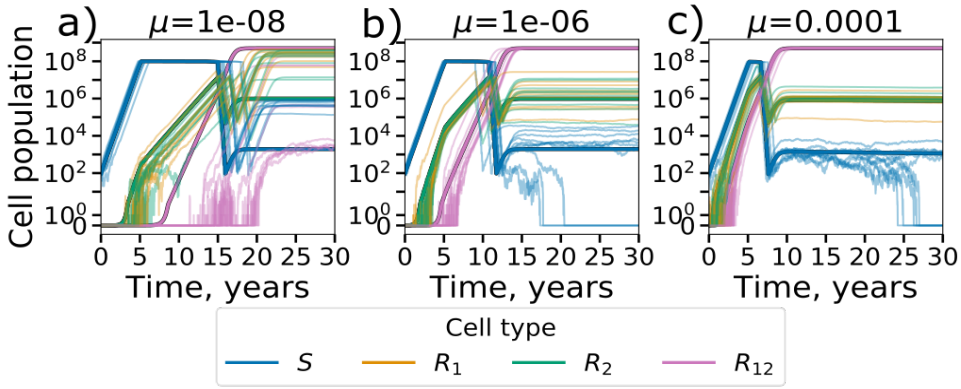


Figure 3.5.1: A selection of 10 stochastic simulations and ODE solutions of treated tumour growth dynamics for 3 values of the mutation probability, μ. Treatment with protocol A commences at diagnosis and is followed by protocol B when progression occurs on protocol A. Key: Blue represents S cells, orange; R_1 cells, green; R_2 cells and pink; R_{12} cells. For a given cell line, block thick lines represent the solution of the deterministic model (see chapter 2). Thin lines represent a stochastic simulations. Parameters: for non-varied model parameters, default values are given in Table 3.2.2 are employed. Ranges of values of μ are also given in Table 3.2.2 and values of μ are indicated in subplot titles.

In Figure 3.5.1, variation in the number of R_1, R_2 and R_{12} cells over time increases as μ decreases. This is because increased randomness in mutation events leads to larger variation in tumour cell dynamics (as seen in section 3.4.2), which, in turn, increases the variability in the times to progression during first-line treatment. In particular, when $\mu = 10^{-8}$, in Figure 3.5.1(a), the proportion of the R_{12} cells predicted by the ODE model is considerably larger than that observed for the stochastic realisations at early times.

In Figure 3.5.1(b), for $\mu = 10^{-6}$, the S population becomes extinct at approximately $t = 17$ years, after the tumour has undergone second-line treatment. This is because the high dose, second-line treatment significantly reduces the size of the S population (10^2 cells at the end of second-line treatment), making them vulnerable to random

fluctuations; a real world phenomenon that is not captured by a deterministic model.

3.5.2 First and second-line treatment response

For each value of μ, we simulate 2623 stochastic simulations. For each realisation, we record the progression times on first (protocol A) and second-line treatment (protocol B) and the proportions of R_{12} cells at these times. Results are presented as box plots in Figure 3.5.2. We also include progression times and $p_{R_{12}}$ values at progression for untreated tumours. For comparison, we include the progression times and $p_{R_{12}}$ values at progression generated from the deterministic model.

When simulating the stochastic model, the computation time for a single simulation varied for different values of μ. For example, for $\mu = 10^{-8}$, 7213 simulations were completed, for $\mu = 10^{-6}$, 3455 were completed and for $\mu = 10^{-4}$, 2623 simulations were completed. We attribute this to the stochastic algorithm itself whereby the τ value is calculated based on the relative sizes of the cell populations in the tumour. Changes in μ will result in different cell population sizes in the tumour and the step sizes will therefore depend on the value of μ. We select 2623 as the number of simulations as this is the maximum number of completed simulations for $\mu = 10^{-4}$ and we require the same number of simulations when analysing the differences in responses as μ varies.

In Figures 3.5.2(b) and (c), we observe larger variation in progression times on first and second-line treatment across stochastic simulations as μ decreases. We also observe larger differences between the ODE and the stochastic model, which partially answers research questions (i) and (iii). We ascribe the variation in progression times to the increased variation in values of $p_{R_{12}}$ at progression on first and second-line treatment, which we investigate further below using Figures 3.5.3 and 3.5.4. In particular, for lower values of μ, the corresponding values of $p_{R_{12}}$ from the stochastic model are much smaller than those from the ODE model.

In Figure 3.5.3, we present a 2D plot showing, for a fixed value of μ, the relationship between the values of $p_{R_{12}}$ at diagnosis and the time to progression on first-line treatment for each of the 2623 stochastic realisations.

In Figure 3.5.3, for $\mu = 10^{-8}$, there are no data points as none of the tumours contained R_{12} cells at diagnosis. For $\mu = 10^{-4}$, quicker times to progression correlate

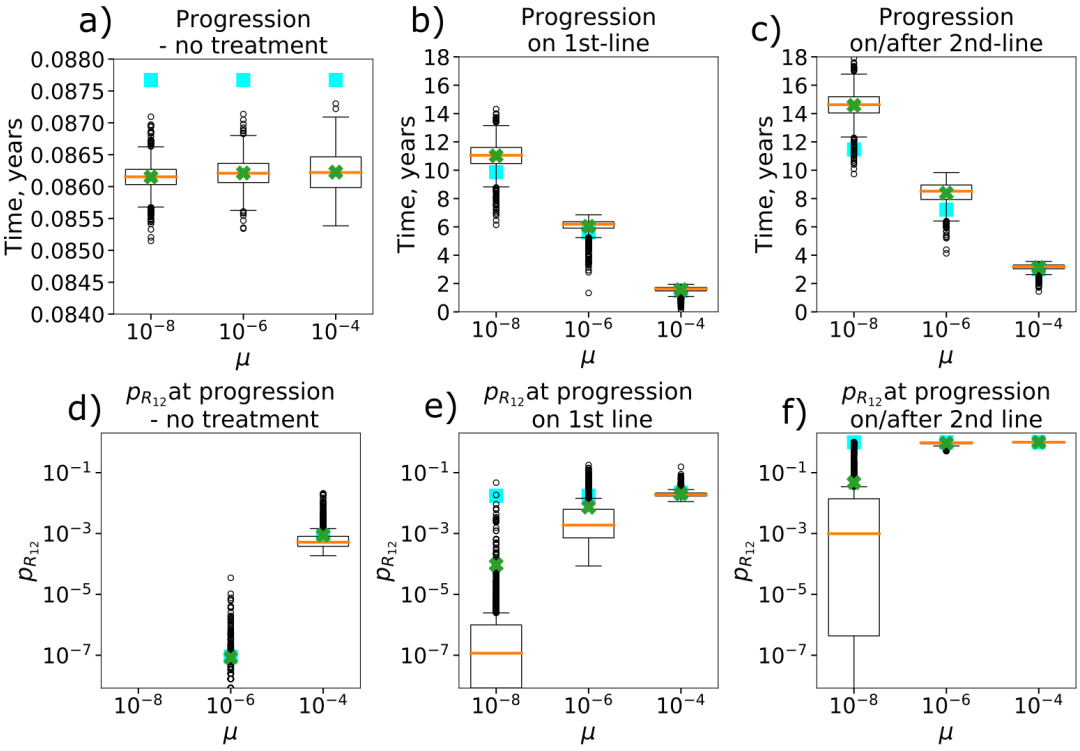


Figure 3.5.2: Box plots of progression times (from diagnosis) and $p_{R_{12}}$ values at progression on first and second-line treatment. Box plots: orange lines represent the median. The box indicates the interquartile range (IQR) ranging from the lower quartile (25th percentile) to the upper (75th percentile) quartile. The bars indicate 1.5× the upper and lower quartiles and black circles indicate values greater or less 1.5× the upper and lower quartiles. Green crosses indicate the mean of the distribution of values recorded from 2623 stochastic simulations. We include progression times and $p_{R_{12}}$ values for untreated tumours generated by the ODE and present these predicted values using cyan squares. Parameters: for non-varied parameters, default model parameters presented in Table 3.2.2 are employed.

with larger values of $p_{R_{12}}$ at diagnosis. For $\mu = 10^{-6}$, the negative correlation between $p_{R_{12}}$ at diagnosis and time to progression is weaker.

In chapter 2, we speculated that tumours with higher levels of resistance at progression on first-line treatment will have reduced response to second-line treatment. We investigate this in Figure 3.5.4 where, for each stochastic realisation, we explore the correlation between the time to progression, on, or after, second-line treatment from the start of second-line treatment and the value of $p_{R_{12}}$ at progression on first-line treatment. For $\mu \geq 10^{-6}$, we observe a strong negative correlation. For $\mu = 10^{-8}$ and values of $p_{R_{12}} \leq 10^{-5}$, we observe progression times of approximately 3.5 years from the start of second-line treatment which suggests that R_{12} cells at low levels at the start of

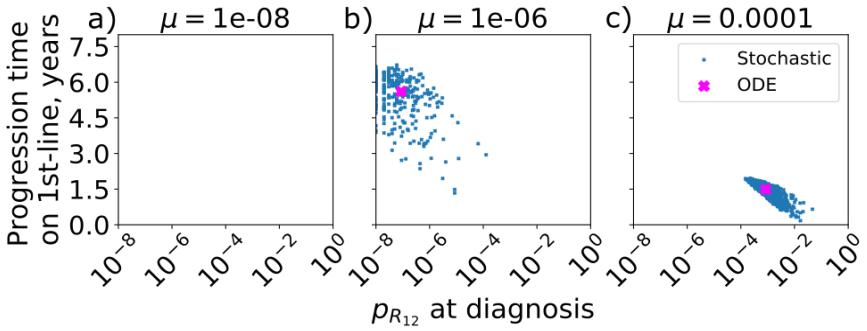


Figure 3.5.3: Progression times on first-line treatment (from the time of diagnosis) versus the proportion of R_{12} values at diagnosis from 2623 stochastic simulations. Blue crosses indicate results from the stochastic framework. Magenta crosses indicate the solution from the ODE. Parameters: for non-varied parameters, default model parameters presented in Table 3.2.2 are employed.

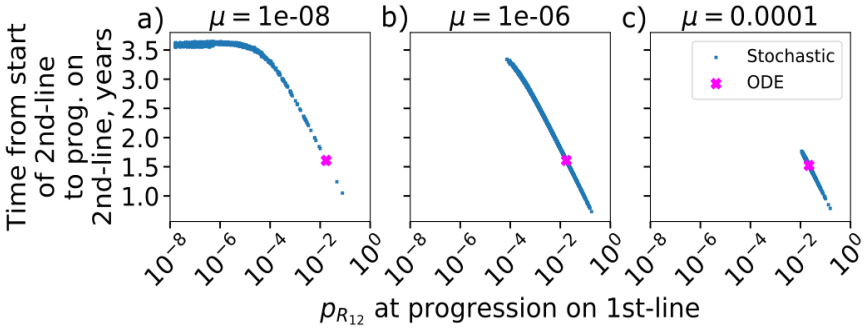


Figure 3.5.4: The duration of time from the start of second-line treatment to progression on, or after second-line treatment versus the proportion of R_{12} values at the start of second-line treatment, generated from 2623 stochastic simulations. Blue crosses indicate results from the stochastic framework. Magenta crosses indicate the solution from the ODE. Parameters: for non-varied parameters, default model parameters presented in Table 3.2.2 are employed.

second-line treatment do not greatly affect the response to second-line treatment. For $\mu = 10^{-8}$, if values of $p_{R_{12}} \geq 10^{-5}$ at progression on first-line treatment, we observe a strong negative correlation between $p_{R_{12}}$ at progression on first-line treatment and time to progression on second-line treatment. For all values of μ, there is a noticeable deviation between the ODE value and the stochastic values, however, markedly different results are shown for values of $\mu \leq 10^{-6}$.

To summarise, with regard research question (ii), we conclude that the mutation rate has a considerable effect on the variability in response across stochastic realisations, with larger variability observed for smaller values of μ. In reference to research question (iii), we conclude that, due to small variability in stochastic realisations and good agreement

with progression times and proportions of R_{12} cells at progression on first and second-line treatment, the ODE provides a good approximation to the stochastic model for values of $\mu = 10^{-4}$. However, for smaller values of μ, the increased variability in treatment response and large discrepancies between the ODE and the stochastic simulations, render the ODE less reliable and the stochastic model more informative.

3.6 Discussion

In this chapter we extended our analysis of NSCLC tumour growth dynamics and the effects of combination TKI treatment by considering the impact of stochasticity. We developed a stochastic model based on the same assumptions as those used for the ODE model. We posed the motivational questions at the start of this chapter and restate them here.

i In an untreated tumour, how does variation (within a set of stochastic repeats) in cellular dynamics over time, and in levels of pre-existing resistance at diagnosis, differ when different values of the mutation probability and the relative growth rates of the resistant cells (to sensitive cells), are employed in the model?

ii How does variation (within a set of stochastic repeats) in treatment response differ for different mutation probabilities?

iii In what regions of the parameter space do outcomes from the stochastic model deviate from those of the deterministic model, developed in chapter 2?

The simulations and analyses performed in this chapter have informed our answers to these questions. We summarise below, our findings along with our answers to these questions.

i Larger variation in cellular dynamics and resistant proportions at diagnosis were observed in tumours with lower propensities to mutate, and, whose resistant cells had higher growth rates relative to the sensitive cells.

ii Changes in μ were shown to greatly influence the variability of progression times and, $p_{R_{12}}$ values at progression on first and second-line treatment, within a set of stochastic simulations.

iii For the untreated and treated cases, we conclude that the ODE is less informative when making modelling predictions for smaller values of μ, as larger variation in pre-existing resistance levels and treatment response was shown.

A large range of mutation probabilities are presented in the literature with values of $\mu \leq 10^{-6}$ [48]. Therefore, more studies considering the effects of stochastics when modelling tumour growth with mutation and response to TKI treatment, would be informative.

We simulated our model using the τ-leap method with adaptive step size calculation [17]. A drawback of this method is the computational cost when simulating tumours with vastly different cell population sizes. However, a benefit of this method, is that physiologically motivated parameter values can be used and tumours with realistic cell numbers can be simulated, which, to our knowledge, in a stochastic setting, has not been implemented before. Therefore, we can directly compare the ODE and stochastic models in which, clinically relevant parameter regimes are employed. Hence, the research carried out in this chapter is of use when selecting an appropriate modelling framework to investigate the response of NSCLC tumour to TKIs treatment.

A recent study presented by Hayford et al. uses in vitro models of tumour heterogeneity and stochastic simulations to understand the extent to which genetic, epigenetic and stochastic effects influence tumour heterogeneity [56]. One conclusion reached was that variability observed in a single cell line could be explained by intrinsic randomness in cell division and death. Therefore, this study demonstrates that stochasticity can drive tumour heterogeneity. In the clinic, diverse outcomes to TKI treatment in EGFR-mutant NSCLC is observed [24]. However, determining the extent to which stochasticity impacts variability in tumour composition and treatment response across patients in the clinic is a difficult task and would require extensive data of tumour composition over time. Gathering biopsies is invasive and such data is not readily available.

Naturally, a broad range of treatments, applying TKIs in combination, can be considered. Due to the symmetric nature of our model, investigations into the alternation of two TKIs would not only be feasible, but relevant as such clinical studies are ongoing [119]. In the previous chapter, for a given set of model parameters employed, the dynamics of R_1 and R_2 cells are the same. When stochastic effects are considered, this is no longer the case. This loss of symmetry may contribute to increased variation in response to treatment schedules in which TKIs are applied in an alternating fashion. This is subject to future work and such treatment schedules are not considered in this thesis.

To summarise, in this chapter we have identified regions of the parameter space in which stochasticity plays a large role in tumour cell dynamics and treatment response. Our modelling predictions have shown that a stochastic model may be more informative, demonstrating variation in treatment response that should not be overlooked. Based on our predictions made in this chapter, we propose ODE models, used to make predictions on treatment response and the emergence of drug resistance, should be applied with caution and qualified by a consideration of stochastic models. However, validation of both models using tumour cell data is required in order to make finite conclusions on model suitability when modelling the emergence of drug resistance and treatment response. In particular, this emphasises that the desired future transition to modelling supporting decision making and personalised medicine in this area, requires the more difficult task of parametrising and validating stochastic models.

In the next chapter, we move away from theoretical modelling frameworks and turn to in vitro and in vivo preclinical data, describing NSCLC tumour growth in the absence of TKIs. The data introduced in the following chapter provides a physical example of the studies we have considered in chapters 2 and 3. We then proceed to develop models with the appropriate level of complexity to represent the data. We also examine the suitability of the models of in vitro cell growth to describe tumour growth in vivo.

Appendix

3.A Convergence testing for identification of the optimal tolerance parameter, ϵ in the τ-leap method with adaptive τ calculation

In this appendix, we present results of convergence testing of the τ-leap approximation method, with adaptive τ calculation, which we use to simulate our stochastic model.

To determine our choice of tolerance parameter, ϵ, we compare distributions of stochastic realisations, generated using the τ-leap approximation method and, using the SSA. We consider values of $\epsilon \in \{0.003, 0.004, 0.005, 0.01, 0.02, 0.03\}$ with $n_c = 50$ (n_c denotes the critical reaction number (see section 3.3)). We use the two-sample Kolmogorov-Smirnov (K-S) test to compare the distributions generated by the SSA and the τ-leap method, for a fixed value of ϵ. We select a high value of $\epsilon = 0.03$ as this value is proposed by Cao et al. who presented the τ-leap approximation method used in this thesis [17]. However, in the study presented by Cao et al., they consider population sizes up to 10^4. In the main text of this thesis, we consider larger population sizes ($> 10^8$) with tumours that consist of multiple cell populations which vary greatly in size. The control parameter, ϵ bounds the calculated value of τ such that no time step is so large that a large change in a small population may occur, resulting in an accurate approximation of the SSA algorithm. As our tumours consist of vastly different cell subpopulation sizes, to ensure we minimise the error in the number of cells in a given cell population relative to the gold standard SSA, we consider lower values of $\epsilon < 0.03$ with a minimum of $\epsilon = 0.003$.

The K-S test is a non-parametric test, used to test whether or not two samples are drawn from the same distribution that is assumed to be continuous. For each time step, we compare the distribution of cell numbers, generated by the SSA and τ-leap

parameter	value	units
α	0.05	day^{-1}
β	1	dimensionless
δ	1	dimensionless
γ	1	dimensionless
σ	3	dimensionless
K	10^6	cells
μ	5×10^{-4}	dimensionless

Table 3.A.1: Parameter values employed in the model for convergence testing of the τ-leap algorithm. Descriptions of model parameters are given in Table 3.2.2.

approximation method, by calculating the K-S statistic which is defined as,

$$D_{n,m} = \sup_x \mid F_n(x) - G_m(x) \mid,$$

where $F_n(x)$ is the cumulative distribution of n samples (samples are cell numbers at a given time) generated using the SSA. $G_m(x)$ is the cumulative distribution of m samples (samples are cell numbers at a given time) that were generated using the τ-leap approximation method. *sup* is the supremum function.

The null hypothesis; stating that two samples are of the same distribution, is rejected at significance level $\hat{\alpha}$ if,

$$D_{n,m} > c(\hat{\alpha})\sqrt{(\frac{n+m}{nm})},$$

where n and m are the sizes of the first and second sample, respectively, and where

$$c(\hat{\alpha}) = \sqrt{(-\ln(\hat{\alpha}/2)/2)} = 1.628$$

is calculated from the inverse Kolmogorov distribution at significance level $\hat{\alpha} = 0.01$.

We use the SSA to generate 7000 simulations of our stochastic model. These simulations describe the growth of an untreated tumour for 150 days. We do the same using the τ-leap method for each value of ϵ. We employ model parameter values presented in Table 3.A.1.

For a given cell line, at each time point, we calculate the K-S statistic of the distributions of cell numbers. Therefore, we yield a set of K-S statistics for a range of time points and we present results in Figure 3.A.1.

In Figure 3.A.1, $\epsilon = 0.005$ (black block line) is the largest value of the tolerance parameter that yields K-S statistics that satisfy the null hypothesis for all cells and for

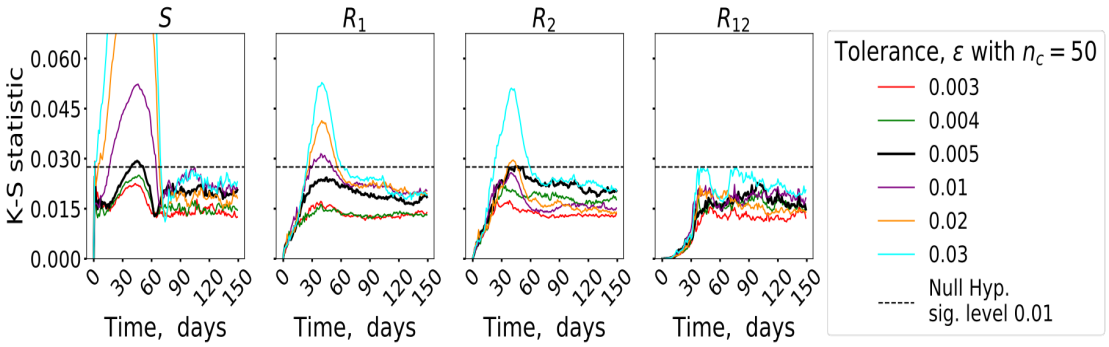


Figure 3.A.1: Kolmogorov-Smirnov (K-S) statistics over time for tumour cell populations. 7000 realisations of the SSA and τ-leap method are simulated where a range of values of ϵ are considered in the τ-leap method. Key: coloured lines; different values of ϵ, black horizontal dashed line; K-S statistic value corresponding to the null hypothesis.

all time. Hence, we employ control parameters, $\epsilon = 0.005$ and $n_c = 50$ in our the τ-leap algorithm when simulating our stochastic model. We note that we applied the same procedure for values of $n_c \in \{2, 10, 20\}$, and found results did not differ markedly to those presented in Figure 3.A.1.

In order to simulate the model using the SSA algorithm with reduced computation time, we have considered smaller tumour populations with lower carrying capacity, $K = 10^6$ (in the main text, $K = 5 \times 10^8$ (cells) is employed) and faster rates of cell growth ($\alpha = 0.05$ days^{-1} and $\sigma = 3$) than in the main text. We do not know whether the same tolerance parameter ($\epsilon = 0.005$) applies to those parameters that are employed in the model. However, Cao et al. suggested a tolerance of $\epsilon = 0.03$ in [17] whereas, we use a value that is more than an order of magnitude smaller. Further, the adaptive τ calculation step, in the algorithm, accounts for all tumour cell population sizes to ensure the leap condition is met.

Chapter 4

Modelling preclinical tumour growth data

4.1 Introduction

In this chapter, our aim is to develop a deterministic model describing the growth of untreated non-small cell lung cancer (NSCLC) cell lines. We consider in vitro tumour growth data for a single cell line, that demonstrates clear sigmoid growth dynamics and consists of large numbers of data points. Therefore, to model this data, we consider models that are simpler than those developed in chapters 2 and 3 in that they possess a level of complexity in line with the data and thus describe the growth of a single cell line with a small number of parameters. We use these models to distinguish the growth dynamics of four NSCLC cell lines by fitting the models to in vitro and in vivo data sets.

Mathematical modelling can be used to infer the underlying biological mechanisms in tumour growth and treatment response via the estimation of unique model parameters when fitting models to data. However, the ability to estimate such parameters for a given model and data set will depend on the following. Firstly, whether the model is structurally identifiable. A structurally identifiable model is one in which distinct sets of model parameters give rise to distinct model outputs in the best-case scenario in which we have an infinite amount of noise-free data [36]. Structural identifiability analysis is performed on a model prior to the consideration of the data set and is a useful and necessary step in determining the appropriate level of model complexity (i.e. number of identifiable model parameters), relative to the observables in the available data [125].

A structurally non-identifiable model is one in which a set of model parameters can be varied without changing the observed model output [115]. Practical identifiability addresses the scenario in which noise within a data set and the number of available data points affect our ability to estimate unique model parameters. Structural identifiability analysis is a prerequisite to practical identifiability analysis and, even for a structurally identifiable model, practical identifiability issues may arise, hindering our ability to estimate parameters [115].

Predictions made using models that contain practically non-identifiable parameters (i.e. parameters that can not be uniquely estimated to a sufficient level of accuracy given the available data), can be misleading [36, 115]. Two solutions to practical non-identifiability involve, (i) reduction in model complexity (e.g. fixing model parameters or reparameterisation of the model), and/or (ii) additional data measurements. We investigate the benefit of measuring additional data on increasing the identifiability of models in the following chapter. In this chapter, we employ identifiability analysis to carry out model selection where we consider a suite of increasingly complex models to describe the in vitro data. We then use the profile likelihood method to assess the practical identifiability of the model [115].

Information criteria and goodness-of-fit metrics are commonly used in the selection of models to describe a given data set [2]. When fitting a model to data, the Akaike information criterion (AIC) statistic of the fit rewards the model for its goodness-of-fit yet, penalises models with increased complexity so as to avoid over-fitting and uses a penalty term $2u$, where u is the number of parameters. An alternative information criterion statistic is the Bayesian information criterion (BIC) [122]. Analogous to AIC, BIC penalises models with higher complexity (number of model parameters) yet, also penalises models based on the amount of data the model is fitted to, whereby a larger penalty is incurred for larger amounts of data. BIC employs a penalty term, $u \ln n$, where n is the number of data points. AIC and BIC give a single statistic of the penalised goodness-of-fit of the models yet do not shed light on the underlying biological mechanisms within the system of interest (e.g. tumour cell growth).

In this chapter, we use parameter estimation, information criteria/goodness-of-fit

metrics and structural and practical identifiability approaches to carry out model selection from a set of five models that describe the growth of NSCLC cell lines in vitro. We aim to select a model with appropriate complexity, given the data and to estimate parameter values to compare the growth properties of different NSCLC cell lines.

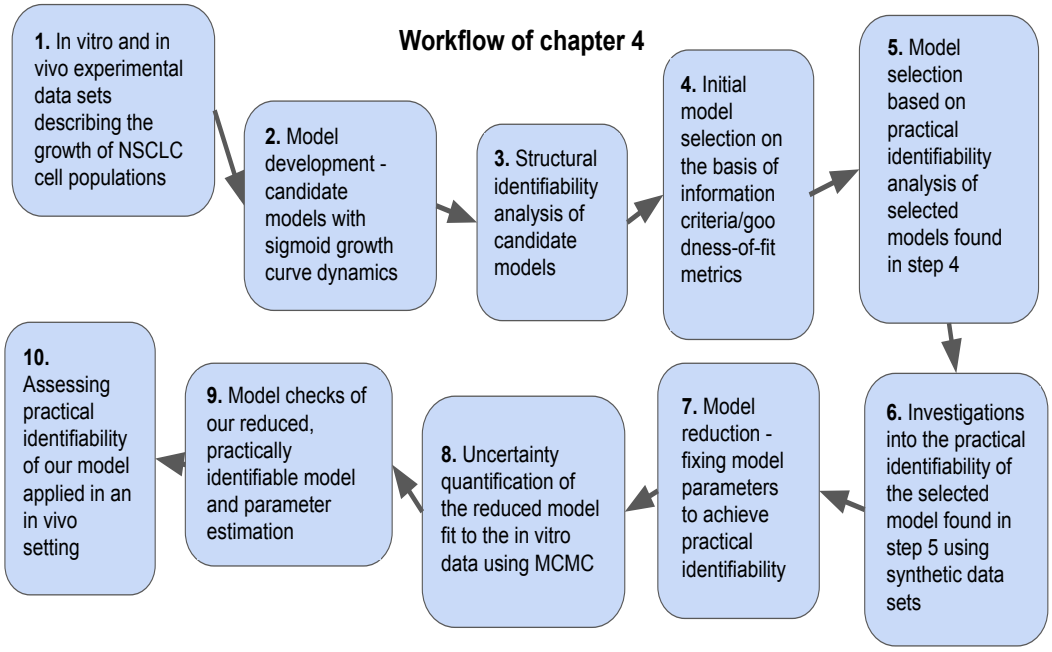


Figure 4.1.1: Schematic of the workflow in this chapter.

The workflow of this chapter is summarised in Figure 4.1.1 and we expand on the workflow here.

1. We begin by presenting experimental datasets describing the growth of NSCLC cell lines cultured in vitro and in vivo.

2. We present candidate models and perform model comparison by fitting the models to the in vitro experimental data and comparing information criteria/goodness-of-fit metrics.

3. We show all models to be structurally identifiable.

4. We select those two models which yield the lowest information criteria/goodness-of-fit metrics.

5. We uncover practical identifiability issues for both models (model parameters are practically non-identifiable given the in vitro data) and, in so doing, illustrate that information criteria/goodness-of-fit metrics alone may not be sufficient for model selection when one aims to infer the biological properties of the model parameters given the data. We proceed with the model which contains fewer practically non-identifiable parameters.

6. We generate synthetic datasets from the selected model to assess the impact of noise on the identifiability of model parameters and we find issues with practical identifiability.

7. We construct parameter likelihood profiles to understand the relationship between model parameters and the goodness-of-fit. This analysis guides our decision to fix the non-identifiable model parameters in order to achieve practical identifiability (i.e. all model parameters are practically identifiable).

8. We then assess the uncertainty of the model fit to the in vitro data using the Bayesian technique for parameter inference, Markov chain Monte Carlo (MCMC). We also assess the practical identifiability of the reduced model with fixed parameters using our MCMC results and we find results on practical identifiability corroborate the results from the profile likelihoods. By sampling from the posterior parameter distributions, we calculate the uncertainty of the model fit to the data.

9. We use the reduced model to compare the growth rates of four NSCLC cell lines in vitro and check the suitability of the model fit to the data by plotting residuals over time.

10. We apply our model to in vivo tumour volume growth curves and use profile likelihoods to compare the practical identifiability of our model in the in vivo and in vitro settings.

In section 1.2.1 in chapter 1, we discussed models that present sigmoid growth dynamics with increasing complexity (number of model parameters). Generalising these

models to describe in vitro NSCLC tumour growth curves represents the novel aspects of the modelling developed in this chapter.

The remainder of this chapter is structured as follows. In section 4.2, we describe the in vitro and in vivo datasets. Model development takes place in section 4.3 where we present five candidate tumour growth models. In section 4.4, we introduce the analytical methods that we use in this chapter. We present our results in section 4.5. The chapter finishes in section 4.6 where we discuss our findings and conclusions.

4.2 Experimental data

4.2.1 In vitro monoculture growth curves

Before we develop our mathematical models, we introduce the series of in vitro data sets which were generated by our industrial collaborators at AstraZeneca and, which motivate the work carried out in this chapter.

The in vitro datasets consist of growth curves of four epidermal growth factor receptor (EGFR) activating, NSCLC PC9 variant cell lines. PC9 is a human lung adenocarcinoma with an EGFR exon 19 deletion that can increase EGFR autophosphorylation and promote cell survival through the selection of alternative survival pathways, therefore driving the disease [135]. EGFR exon 19 deletions include many variants of the deleted amino acid [135].

The four cell lines used in the in vitro experiments differ in their sensitivity to tyrosine kinase inhibitors (TKIs), osimertinib and gefitinib. The parental cell line PC9 (with the EGFR exon 19 deletion mutation) is sensitive to gefitinib and osimertinib; the PC9VanR cell line carries the T790M mutation and an EGFR exon 19 deletion mutation (the T790M mutation confers resistance to gefitinib). The PC9VanR cell line was established in vitro from the PC9 cell line, by chronic exposure to vandetanib (tradename Caprelsa), a TKI applied in thyroid cancer [65].Using the genome editing tool, CRISPR, the PC9VanR cell line was modified to produce two mutants, both expressing the C797S mutation which confers resistance to osimertinib [114]. These mutant cell lines are labelled SCD4 and SCB5 and both have T790M and C797S mutations. SCD4 has two copies of the EGFR gene, one has the C797S mutation included, the other does

not. Hence, SCD4 cells express two versions of EGFR on the cell surface, one with T790M and C797S mutations, making them resistant to both drugs, and the other with the T790M mutation only, making them resistant to gefitinib. SCB5 cells have one type of EGFR with the T790M and C797S mutations. A summary of these four cell lines, their respective mutations and drug-sensitivities is given in Table 4.2.1. PC9VanR cells are referred to as T790M cells for the remainder of this chapter.

Cell line	Resistance mutations	Drug-sensitivity	Notation
PC9	–	gefitinib and osimertinib	PC9
PC9VanR	T790M	osimertinib	T790M
SCD4	T790M+C797S	–	SCD4
SCB5	T790M+C797S	–	SCB5

Table 4.2.1: Summary of the properties of the four NSCLC cell lines used to generate the in vitro datasets: cell lines and their respective resistance conferring mutations to two TKIs, gefitinib and osimertinib.

4.2.1.1 In vitro experiment protocol

1. Cells were cultured on a 384 well plate.

2. Cells were seeded onto new plates at $\sim 4\%$ confluence (4% of the area of the plate was covered by the cells at the start of the experiment).

3. The cells were cultured in an incucyte machine for a period of 6 days. The degree of confluence was measured every 4 hours and cells were grown to 100% confluence. Live and dead cells contribute to confluence; however, individual live and dead cell numbers are not recorded.

4. Each experiment was repeated three times.

We note that the in vitro data have not been published.

4.2.1.2 Data processing

For the purpose of studying the tumour cell dynamics, each data set (in units of confluence) was converted to cell numbers by assuming that 4% confluence corresponds to 2000 cells (personal communication, AstraZeneca). At 100% confluence, cells have

formed a monolayer, covering the entire plate. The carrying capacity, K, is defined to be the number of cells at 100% confluence.

Hence, we estimate

$$K = \frac{2000}{0.04} = 50000 \text{ cells.} \tag{4.2.1}$$

The initial number of cells, y_0, is estimated for each experiment,

$$y_0 = K \times \text{initial confluence.} \tag{4.2.2}$$

Cell proliferation assays typically span periods of up to 24 hours for the purpose of estimating cell proliferation rates at low cell densities [144]. It is less common that cells are cultured over large time periods such that growth is limited due to contact between the cells and the walls of the well plate.

Mathematical models with competition, for example the logistic growth curve, can be used to describe in vitro proliferation assays where contact inhibition takes place [70, 121, 126, 144].

4.2.1.3 Discussion on in vitro datasets

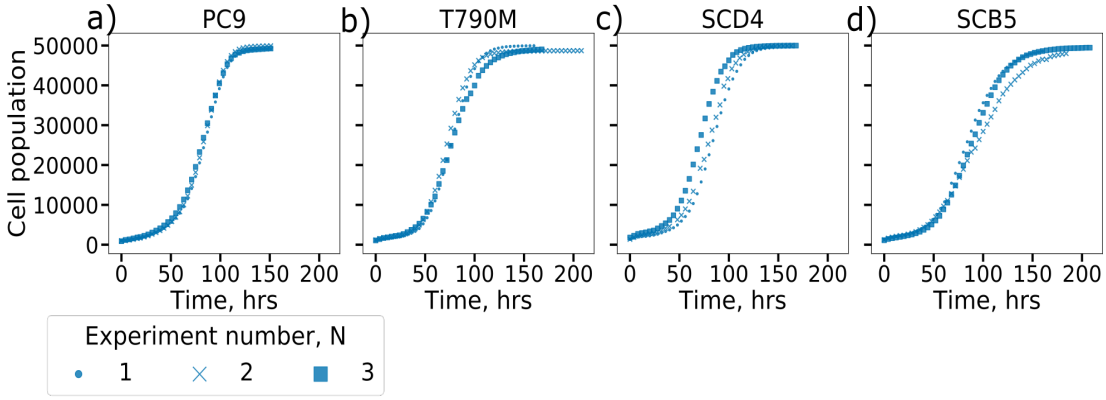


Figure 4.2.1: Plots of in vitro growth curves of the 4 cell lines: PC9, T790M, SCD4 and SCB5 (see Table 4.2.1). Marker shapes: circles, crosses and squares are representative of experiment repeats, $N = 1$, 2 and 3, respectively. The legend presents the marker shape corresponding to each experimental replicate.

In Figures 4.2.1(a)-(d) we present the data from the in vitro experiments. Across all four cell lines (see Table 4.2.1), the growth is approximately exponential during the first

75 hours and then slows to linear growth. After 100 hours, growth starts to saturate, as the cell populations reach their carrying capacity of 50,000 cells.

The results presented in Figures 4.2.1(a)-(d) highlight variability in the replicates for each cell line. We speculate that this may be due variation in the view of the well of the incucyte machine's microscope which may not capture the same region across repeats. Other sources of variability may be in cells themselves, fluctuations in culture media and variability in culture conditions.

4.2.2 In vivo mouse xenograft tumour volume growth curves

4.2.2.1 In vivo experiment protocol

Human tumour cells were transplanted into the flank of Nude (NuNu) immunocompromised mice. These mice were used because they do not reject the human cells. Tumours consisting of 1 million cells of a single type were implanted subcutaneously into the mice's flanks.

The same four cell lines were used as in the in vitro experiments: PC9 cells, T790M cells and a 1:1 mixture of SCD4 and SCB5 cells was considered. The volume of the tumours was measured after 216 hours from implantation to 648 hours for T790M tumours and to 720 hours for PC9 and 1:1 SCD4:SCB5 tumours. Volumes were measured at times, 336, 408, 480, 552, 648 hours for all tumours and at 720 hours for PC9 and 1:1 SCD4:SCB5 tumours. For each replicate, the same tumour was measured at each time point. Five experimental repeats were carried out for each cell type.

4.2.2.2 Discussion on in vivo datasets

The raw data presented in Figure 4.2.2(a) show how the in vivo tumour volume changes over time, for each cell line and each experimental repeat.

In Figure 4.2.2(b) we present, for a given cell line, the mean tumour volume averaged across experimental repeats at each time point and indicate standard deviations by bars. Figure 4.2.2(b), shows that at time, $t = 648$ hours, the mean volume of tumours initiated with a 1:1 mixture of SCD4 and SCB5 cells (green triangle) is approximately 100mm^3 smaller than the mean T790M (yellow star) tumour volume. The mean initial size of tumours comprising only PC9 cells (blue diamond) are the smallest. These tumours are,

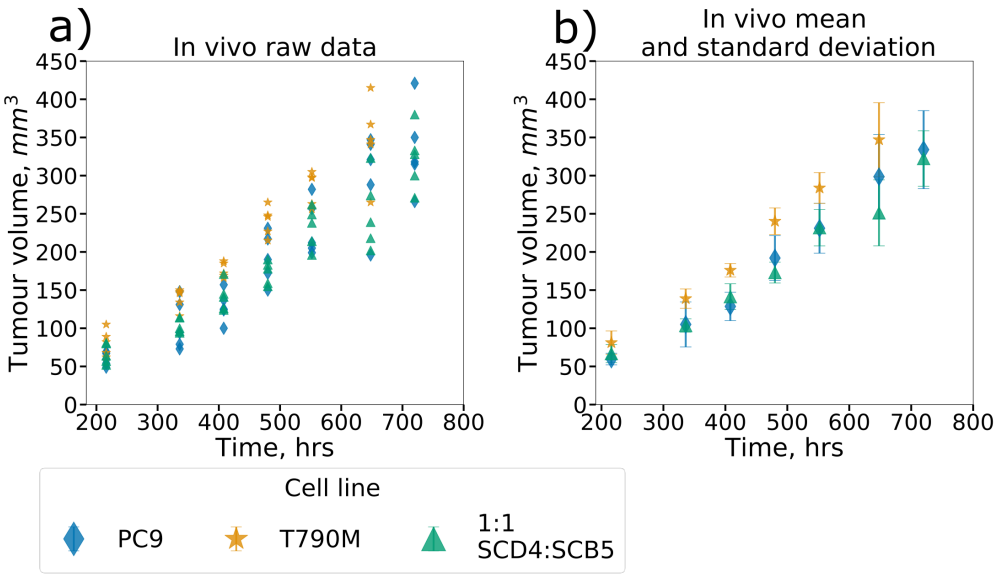


Figure 4.2.2: In vivo control experimental mouse xenograft tumour volume data (units of mm^3). a) Raw in vivo tumour volume data. Marker shapes: diamonds, stars and triangles represent PC9, T790M and 1:1 SCD4:SCB5 tumours, respectively. Five experimental repeats were carried out for each cell line. b) Mean in vivo tumour volume data for each cell line (maker shape) and standard deviation (bars) across five experimental repeats.

however, larger than the mean SCD4:SCB5 tumour volume (green triangles) at time $t = 648$ hours which suggests that tumours consisting of the SCD4 and SCB5 cells grow more slowly than those consisting of PC9 cells.

The results presented in Figures 4.2.2(a)-(b) show large variability across the five repeats for each cell line. The increased variability may be due to an increased number of factors influencing tumour growth such as interactions between the tumour and the vasculature and connective tissue.

4.3 In vitro tumour growth models

We propose a series of ODE models, with varying levels of complexity to describe the growth of the NSCLC cell lines in vitro (see Figures 4.2.1(a)-(d)).

4.3.1 Model development

Motivated by the experimental data, we consider the following four models that, at long times, evolve to a non-zero steady state. The models include the generalised-logistic

growth model, Richards growth equation, logistic growth, and the Gompertz model which present sigmoid shape growth curves [79, 117, 136]. We include the exponential growth model for comparison.

The five candidate models are presented in Table 4.3.1. Each model describes the time rate of change in cell number where $y(t)$ denotes the number of cells at time t. For all models, the growth rate is denoted by $r > 0$ (units of hour^{-1}), K denotes the carrying capacity (units of cells) and we fix $K = 50,000$ cells (see equation (4.2.1)).

The dimensionless parameter, $\beta > 0$, in the Richards growth equation model denotes the strength of competition; competition increases for smaller values of β. The dimensionless parameters α, β and γ in the generalised-logistic model are positive real exponents. A summary of the model parameters and their respective units is given in Table 4.3.2.

Models 2-5 can be derived from the generalised-logistic equation for specific model parameter values, as shown in Table 4.3.1. We aim to assess how well each model describes the in vitro data and the extent to which model parameters can be estimated given the data. The Gompertz growth model is commonly used to model in vivo tumour growth [79, 121]. The logistic growth model is commonly used to model cell proliferation assays and scratch assays for studies on wound healing [70, 87, 144]. The Richards equation was first used to model plant data and is less well known in modelling tumour growth [117]. In a recent study, Simpson et al. considered the application and practical identifiability of these three models when fitted to growth curve data describing coral cover off the coast of Australia [127]. A useful and informative review and analysis of the generalised-logistic, Richards growth equation, logistic growth and Gompertz model is given by Tsoularis et al. [136].

4.3.2 Initial conditions

For all models, we prescribe the initial number of cells so that

$$y(t = 0) = y_0 > 0, \tag{4.3.6}$$

where y_0 is a parameter to be estimated.

Model ref. number	Tumour growth model	Equation
1	Generalised-logistic	$\frac{dy}{dt} = ry^{\alpha}\left(1 - \left(\frac{y}{K}\right)^{\beta}\right)^{\gamma}$ (4.3.1)
2	Richards growth equation	$\frac{dy}{dt} = ry\left(1 - \left(\frac{y}{K}\right)^{\beta}\right)$ (4.3.2) $\alpha = 1$, $\gamma = 1$ in (4.3.1)
3	Logistic	$\frac{dy}{dt} = ry\left(1 - \left(\frac{y}{K}\right)\right)$ (4.3.3) $\alpha = 1$, $\beta = 1$, $\gamma = 1$ in (4.3.1)
4	Gompertz	$\frac{dy}{dt} = ry\ln\left(\frac{K}{y}\right)$ (4.3.4)
5	Exponential	$\frac{dy}{dt} = ry$ (4.3.5) $\alpha = 1$, $K \to \infty$ in (4.3.1)

Table 4.3.1: Table presenting candidate models of in vitro tumour growth: generalised-logistic, Richards growth equation, logistic, Gompertz, and exponential growth. $y(t)$ denotes the number of cells at time, t. For a summary of model parameters, see Table 4.3.2.

To summarise, we consider a selection of ODEs presented in Table 4.3.1 defined by equations (4.3.1)-(4.3.5) and initialised in equation (4.3.6).

4.3.3 Parameters

Model parameters for each candidate model are summarised in Table 4.3.2 in their dimensional form. We present biologically motivated bounds for model parameters in section 4.4.1.

In our models, we fix the carrying capacity parameter, K. Fixing K is a result of the in vitro experimental setup such that, when 100% confluence is reached and the

cells cover the entire plate, a fixed maximum number of cells is reached. Therefore, constraining K is appropriate. We acknowledge that, when applying models to in vivo settings, the consideration of unconstrained K should take place. In section 4.5.6, we consider in vivo data and, when fitting models to this data, we do not constrain K.

Parameter	Description	Units	Model	Estimated/fixed value
K	Carrying capacity	cells	1-4	fixed; 50000 cells
r	Growth rate	hour^{-1}	All	estimated
α	Exponent of exponential growth	dim.	1	estimated
β	Competition strength parameter	dim.	1,2	estimated
γ	Exponent of growth-limiting term	dim.	1	estimated
y_0	Initial cell number	cells	All	estimated

Table 4.3.2: Table of estimated and fixed parameters and their respective units, corresponding to in vitro candidate models given in Table 4.3.1 which are fitted to in vitro data. The column labelled Model refers to the model reference in which a given parameter is present. Carrying capacity, K is fixed at $K = 50000$ cells in candidate models 1-4 and all other parameters are estimated by fitting each model to the in vitro datasets. Dim. is an abbreviation for dimensionless.

4.3.4 Typical simulations of candidate ODE models

Typical solutions of all candidate ODEs are presented in Figure 4.3.1. In Figure 4.3.1, unconstrained cell growth is observed for the exponential growth model (blue line). For all other models, sigmoid shape growth curves are observed.

Consider the Richards growth equation defined by equation (4.3.2) in Table 4.3.1. The positive parameter, β indicates the strength of competition. We pause to discuss the implications of the value of β on tumour cell growth dynamics; referring to Figure 4.3.1, when $\beta = 1$, the Richards growth equation (black with varied line styles) reduces to the logistic model (grey) and the inflection point (time of most rapid growth) occurs when $y(t) = K/2$, i.e., when the cell population is half its carrying capacity. When $\beta < 1$, the limiting-effect of competition is stronger, the logarithmic growth phase is reduced, and the inflection point occurs at a smaller population size. For increasing values of β and $\beta > 1$, the inflection point approaches the carrying capacity and the cell dynamics tend towards exponential growth (blue). For $\beta \geq 3$, growth halts abruptly when the cell

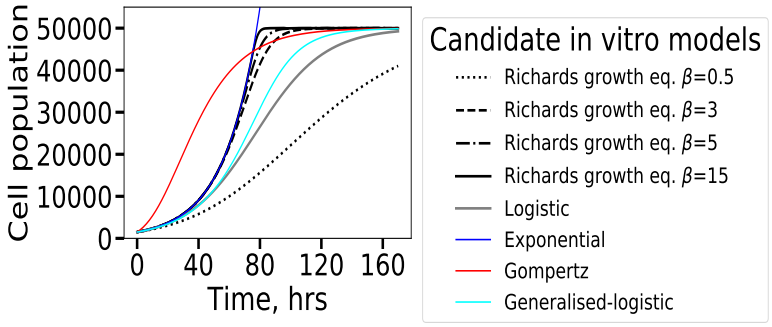


Figure 4.3.1: Illustrative simulations of the candidate ODE models defined by equations (4.3.1)-(4.3.5) (see Table 4.3.1). Key: Colour black; solution of Richards growth equation with varying line styles corresponding to a range of values of β, dark blue; the exponential growth model, grey; logistic growth model, red; Gompertz model, cyan; generalised-logistic model. Common parameters, $r = 0.045$, $K = 50000$. Parameter values specific to the generalised logistic model: $\alpha = 0.99$, $\beta = 2$, $\gamma = 1.1$. Parameter values specific to Richards growth equation: β varies for $\beta \in [0.5, 3, 5, 15]$. We fix $y_0 = 1500$ cells for all model solutions.

population reaches values close to the carrying capacity.

4.4 Methods

In this section, we begin by defining biologically motivated bounds on the model parameter values. We then outline the parameter estimation techniques that we use to estimate model parameters. We then introduce the information criteria and goodness-of-fit metrics that we use to compare the goodness-of-fits of the models when fitted to the data. Thereafter, we define structural identifiability and present a technique for applying it to our models. We then introduce practical identifiability and the profile likelihood method which we use for detecting parameters that are practically non-identifiable given the data. Next, we define the Bayesian parameter inference technique, Markov chain Monte Carlo (MCMC), a method we use to quantify the uncertainty of the model fits to the in vitro data. We present the diagnostic tests that we use to check for convergence of our MCMC. We use these tests to confirm the practical identifiability results obtained using the profile likelihood method.

We implement our models in Python 3.7.6. We solve our ODE models using the solve_ivp function from the Python library, SciPy 1.4.1. For plotting, we use Python libraries Matplotlib 3.1.3 and Seaborn 0.11.1. To combine and modify plots, we use

Inkscape.

4.4.1 Bounds on model parameters for in vitro data

We present biologically motivated bounds for model parameters in equations (4.3.1)-(4.3.5) (see Table (4.3.1)).

Model (ref. number)	Param.	Description	Units	Range
All models (1-5)	r	Growth rate	hour^{-1}	[0.001, 0.1]
All models (1-5)	y_0	Initial cell number	cells	[100, 2500]
Generalised-logistic (1)	α	Exponent of exponential growth	dim.	[0.5, 5]
Generalised-logistic (1)	β	Competition strength parameter	dim.	[0.1, 10]
Generalised-logistic (1)	γ	Exponent of growth-limiting term	dim.	[0.5, 5]
Richards growth equation (2)	β	Competition strength parameter	dim.	[0.1, 10]

Table 4.4.1: Parameter bounds for candidate models (with reference numbers 1-5 as in Table 4.3.1) when fitting to the in vitro data. Candidate models are given in Table 4.3.1. Dim. denotes dimensionless.

In Table 4.4.1, we present the ranges of the parameter values that appear in the candidate models (see Table 4.3.1). In Table 4.4.1, bounds on the growth rates, r (hour^{-1}) of EGFR lung adenocarcinoma sensitive and resistant cells are inferred from previous experimental measurements where values of r range from 0.0025 to 0.0125 (hour^{-1}) [72]. We take the conservative range of $0.001 < r < 0.1$. For an exponentially growing cell population, these values correspond to population doubling times ranging from 7 to 693 hours (doubling time$= (1/r)\ln 2$). In the context of the in vitro data, low values of r ($r = 0.001$) correspond to minimal growth over the period of the experiment (up to 200 hours), while for high values of r ($r = 0.1$) the cell population reaches its carrying capacity (50,000 cells) within approximately 35 hours if exponential growth is assumed and $y_0 = 1500$.

The in vitro data presented in Figures 4.2.1(a)-(d) show that the initial number of cells range between 1000 and 2250 cells across all cell lines. Therefore, we introduce the conservative bounds of $100 < y_0 < 2500$.

Parameters α, β and γ in the generalised-logistic model (see Table 4.3.1) are assumed to be positive; negative values do not provide a biologically plausible model [136]. We

take the conservative bounds of $0.5 < \alpha < 5$, $0.1 < \beta < 10$ and $0.5 < \gamma < 5$ so that a large region of parameter space is available when performing parameter estimation techniques. When performing initial fits of the generalised-logistic model to the in vitro data where a range of initial parameter value guesses were considered, the optimal α, β and γ values were within the ranges given in Table 4.4.1.

When using the Richards growth equation, we use the same bound for β as in the generalised-logistic equation, so that $0.1 < \beta < 10$. This range is consistent with the range of values of β considered by Tsoularis et al. [136].

4.4.2 Bounds of model parameters for in vivo data

Parameter	Units	Range
r	hour^{-1}	[0.0005, 0.025]
y_0	mm^3	[30, 110]
K	mm^3	[200, 600]
β	dimensionless	[1, 5]

Table 4.4.2: Parameter bounds for Richards growth equation (equation (4.3.2) in Table 4.3.1) when fitted to in vivo tumour volume data and parameters r, β, y_0 and K are estimated, where y_0 denotes the initial tumour volume.

When we fit the Richards growth model (equation (4.3.2)) to in vivo tumour volume data, we estimate r, y_0, β and the carrying capacity, K and we consider ranges of these parameters presented in Table 4.4.2.

We take the conservative bounds of $0.0005 < r < 0.025$ for the growth rates, r (hour^{-1}) whose values result in cellular dynamics similar to those of the in vivo tumour growth curves (see Figures 4.2.2(a)-(b)). We note that these values are lower than those for the in vitro data. We justify this difference by noting that, in vivo, death rates have been estimated to be close to proliferation rates. Higher death rates result in lower values of the net growth rate, r, than for in vitro growth [131].

Referring to the in vivo data presented in Figures 4.2.2(a)-(b), initial recorded in vivo tumour volumes range between 90mm^3 and 110mm^3 across all cell lines. Therefore, we consider the conservative bounds of $30 < y_0 < 110$.

For the carrying capacity K (units of tumour volume mm^3), we employ the conservative bounds of $200 < K < 600$.

We do not observe sigmoid tumour growth curves in Figures 4.2.2(a)-(b). Consequently, we are unable to determine the tumour volume at which growth saturates. We note that, at times $t = 552$ hours, the recorded in vivo tumour volumes exceed 200mm^3. The lower bound of $K = 200\text{mm}^3$ is chosen in order to assess the practical identifiability of K when the model is fitted to the in vivo data.

We consider a similar range of values of $1 < \beta < 5$ as the in vitro model parameter bounds when fitting the model to the in vivo data. This facilitates comparison of the practical identifiability of β for the in vitro and in vivo data.

4.4.3 Parameter estimation

To fit models to data and estimate parameter values, we use the non-linear least-squares minimisation method. We implement the method using, LMFIT: Non-Linear Least Squares Minimisation and Curve-Fitting for Python [103]. We calculate estimates of the parameter values, the uncertainties of these estimates, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics and the goodness-of-fit metric (the residual sum of squared errors whose formal definitions are given below) [2, 122].

Non-linear least squares minimisation method (NLLS)

For the in vitro datasets, three experimental repeats for a given cell line were conducted. For the in vivo datasets, five experimental repeats for a given cell line were conducted.

For the in vitro data (see Figures 4.2.1(a)-(d)), each experimental replicate is grown in an independent well. Therefore, at each time point, data points across the three experimental replicates are independent. In order to use the non-linear least squares method, we make the following assumptions about the experimental data:

1. Exact time points are known.
2. Observable values (data points) at a given time point are Gaussian distributed.
3. Observations at a given time point are independent.
4. The amount of scatter is the same at each time point.

With only three replicates for the in vitro data (five for the in vivo data), calculating standard deviations and means is not reliable. Hence, we minimise across all three repeats (or all five in the in vivo data) rather than using the mean as replicates are independent and should be treated as separate points. We minimise across all repeats for a given cell line, drug and dose as follows:

Suppose we have three experimental repeats as in the in vitro experiments. Consider the dataset corresponding to experiment repeat N, for $N = 1, 2, 3$. Assume this repeat dataset has m data points, (t_{N_1}, y_{N_1}), (t_{N_2}, y_{N_2}),..., (t_{N_m}, y_{N_m}), where y_{N_m} is the m^{th} data point for experimental repeat, N and assume that we have a model, $y = f(t, \boldsymbol{\theta})$, that depends on time, t and is non-linear in the model parameters, $\boldsymbol{\theta} = (\theta_1, ..., \theta_l)$. We seek to estimate the vector of parameters, $\hat{\boldsymbol{\theta}}$ that minimises, $\chi^2(\hat{\boldsymbol{\theta}})$, the residual sum of the squared errors, χ^2 where,

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{m} ({r_{1i}}^2 + {r_{2i}}^2 + {r_{3i}}^2), \tag{4.4.1}$$

with

$$\hat{\boldsymbol{\theta}} = \text{argmin}[\chi^2(\boldsymbol{\theta})] \tag{4.4.2}$$

and where

$$r_{Ni} = \frac{(y_{Ni} - f(t_i, \boldsymbol{\theta}))}{y_{Ni}}, \tag{4.4.3}$$

is the residual for experiment repeat, N and the i^{th} data point. Since we are working with large cell numbers, at each time point, t_i, we impose a relative weighting of the residuals and normalise with respect to the data point, $y(t_i)$. This ensures that the minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ is not skewed by the data. Otherwise, at large cell populations, larger deviations between the model predictions and the data would have a large effect on $\chi^2(\hat{\boldsymbol{\theta}})$. A disadvantage of this approach is that noise in data associated with small tumour sizes is amplified.

The goodness-of-fit of a model to the data is the minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ (equations (4.4.1)-(4.4.2)). As we are assuming normally distributed observational noise, $\chi^2(\hat{\boldsymbol{\theta}})$ corresponds to the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$

and is written as

$$\chi^2 = const - 2log(L(\boldsymbol{\theta})),$$

where $L(\boldsymbol{\theta})$ is the likelihood. In this chapter, $\chi^2(\boldsymbol{\theta})$ will represent the likelihood.

If $f(t, \boldsymbol{\theta})$ is non-linear in $\boldsymbol{\theta}$, the minimisation of χ^2 must be performed iteratively. Starting from a vector of initial parameter value guesses, we use the Levenberg-Marquardt Algorithm (LMA) to perform minimisation [83, 91]. On each iteration of the LMA, the goal is to find a perturbation, $\boldsymbol{\delta}$, to the parameters, $\boldsymbol{\theta}$, that minimises $\chi^2(\boldsymbol{\theta} + \boldsymbol{\delta})$.

We perform a non-linear least squares optimisation as follows:

1. We define a uniform distribution for each parameter (see Table 4.4.1) and randomly draw 30 initial guess parameter values for each model parameter, yielding 30 sets of model parameter values. We draw 30 samples to ensure the parameter space is adequately searched; multiple minima in $\chi^2(\boldsymbol{\theta})$ may occur and correspond to different sets of parameters.

2. We implement each set of initial guess model parameters into the LMA algorithm and find the local minima in $\chi^2(\boldsymbol{\theta})$.

3. We calculate the minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ for the 30 local minima found in step 2 and we define the parameter estimates associated with the minimum of the local minima as the optimal parameters, $\hat{\boldsymbol{\theta}}$.

Asymptotic standard errors, e_p

Uncertainties in estimated parameter values are calculated using the LMFIT package. The vector of asymptotic standard parameter errors, e_p is related to the covariance matrix, V as follows,

$$e_p = \sqrt{diag(V^{-1})}. \tag{4.4.4}$$

If a parameter has no effect on the fit, the covariance matrix will be singular and the corresponding values of the standard errors will be large relative to the estimated value of that parameter.

4.4.4 Akaike and Bayesian information criterion for model comparison

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics are used to quantify the quality of a model fit to the data [2, 122]. AIC and BIC reward goodness-of-fit and account for over-fitting by penalising models with too many parameters. When comparing models, the preferred model is the one with the lowest AIC and BIC values. The formal definition of the AIC statistic is,

$$\mathrm{AIC} = m \ln\left(\frac{\chi^2(\hat{\boldsymbol{\theta}})}{m}\right) + 2l, \tag{4.4.5}$$

and the formal definition of the BIC statistic is,

$$\mathrm{BIC} = m \ln\left(\frac{\chi^2(\hat{\boldsymbol{\theta}})}{m}\right) + l \ln(m) \tag{4.4.6}$$

where m denotes the number of data points and l, the number of estimated model parameters.

4.4.5 Structural and practical identifiability

An aim of this chapter is to determine the simplest mathematical model that can describe the in vitro experimental data so that model parameters, describing the growth properties of NSCLC cell lines can be estimated. We must first use structural and practical identifiability analysis to assess whether it is possible to uniquely estimate model parameters for a given model and data set. Parameter estimation techniques will struggle to estimate parameters that are not identifiable [22].

4.4.5.1 Structural identifiability

Structural identifiability relates to the structure of the model and is carried out prior to the consideration of data [115]. We refer to the definition of structural identifiability given by Chis et al. in [22]. We consider the following biological system that is described using a set of ordinary differential equations:

$$\sum(\boldsymbol{\theta}) : \begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) + \sum_{j=1}^{n_u} \boldsymbol{g}_j(\boldsymbol{x}, \boldsymbol{\theta}) \boldsymbol{u}_j, \\ \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta}),\ \boldsymbol{x}(t_0) = \boldsymbol{x_0}(\boldsymbol{\theta}) \end{cases} \tag{4.4.7}$$

- $\boldsymbol{x} = (x_1, ..., x_{n_x}) \in \boldsymbol{M} \subset \boldsymbol{R}^{n_x}$ is the state variable, n_x is the number of states in the system and $\boldsymbol{M}$ is a subset of $\boldsymbol{R}^{n_x}$;
- $\boldsymbol{f}$, $\boldsymbol{g} = (g_1, ..., g_{n_u})$ and $\boldsymbol{h}$ are analytical functions of their arguments;
- $\boldsymbol{u} = (u_1, ..., u_{n_u}) \in \boldsymbol{R}^{n_u}$ is the n_u-dimensional input (control) vector of smooth functions;
- $\boldsymbol{y} = (y_1, ..., y_{n_y}) \in \boldsymbol{R}^{n_y}$ is the n_y-dimensional output (experimentally observed quantities);
- $\boldsymbol{\theta} = (\theta_i, ..., \theta_{n_\theta}) \in \boldsymbol{\Theta}$ is the vector of unknown parameters where $\boldsymbol{\Theta}$ is a open and connected subset of $\boldsymbol{R}^{n_\theta}$;
- $\boldsymbol{x_0}(\boldsymbol{\theta})$ is the vector of initial conditions chosen so that the model is well-defined, though its coefficients may be dependent on unknown parameters.

These functions and initial conditions may depend on the parameter vector, $\boldsymbol{\theta} \subset \boldsymbol{\Theta}$. A vector $s(\boldsymbol{\theta})$, is an **exhaustive summary** of the experiment if it contains only the information about the parameters, $\boldsymbol{\theta}$ that can be extracted from knowledge of the inputs, $\boldsymbol{u}(t)$ and outputs, $\boldsymbol{y}(t, \boldsymbol{\theta})$ of the experiment.

Definition: Structural identifiability

Structural identifiability analysis addresses whether unique values, or otherwise, of model parameters can be determined from the available observables where, perfect experimental data is assumed, which is continuous in time and noise-free.

- A parameter, θ_i, $i = 1, ..., n_\theta$, is *structurally globally (or uniquely) identifiable* if for almost any $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$,

$$\sum(\boldsymbol{\theta}) = \sum(\boldsymbol{\theta}^*) \implies \theta_i = \theta_i^*, \tag{4.4.8}$$

  whereby there exists a unique solution for θ_i that can be computed from the exhaustive summary;

- A parameter, θ_i, $i = 1, ..., n_\theta$, is *structurally locally identifiable* if for almost any $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$, there exists a neighbourhood $\boldsymbol{V(\theta^*)}$ such that,

$$\boldsymbol{\theta} \in \boldsymbol{V(\theta^*)} \text{ and } \sum(\boldsymbol{\theta}) = \sum(\boldsymbol{\theta^*}) \implies \theta_i = \theta_i^*. \tag{4.4.9}$$

  whereby more than one solution for θ_i can be computed from the exhaustive summary;

- A parameter, θ_i, $i = 1, ..., n_\theta$, is *structurally non-identifiable* if for almost any $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$, there exists no neighbourhood $\boldsymbol{V(\theta^*)}$ such that,

$$\boldsymbol{\theta} \in \boldsymbol{V(\theta^*)} \text{ and } \sum(\boldsymbol{\theta}) = \sum(\boldsymbol{\theta^*}) \implies \theta_i = \theta_i^*. \tag{4.4.10}$$

  whereby, no solution can be computed from the exhaustive summary for θ_i.

A broad range of analytical methods has been developed to test structural identifiability of ODE models and we refer the reader to the reviews [22, 97] for a summary and comparison of structural identifiability methods applied to a selection of models commonly used to describe biological systems [18]. Software packages have been developed to implement structural identifiability analysis techniques [8, 18, 23, 84, 95, 97, 140, 139].

In this chapter, we perform structural identifiability analysis using the MATLAB software package, GenSSI2 which employs a combination of the generating series approach and identifiability tableaus [23, 84]. GenSSI2 computes successive Lie derivatives to generate a non-linear system of equations in the parameters [23, 84]. If a unique solution to these equations is found, then the parameters are structurally globally identifiable. If global identifiability is not proved, GenSSI2 will assess local identifiability by determining whether there is more than one solution of the system of equations. If no solution exists, parameters are not identifiable.

We also use DAISY (Differential Algebra for Identifiability of SYstems) to assess structural identifiability [8]. DAISY is a software package which employs a differential algebra approach and determines the structural identifiability and identifiable combinations of parameters for models that are in the form of polynomial and rational functions.

The differential algebra approach generates a system of *input-output equations* for the model using Groebner bases [8, 36]. These equations are monic differential polynomials in the output and input variables and their derivatives and have rational coefficients consisting of parameters. The coefficients of these equations are a complete set of identifiable combinations of parameters for the model and the observations of the model. The structural identifiability of the parameters depends on the injectivity of the map from the parameters to the coefficients.

4.4.5.2 Practical identifiability

A model can be structurally identifiable yet *practically non-identifiable. Practical identifiability* relates to the ability for parameters to be estimated from finite noisy data [115, 116, 125]. We refer to the definition of practical identifiability, inspired by likelihood-based confidence regions, proposed by Raue et al. in [115] and we will present the formal definition below, when we present the profile likelihood method.

Methods for testing parameter identifiability include, but are not limited to, the profile likelihood method and the Bayesian inference technique, Markov chain Monte Carlo (MCMC) [58, 138]; both of which have been applied in a biological setting [13, 28, 75, 115, 125, 127]. Simpson et al. obtained similar results using the two methods to investigate the practical parameter identifiability of spatio-temporal models of cell invasion [125]. They found the profile likelihood method to be less computer intensive than MCMC and suggested that the profile likelihood method be used to search for optimal parameters before computationally intensive MCMC simulations are performed [125].

In this chapter, we use the profile likelihood method to investigate practical identifiability of model parameters, following the approach presented by Raue et al. [115]. We also employ MCMC, primarily to quantify the uncertainty of model fits to the in vitro data. By way of carrying out diagnostic checks, we also assess practical identifiability using this method [116, 125]. Descriptions of MCMC and diagnostic plots are given below. We focus primarily on profile likelihood results for a practical identifiability analysis of the unknown model parameters.

4.4.5.3 Profile likelihood, χ^2_{PL} and likelihood-based confidence intervals

The *profile likelihood*, $\chi^2_{PL}(\theta_i)$ of a given parameter θ_i is calculated by fixing θ_i across a range of values and, estimating all remaining parameters, $\theta_{j\neq i}$ for each fixed value of θ_i, [100, 115, 138]. The profile likelihood is defined by,

$$\chi^2_{PL}(\theta_i) = \underset{\theta_{j\neq i}}{min}[\chi^2(\theta)]. \tag{4.4.11}$$

Likelihood-based confidence intervals are determined by imposing a threshold on the likelihood and are defined as follows [115, 75],

$$\{\boldsymbol{\theta}|\chi^2(\boldsymbol{\theta}) - \chi^2(\hat{\boldsymbol{\theta}}) < \triangle_\alpha\}, \tag{4.4.12}$$

where the threshold is defined by the α quantile of the χ^2-distribution with *df* degrees of freedom (the number of degrees of freedom is equal to the number of fixed parameters, for example, $df = 1$ for one-dimensional profile likelihoods where one parameter is fixed at a time),

$$\triangle_\alpha = icdf(\chi^2_1, \alpha). \tag{4.4.13}$$

Asymptotically (i.e. for a sufficiently large number of data points), the threshold is calculated by the inverse cumulative density function denoted by *icdf* [75]. Hence, $\triangle\alpha = \chi^2(\alpha, 1) = 3.841$ at the 95% confidence level [75]. We employ this threshold as we have a large number of data points in the in vitro dataset.

The intersection between the profile likelihood, $\chi^2(\theta_i)$, and the threshold, $\triangle\alpha$, defines the likelihood-based confidence interval to a confidence level, α. The confidence interval is denoted by $[\sigma_i^-, \sigma_i^+]$ where σ_i^- and σ_i^+ represent the lower and upper interval bounds, respectively [115]. Cubic interpolation of the profile likelihood is used for ease of visualisation and to calculate values of σ_i^- and σ_i^+.

Two-dimensional profile likelihoods are calculated by defining a finite 50x55 grid of pairs of parameter values. At each grid point, the profile likelihood is calculated

by optimising with respect to all other parameters. $\triangle\alpha = icdf(\chi^2_2, \alpha) = 5.99$ for two-dimensional profile likelihoods at the 95% confidence level. We use the Python contour package to plot two-dimensional profile likelihoods and 95% confidence thresholds.

4.4.5.4 Practical identifiability analysis using likelihood-based confidence intervals

We use the upper and lower bounds of the likelihood-based confidence interval to assess the practical identifiability of a model parameter. We use the formal definition of practical non-identifiability with respect to confidence intervals given by Raue et al. as follows [115]:

Definition: A *practically non-identifiable* parameter is a parameter whose likelihood attains a minimum value, but does not cross the threshold, $\triangle_\alpha$, for increasing and/or decreasing values of θ_i. The likelihood-based confidence region, therefore, has at least one infinite bound, i.e. either $\sigma_i^- = -\infty$ and/or $\sigma_i^+ = +\infty$. When two-dimensional profile likelihoods are calculated, practical non-identifiability can be visualised as a flat valley that is infinitely extended along either the upper or lower bound, or both [115]. For a practically identifiable parameter, the intersection between the cubic interpolated profile likelihood, χ^2_{PL}, and $\triangle_\alpha$ defines the lower and upper bounds of the confidence interval, $[\sigma_i^-, \sigma_i^+]$. Such finite intervals can not be determined for non-identifiable parameters. We refer to an identifiable model as one in which all its parameters are structurally and practically identifiable.

When a model is structurally non-identifiable, it is possible to improve its identifiability by: (i) reducing the number of states and parameters in the model, (ii) fixing parameters that are shown to be less relevant using additional analysis such as sensitivity analysis, and (iii) to design new experiments to increase the number of model outputs (measured quantities) [22]. Experimental design can be used to tackle practical non-identifiability. This involves proposing additional measurements that will reduce parameter uncertainty by increasing the degree of curvature of the likelihood profile. Additional measurements may also enable finite confidence intervals to be determined for parameters that were previously non-identifiable. Studies involving model based experimental design studies have been performed to decrease parameter uncertainty [76].

4.4.5.5 Algorithm used for the calculation of profile likelihoods

The algorithm used to calculate $\chi^2_{PL}(\theta_i)$ is as follows [115, 125]:

1. For parameter θ_i, we define a uniformly spaced grid of 30 points spanning the interval in Table 4.4.1. We consider 30 points to make it more likely that the profile likelihood attains a minimum close to the optimal parameter value.

2. For each fixed value of θ_i, we optimise with respect to the remaining model parameters. We repeat this optimisation 15 times, with each optimisation starting from a different set of randomly drawn initial guess parameter values (see description of parameter estimation techniques presented in subsection 4.4.3). This yields 15 local minima and we take the minimum value of these local minima. We choose 15 sets of initial guess parameter values to ensure the parameter space has been adequately searched as multiple minima values of $\chi^2(\theta_i)$ corresponding to the sets of parameters, may be present. In summary, the profile likelihood, $\chi^2_{PL}(\theta_i)$, consists of 30 values of $\chi^2(\theta_i)$, one for each value of θ_i on the grid constructed in step 1.

3. We perform cubic interpolation of the profile likelihood, χ^2_{PL} to allow for smoothness of the calculated profile likelihood points. We plot the interpolated profile likelihood and then calculate the intersection of the line, which we plot at the goodness-of-fit plus the threshold value $\chi^2(\hat{\boldsymbol{\theta}}) + \triangle\alpha$, and the interpolated likelihood to obtain likelihood-based confidence intervals of θ_i with a confidence level of 95%.

4.4.6 Markov chain Monte Carlo (MCMC)

Bayesian techniques, such as Markov chain Monte Carlo (MCMC), for parameter inference, generate probability distributions as opposed to the single point estimates, which are generated using non-linear least squares (see chapter 1 for more details about Bayesian inference and MCMC). A common algorithm used to generate a Markov chain is the Metropolis-Hastings (M-H) algorithm and many algorithms are based on it [96]. We summarise the (M-H) algorithm here. For iteration i, the Markov chain is at the

position, θ_i in the parameter space. A new potential position in the parameter space, θ^*, is proposed. The posterior probability of this new position, $p(\theta^*|y_N)$, is calculated using Baye's rule, where y_N denotes the data. The new position is accepted if the posterior probability is greater than the posterior probability of the current position and θ^* is added to the Markov chain. If the probability is less than the current probability, θ^* is rejected and a new proposed position is sampled. Hence, transitions in the Markov Chain are accepted with probability α where,

$$\alpha = min\left(1, \frac{p(\theta^*|y_N)}{p(\theta_i|y_N)}\right) \tag{4.4.14}$$

These steps are repeated for a predefined number of iterations and for each iteration, the Markov chain moves towards areas of higher posterior probability [58].

We use the Python package, Pints to perform MCMC. Pints performs MCMC sampling and explores the parameter space using the differential evolution algorithm "emcee: the MCMC hammer" [39]. This MCMC sampler differs from the typical Metropolis Hastings algorithm by employing an ensemble of 'walkers' that search the parameter space in parallel, making for more efficient sampling of the parameter space [50]. We implement 10 independent MCMC chains, starting from different positions in the parameter space, to allow for more efficient sampling of the parameter space. We provide more details of the construction of our MCMC in the appendix of this chapter where we also present results of our MCMC analysis.

Since we are interested in information about model parameters which depends on the data and not the priors, we use uninformative uniform priors for all parameters. We introduce a noise parameter to infer the noise in the data and assume noise is additive and normally distributed, with zero mean and common variance, σ^2.

4.4.6.1 Diagnostic tests and identifiability analysis

We use a range of diagnostic tests to check for convergence of our MCMC. These tests can also be used to assess practical identifiability of model parameters [116]. In a Bayesian setting, poor identifiability of model parameters is characterised by MCMC chains that do not converge, multi-modal posterior parameter distributions that can present as broad

distributions of parameter values, and non-decaying autocorrelation times [125, 116, 58]. We carry out diagnostic tests and assess practical non-identifiability numerically and graphically by:

1. calculating the Gelman-Rubin R statistic, denoted by $\hat{R}$, judging that convergence has been reached if $\hat{R} \leq 1.01$ for all model parameters [43],

2. observing trace plots that present the accepted parameter values for each iteration when sampling from the parameter posterior distribution, to see if MCMC chains converge,

3. determining whether posterior distributions are multi-modal.

To increase independence between samples, we thin the MCMC chain by selecting every 1/100 sample. We also discard the first 5000 samples, known as the burn-in. We note that it is not possible to make a direct comparison of practical identifiability results generated from the profile likelihood method and MCMC since, for the profile likelihood method, we assume a fixed noise of $\sigma = 1$, whereas for MCMC, noise is a parameter, whose values are inferred from the data.

4.4.6.2 Posterior predictive checks

If identifiability is achieved, we carry out posterior predictive checks to test whether the model is a good fit to the data. We do this by sampling from the posterior predictive distribution. The posterior predictive distribution is a probability distribution describing possible future data sets. To sample from the posterior predictive distribution to carry out the posterior predictive check, we first combine all the samples from the independent MCMC chains (10 in total) and then randomly sample 30 samples from the combined MCMC chain. For each sample, we simulate the model. We then add Gaussian noise at every time point, using the σ value corresponding to the randomly selected sample of the combined MCMC chains. If, when adding noise, cell populations reach negative values, we set these values to zero. We carry out visual posterior predictive checks where we determine whether the model describes the data well, when the 30 samples from the posterior predictive distribution and the data are plotted together.

4.4.6.3 Calculating 95% credible intervals of the model fit to data

To calculate 95% credible intervals of the model fit to the in vitro data, we carry out the same procedure as for the predictive checks. However, we now employ all the samples from the MCMC chain. At each time point, we calculate the median and the 95% credible interval of the model by calculating the 2.5, 50 and 97.5th quantiles of the posterior predictive distribution.

4.5 Results

In this section, we present results in line with the workflow presented in Figure 4.1.1.

4.5.1 Structural identifiability analysis

We examine the structural identifiability of the candidate models. We carry out structural identifiability analysis using the GenSSI2 software in MATLAB and find the exponential, Gompertz and logistic growth models are globally identifiable (i.e. all parameters are globally identifiable) [84].

When the generalised-logistic model and the Richards growth equation model are written in the forms presented in Table 4.3.1, using GenSSI2, the models were found to be structurally locally identifiable however, no results on global identifiability for said models could be given due to the complexity of the algebraic expressions of parameters. In response to this, we added additional variables and augmented our state-space generalised-logistic and Richards growth equation models. In doing so, we are able to obtain systems of rational and polynomial functions for both models (for augmented state-space models, see section 4.A in the appendix). When the Richards growth equation is presented as polynomial and rational functions, all model parameters are found to be globally structurally identifiable by GenSSI2 and DAISY. The results from the Richards growth model confirm previous results presented by Simpson et al. using GenSSI software [127]. DAISY also confirmed global identifiability for the generalised-logistic model when written as the augmented state-space model with additional variables (see section 4.A in the appendix).

4.5.2 Model calibration

We begin by fitting each candidate model to the in vitro data for each cell line. For each candidate model, we estimate model parameter values and the initial cell population, y_0. For a given fit, we calculate the AIC and BIC statistics (see equations (4.4.5) and (4.4.6)) and the minimum residual sum of squared errors, denoted by $\chi^2(\hat{\boldsymbol{\theta}})$ (see equation (4.4.1)). Estimated parameter values and information criteria/goodness-of-fit statistics are summarised in Table 4.5.1. Figure 4.5.1 shows the best fit model solutions for each candidate model and each in vitro data set. For each candidate model fitted to each dataset, correlations between model parameters are given and are presented in Tables (4.5.2)-(4.5.6).

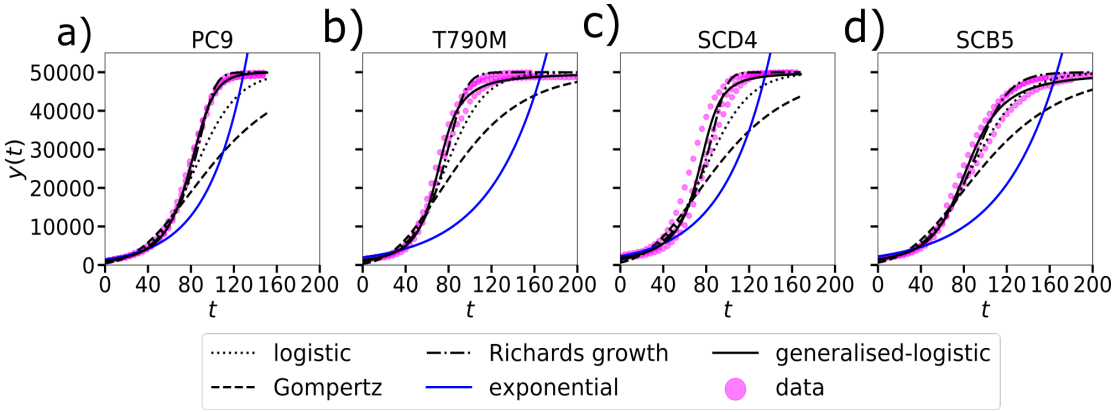


Figure 4.5.1: Best-fit model solutions of candidate models (see Table 4.3.1 for models) fit to the in vitro data for the four cell lines (PC9, T790M, SCD4 and SCB5). Pink dots represent the data. Block black lines represent the best fit solution of the generalised-logistic model, and similarly for the following models: black dashed line; Gompertz, black dashed-dotted line; Richards growth equation, blue block line; exponential, black dotted line; logistic. A summary of the estimated parameters for all models are given in Table 4.5.1 for each cell line.

In Table 4.5.1, for all cell lines, the generalised-logistic model yields the lowest $\chi^2(\hat{\boldsymbol{\theta}})$ value, as well as the lowest AIC and BIC statistics and the Richards growth equation model yields the second lowest $\chi^2(\hat{\boldsymbol{\theta}})$ value, as well as the second lowest AIC and BIC statistics. The other models do not fit the data well; their information criteria/goodness-of-fit statistics are higher than those of the best fit models. Based on information criteria/goodness-of-fit statistics alone, we conclude that the generalised-logistic model best describes the in vitro data for these cell lines and for this set of 2D experiments.

Cell line	Model	$\chi^2(\hat{\theta})$	AIC	BIC	r (hour^{-1})	α (dim.)	β (dim.)	γ (dim.)	y_0 ($\times 10^3$ cells)
PC9	Gen-logistic	0.28	-676.0	-662.0	0.0041±0.0019	1.28±0.06	1.4±0.5	1.08±0.15	1.115±.020
	Richards	0.67	-579.0	-571.0	0.0405±0.0004	–	4.6±0.8	–	0.940±0.019
	Logistic	2.76	-420.0	-415.0	0.0496±0.0010	–	–	–	0.765±0,033
	Gompertz	11.05	-262.0	-257.0	0.0198±0.0008	–	–	–	0.44±0.06
	Exponential	13.18	-242.0	-236.0	0.0277±0.0006	–	–	–	1.39±0.08
T790M	Gen-logistic	0.85	-668.0	-653.0	0.001±0.00033	1.46±0.11	1.5±0.6	1.75±0.26	1.28±0.04
	Richards	2.29	-539.0	-531.0	0.0439±0.0009	–	3.5±0.8	–	0.97±0.04
	Logistic	4.07	-464.0	-458.0	0.0541±0.0013	–	–	–	0.77±0.04
	Gompertz	12.54	-313.0	-308.0	0.0229±0.0010	–	–	–	0.34±0.06
	Exponential	34.85	-176.0	-171.0	0.0194±0.0007	–	–	–	1.96±0.15
SCD4	Gen-logistic	3.26	-409.0	-395.0	0.001±0.00002	1.42±0.20	2.0±1.8	1.6±0.7	1.77±0.11
	Richards	4.38	-378.0	-370.0	0.0375±0.0012	–	5.4±3.0	–	1.39±0.07
	Logistic	6.73	-330.0	-325.0	0.0480±0.0017	–	–	–	1.10±0.08
	Gompertz	14.10	-244.0	-238.0	0.0208±0.0011	–	–	–	0.63±0.12
	Exponential	17.89	-216.0	-210.0	0.0229±0.0007	–	–	–	2.23±0.14
SCB5	Gen-logistic	1.14	-647.0	-632.0	0.0010±0.0020	1.45±0.28	0.9±0.9	1.57±0.27	1.51±0.05
	Richards	1.78	-589.0	-581.0	0.0363±0.0009	–	1.92±0.22	–	1.28±0.04
	Logistic	2.40	-550.0	-544.0	0.0433±0.0007	–	–	–	1.11±0.04
	Gompertz	8.64	-375.0	-369.0	0.0192±0.0007	–	–	–	0.62±0.07
	Exponential	24.92	-229.0	-224.0	0.0187±0.0005	–	–	–	2.23±0.14

Table 4.5.1: Table of estimated values of model parameters (± standard errors) and goodness-of-fit statistics, sum of squared errors ($\chi^2(\hat{\boldsymbol{\theta}})$), AIC and BIC information criterion statistics for each candidate model (see Table 4.3.1) when fitted to the in vitro data for each cell line. Dim. denotes dimensionless.

In Tables (4.5.4)-(4.5.6), correlations between the estimated parameters r and y_0 for the logistic, Gompertz and exponential models are shown to be strongly negatively correlated. These models are sensitive to initial conditions and the correlations show that the estimated value of y_0 will have a large effect on the estimated value of r.

For correlations between estimated parameters in the Richards growth equation, given in Table 4.5.3, again, r and y_0 are strongly negatively correlated, demonstrating the sensitivity of growth rate estimates, r, to the initial conditions for all cell lines. We observe a weaker positive correlation between the parameters β and y_0 and a negative correlation between r and β for all cell lines. Based on these results, we suspect practical non-identifiability issues may arise for parameters, r and β, as they are negatively correlated. A possible practically identifiable parameter grouping could be $r\beta$, however we do not consider this grouping in this thesis as we wish to uniquely estimate cell growth rate parameters in order to characterise the four cell lines.

For the generalised-logistic model, correlations in Table 4.5.2 show all parameters to

Generalised-logistic						
Cell line		r	α	β	γ	y_0
PC9	r	1.000	-0.999	0.938	0.708	-0.689
	α	-0.999	1.000	-0.949	-0.727	0.665
	β	0.938	-0.949	1.000	0.888	-0.499
	γ	0.708	-0.727	0.888	1.000	-0.291
	y_0	-0.689	0.665	-0.499	-0.291	1.000
T790M	r	1.000	0.999	-0.941	-0.752	0.719
	α	0.999	1.000	-0.952	-0.769	0.696
	β	-0.941	-0.952	1.000	0.908	-0.532
	γ	-0.752	-0.769	0.908	1.000	-0.359
	y_0	0.719	0.696	-0.532	-0.359	1.000
SCD4	r	1.000	0.999	-0.903	-0.664	0.730
	α	0.999	1.000	-0.915	-0.680	0.709
	β	-0.903	-0.915	1.000	0.891	-0.492
	γ	-0.664	-0.680	0.891	1.000	-0.293
	y_0	0.730	0.709	-0.492	-0.293	1.000
SCB5	r	1.000	-0.999	0.978	0.808	-0.664
	α	-0.999	1.000	-0.984	-0.825	0.642
	β	0.978	-0.984	1.000	0.905	-0.547
	γ	0.808	-0.825	0.905	1.000	-0.339
	y_0	-0.664	0.642	-0.547	-0.339	1.000

Table 4.5.2: Table of correlations between estimated model parameters for the generalised-logistic model (see equation (4.3.1) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}), α (dim.), β (dim.), γ (dim.) and y_0 (cells) where dim. stands for dimensionless.

Richards growth				
Cell line		r	β	y_0
PC9	r	1.000	-0.550	-0.871
	β	-0.550	1.000	0.356
	y_0	-0.871	0.356	1.000
T790M	r	1.000	-0.618	-0.887
	β	-0.618	1.000	0.431
	y_0	-0.887	0.431	1.000
SCD4	r	1.000	-0.545	-0.882
	β	-0.545	1.000	0.369
	y_0	-0.882	0.369	1.000
SCB5	r	1.000	-0.830	-0.868
	β	-0.830	1.000	0.581
	y_0	-0.868	0.581	1.000

Table 4.5.3: Table of correlations between estimated model parameters for the Richards growth equation model (see equation (4.3.2) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}), β (dim.), and y_0 (cells) where dim. stands for dimensionless.

Logistic growth			
Cell line		r	y_0
PC9	r	1.000	-0.866
	y_0	-0.866	1.000
T790M	r	1.000	-0.880
	y_0	-0.880	1.000
SCD4	r	1.000	-0.875
	y_0	-0.875	1.000
SCB5	r	1.000	-0.848
	y_0	-0.848	1.000

Table 4.5.4: Table of correlations between estimated model parameters for the logistic growth model (see equation (4.3.3) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells).

Gompertz			
Cell line		r	y_0
PC9	r	1.000	-0.849
	y_0	-0.849	1.000
T790M	r	1.000	-0.876
	y_0	-0.876	1.000
SCD4	r	1.000	-0.861
	y_0	-0.861	1.000
SCB5	r	1.000	-0.838
	y_0	-0.838	1.000

Table 4.5.5: Table of correlations between estimated model parameters for the Gompertz model (see equation (4.3.4) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells).

Exponential			
Cell line		r	y_0
PC9	r	1.000	-0.796
	y_0	-0.796	1.000
T790M	r	1.000	-0.742
	y_0	-0.742	1.000
SCD4	r	1.000	-0.790
	y_0	-0.790	1.000
SCB5	r	1.000	-0.763
	y_0	-0.763	1.000

Table 4.5.6: Table of correlations between estimated model parameters for the exponential model (see equation (4.3.5) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells).

be strongly correlated to each other. This could suggest that practical non-identifiability issues may arise as values of estimated parameters can be compensated by changes in values of another parameter that they are strongly correlated with. In particular, parameters r and α have correlations -0.999 for PC9 and SCB5 and 0.999 for T790M and SCB4 which could suggest that either r or α is practically non-identifiable. Overall, the correlation results could suggest that the generalised-logistic model has too many parameters for the data available to us and a simpler model, such as the Richards growth equation would be more appropriate.

We aim to find the simplest model that describes the in vitro data and to estimate the growth rates of each cell type. Therefore, we reject the Gompertz, logistic and exponential growth models from herein due to poorly fitting solutions with these models.

4.5.3 Practical identifiability analysis

4.5.3.1 Profile likelihoods

We assess the practical identifiability of the parameters of the generalised-logistic and Richards growth equation models using the profile likelihood method [115]. Profile likelihoods are presented in Figure 4.5.2 for all cell lines.

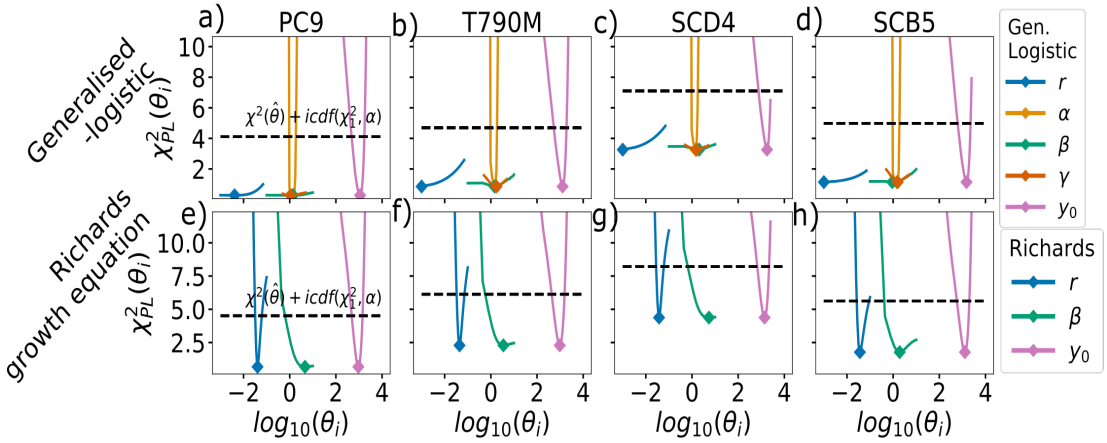


Figure 4.5.2: One-dimensional parameter profile likelihoods for the generalised-logistic model and the Richards growth equation (see Table 4.3.1) when fitted to each in vitro cell line data set. The x-axis describes the $log_{10}(\theta_i)$ profiled parameter values. Colours represent different model parameters (see figure legend). Coloured diamonds represent the $log_{10}(\hat{\boldsymbol{\theta}})$ value of the optimal parameter value. Black dashed lines represent the 95% confidence threshold and is plotted at the value $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$ where $icdf$ denotes the inverse cumulative density function. Coloured lines represent the interpolated profile likelihoods for each model parameter.

In Figure 4.5.2(a)-(d), the profiles for the generalised-logistic model, for all cell lines, show practical non-identifiability for parameters r, β and γ, with profiles that do not intersect the 95% confidence level. Finite upper and lower confidence bounds are obtained for α and y_0, demonstrating practical identifiability of these parameters. For the Richards growth equation (Figures 4.5.2(e)-(h)), for all cell lines, β is practically non-identifiable; its profile flattens out as β increases, yielding a one-sided confidence interval. We deduce that parameters r and y_0 in the Richards growth equation are practically identifiable when fit to the in vitro data because the corresponding profile likelihoods reach the 95% confidence interval twice. Based on the results presented in Figure 4.5.2, we conclude that the generalised-logistic and Richards growth equation models are practically non-identifiable when fitted to the in vitro data.

Recall that, when we base our choice of model on the $\chi^2(\hat{\boldsymbol{\theta}})$, AIC and BIC values, the generalised-logistic model gives the best results. However, based on results in Figure 4.5.2, the generalised-logistic model contains a higher number of practically non-identifiable parameters than the Richards growth equation model. Therefore, we may require additional data and/or extensive reduction in model complexity to achieve practical identifiability of the generalised-logistic model. In addition, the biological motivation for the generalised-logistic model parameters is limited. Hence, we reject the generalised-logistic model and consider the Richards growth model as our preferred model of tumour growth from this point forwards as it yields similar, albeit larger goodness-of-fit statistics as the generalised logistic model and has fewer practically non-identifiable model parameters (see Table 4.5.1).

The practical non-identifiability of the competition strength parameter, β, in the Richards growth equation model may be caused by noise in the data. Alternatively, the non-identifiability may be due to the behaviour of the Richards growth equation itself since its dynamics do not change significantly for sufficiently large values of β.

4.5.3.2 The impact of noise on the recovery and practical identifiability of model parameters

In this subsection, we use synthetic data sets to investigate how noise affects our ability to recover parameters r, β and y_0 in the Richards growth equation model. We also

investigate the effect of noise on likelihood-based confidence intervals, calculated using the profile likelihood method.

We generate synthetic data sets using the approach presented by Cho et al. [25]. We solve the Richards growth equation with parameters, $r = 0.04$ (hour^{-1}), $\beta = 3$ (dimensionless) and $y_0 = 1500$ (cells) and compute the solution at the same time points as in the in vitro datasets. At each time point, we add noise to the data as follows:

$$y_{noise}(t) = y_{exact}(t)(1 + \epsilon)$$

where $\epsilon \sim U(-x, x)$ for $100x\%$ noise. For example, for 5% noise, $x = 0.05$.

We assume multiplicative noise so that the relative noise level is preserved as the population changes over orders of magnitude. An alternative is to consider additive noise. However, if we assume additive noise with $\epsilon = 1000$ for example, the effects of noise on the model fit would be substantially more pronounced at low cell populations compared to high population cell numbers.

We consider noise levels, 0%, 1%, 5%, 10%, 20%. For each noise level, we generate 10 synthetic datasets. We then fit the Richards growth equation to each synthetic data set and estimate r, y_0 and β values and their corresponding standard errors. In Figure 4.5.3, we plot a single synthetic data set for noise levels 1%, 10% and 20% to illustrate the behaviour of the synthetic data sets as noise levels change. We also plot the solution of the Richards equation when fitted to the single synthetic data set.

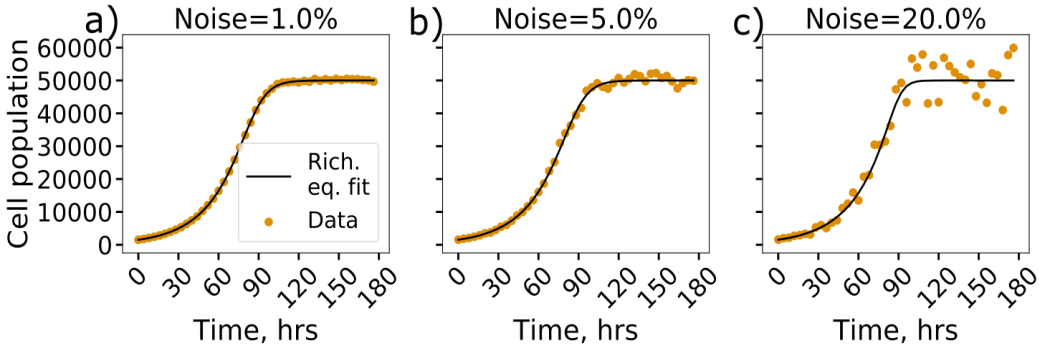


Figure 4.5.3: Synthetic data sets (orange dots) with noise levels, 1%, 5% and 20% and solutions of the Richards growth equation model fit to each data set (black lines).

Results of estimated model parameters and standard errors across the 10 synthetic datasets, for a given noise level, are presented in Figure 4.5.4.

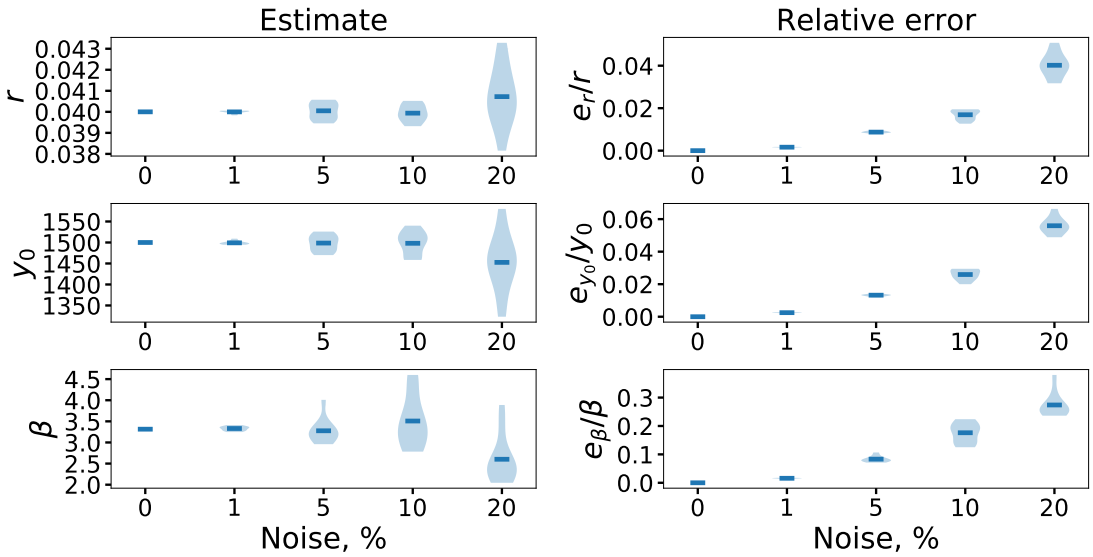


Figure 4.5.4: Violin plots of estimated parameter values and relative standard errors (standard error/estimate) of model parameters, r, β and y_0 estimated by fitting the Richards growth equation to each of the the 10 synthetic data sets at different noise levels. The synthetic data is generated from the Richards growth equation with true model parameters: $r = 0.04$ (hour^{-1}), $\beta = 3$ (dimensionless) and $y_0 = 1500$ (cells). Horizontal block lines in each violin plot represent the mean estimated values and relative errors across the 10 fits for a given noise level.

In Figure 4.5.4, there is a positive correlation between the noise levels and the mean estimated relative standard errors for all parameters. As noise levels increase, we observe larger variability in estimated parameter values and relative errors across the 10 runs. Profile likelihoods for the model parameters, generated from data with 0%, 5% and 20% noise levels are plotted in Figure 4.5.5. The profiles do not differ markedly as noise levels increase and, in all cases, β is practically non-identifiable. The profile of β is flat for increasing values of β, even with 0% noise and there is a slight increase in the likelihood values for lower values of β. We attribute this behaviour, in part, to the dynamics of the Richards growth equation model whereby similar dynamics are observed for large values of β, although the dynamics are sensitive to changes in β at lower values (see Figure 4.3.1).

4.5.4 Model reduction

So far, our results have demonstrated that the Richards growth equation suffers practical non-identifiability issues, even when fitted to synthetic data with 0% noise. In this

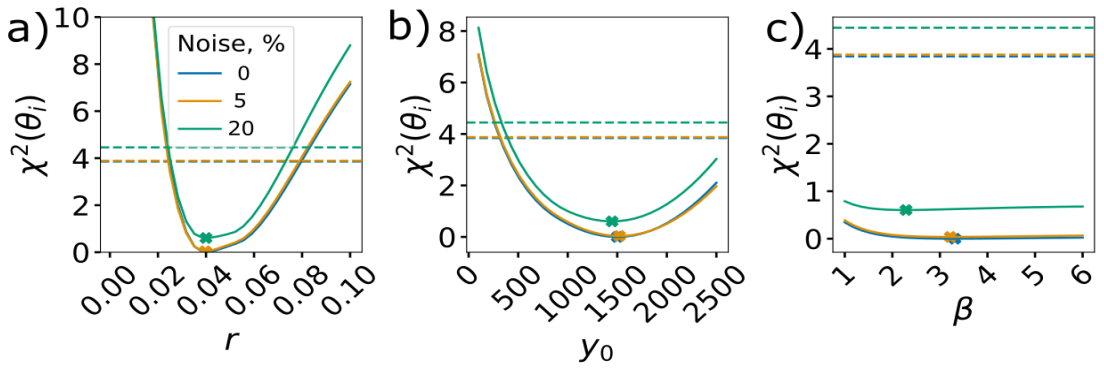


Figure 4.5.5: One-dimensional profile likelihoods of r, y_0 and β when fitting the Richards growth equation to synthetic data for 0%, 5% and 20% noise. Colour block lines indicate profiles of parameters for different noise levels. Coloured dashed lines indicate the 95% confidence level plotted at $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$ for different noise levels.

subsection, we look to fix model parameters to achieve practical identifiability of the model. Before doing so, we calculate two-dimensional profile likelihoods for parameters r and β for each in vitro cell line data set. By doing so, we can assess the relationship between these parameters and establish whether fixing β will not greatly affect the goodness-of-fit of the model [115]. Two-dimensional profile likelihoods are plotted in Figure 4.5.6 for all cell lines. The two-dimensional likelihoods are calculated by fixing values of r and β, estimating y_0 and calculating the minimum residual sum of squared errors, $\chi^2(\theta_i)$.

In Figure 4.5.6, for all cell lines, the likelihood-based confidence region is infinitely extended as $\beta \to \infty$, demonstrating the practical non-identifiability of the model. We conclude that, for a fixed value of r, increasing the value of β above 2 has little effect on the fit, for all cell lines. Based on these results, we reduce the complexity of the Richards growth equation model by first finding a universal value of β for all cell lines and then fixing β at this value. We note that, when using GenSSI2 to perform structural identifiability analysis, we found the Richards growth equation model to be globally identifiable when β is fixed.

We estimate a universal value of β by pooling in vitro datasets for all cell lines; we estimate parameters r and y_0 for each cell line and estimate a shared value of β across all cell lines for $\beta = 3.315$ (dimensionless).

Practical identifiability of the Richards growth equation model, with fixed $\beta = 3.315$,

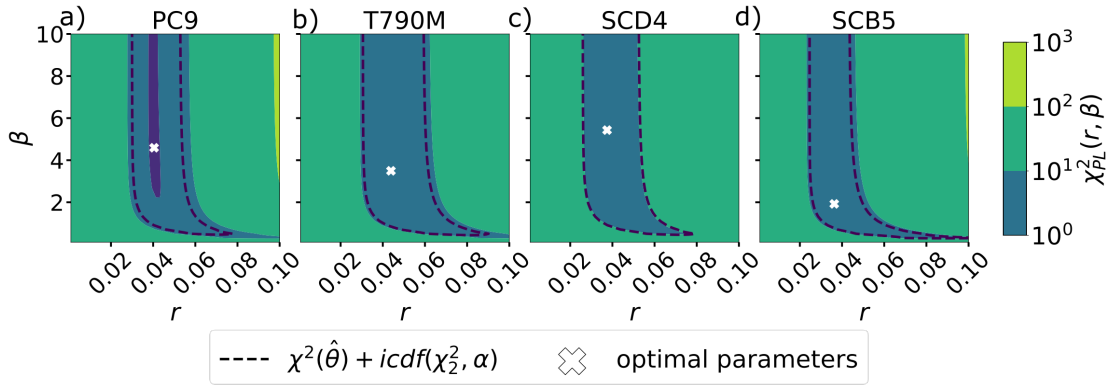


Figure 4.5.6: Two-dimensional profile likelihoods for parameters r and β ($\chi^2_{PL}(r,\beta)$) plotted as contour plots. Dark green and light green regions represent low and high $\chi^2(\theta)$ values, respectively. Black dashed lines represent the 95% confidence level calculated as $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_2, \alpha)$, where $icdf(\chi^2_2, \alpha)$ represents the $95th$ percentile for the χ^2 distribution with 2 degrees of freedom at the 95% confidence level. Crosses represent the coordinates of the optimal parameter values of r and β. For each cell line, the likelihood-based confidence region is infinitely extended for β demonstrating that an increase in β does not markedly alter the goodness-of-fit of the Richards equation to the in vitro data.

is verified in Figure 4.5.7. There, we present one-dimensional profile likelihoods of the Richards growth equation model, with fixed $\beta = 3.315$, when fit to in vitro data for each cell line; we note that the likelihood-based confidence regions for the parameters, r and y_0, are finite.

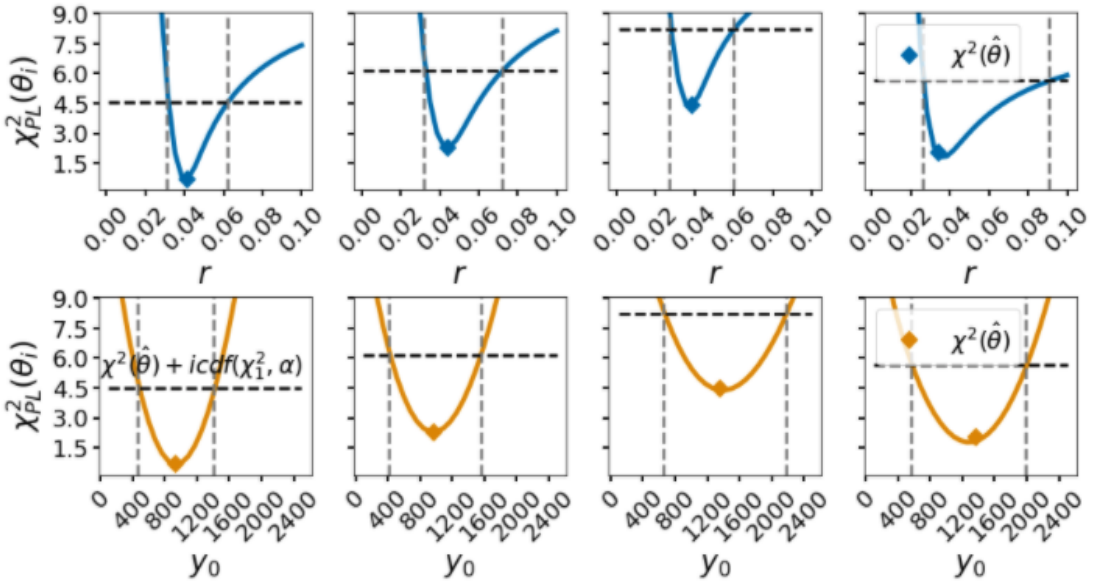


Figure 4.5.7: One-dimensional profile likelihoods of the reduced Richards growth equation model with fixed values of $\beta = \beta_F$ ($\beta_F = 3.315$). Black dashed lines represent the 95% confidence level calculated (see Figure 4.5.2 for details). The endpoints of the likelihood-based confidence intervals lie at the intersection between the likelihood and the 95% confidence level threshold and are indicated by vertical grey dashed lines (see Table 4.5.7). For all cell lines, r and y_0 is practically identifiable and we say that the model is identifiable.

In summary, by fixing the non-identifiable parameter, β, we have reduced the complexity of the Richards growth equation model and achieved practical identifiability.

4.5.5 Parameter estimation and model checks

Estimates of model parameter, goodness-of-fit values ($\chi^2(\hat{\boldsymbol{\theta}})$) and likelihood-based confidence intervals are found by fitting the reduced Richards growth equation model (with fixed β) to the in vitro data for each cell line and are presented in Table 4.5.7.

In Table 4.5.7, the $T790M$ cells are estimated to have the highest growth rate, with $r = 0.044$ hour^{-1}, followed by the PC9, SCD4 and SCB5 cells with estimated growth rates, 0.041, 0.038 and 0.034, respectively. We record large confidence intervals for r and y_0. We note that the confidence bounds for r and y_0 are close in value to those presented in subsection 4.5.2 where the Richards growth equation model, with free β, was considered, although, we do not present these bounds in this chapter. Thus, by fixing β, our model becomes practically identifiable, yet, we do not observe an increase in the precision of the parameter estimate as likelihood-based confidence intervals are

Cell line	$\boldsymbol{r}$ (hour^{-1}) Estimate ± std. err	Profile likeli. CI	$\boldsymbol{y_0}$ (cells) Estimate ± std. err	Profile likeli. CI	$\chi^2(\hat{\boldsymbol{\theta}})$
PC9	0.0410±0.0004	[0.032, 0.062]	924±18	[481.982, 1423.724]	0.711
T790M	0.0440±0.0007	[0.033, 0.072]	964±33	[431.532, 1555.856]	2.289
SCD4	0.0382±0.0010	[0.028, 0.06]	(1.36±0.07)e+03	[669.369, 2185.285]	4.465
SCB5	0.0341±0.0005	[0.026, 0.092]	(1.37±0.04)e+03	[575.676, 1997.898]	2.039

Table 4.5.7: Estimated values of parameters, r and y_0 (± standard error) when the reduced Richards growth equation (fixed $\beta = 3.315$) is fitted to the in vitro data for each cell line. The likelihood-based confidence intervals are calculated using the profile likelihood method (CI stands for confidence interval). The goodness-of-fit values, $\chi^2(\hat{\boldsymbol{\theta}})$ are also given. Std. err stands for standard error. Likeli. stands for likelihood.

Reduced Richards growth			
Cell line		r	y_0
PC9	r	1.000	-0.864
	y_0	-0.864	1.000
T790M	r	1.000	-0.875
	y_0	-0.875	1.000
SCD4	r	1.000	-0.875
	y_0	-0.875	1.000
SCB5	r	1.000	-0.848
	y_0	-0.848	1.000

Table 4.5.8: Table of correlations between estimated model parameters for the reduced Richards growth model with fixed $\beta = 3.315$ (see equation (4.3.2) in Table 4.3.1) when fitted to the in vitro data for each cell line. The units of the parameters are, r (hour^{-1}) and y_0 (cells).

not narrower than when β is also estimated. Although practically identifiable, large confidence intervals in Table 4.5.7 demonstrate the difficulty of reliably estimating growth rates given the available data.

In Table 4.5.8, correlations between the estimated parameters, r and y_0 are given. We observe a strong negative correlation between r and y_0, suggesting estimated values of the growth rate, r strongly depend on the initial cell numbers. We observed similar results in Tables (4.5.2)-(4.5.6). To conclude, the sensitivity of the estimated growth rate values to the initial conditions is great. Our results suggest that experimental procedures should be put in place to minimise the error in the initial number of cells across experimental repeats. We hypothesise that this would give rise to more accurate results when characterising the different cell lines based on their growth properties. To conclude, the sensitivity of the estimated growth rate values to the initial conditions is extensive.

We compare $\chi^2(\hat{\boldsymbol{\theta}})$ values of the Richards growth equation model, with free (see Table 4.5.1) and fixed values of β (dimensionless) (see Table 4.5.7), for each cell line. We observe slightly lower $\chi^2(\hat{\boldsymbol{\theta}})$ values when β is estimated. However, the difference is small, with $\chi^2(\hat{\boldsymbol{\theta}})$ values of 0.71, 2.29, 4.47 and 2.04 for PC9, T790M, SCD4, SCB5 cell lines, respectively, when β is fixed compared to lower values of 0.67, 2.29, 4.38 and 1.78, respectively, when β is estimated. Estimated values of r are the same for free and fixed β for PC9, T790M and SCD4 cell lines (see Tables 4.5.1 and 4.5.7). A lower value of $r = 0.034$ hour^{-1} is estimated for SCB5 cells when β is fixed as opposed to when β is also estimated ($r = 0.036$ hour^{-1}). Estimated values of y_0 are also similar, as are the standard errors associated with the estimated values of r and y_0. These results suggest that the estimated values of r and y_0 are insensitive to the fixing of β.

In Figure 4.5.8, we present the best fit model solutions and residual plots for each cell line. Model solutions were constructed using the optimal parameter values, found using non-linear least squares (NLLS). We sample from the posterior predictive distribution (PPD), generated from our MCMC analysis, and, for each time point, plot the median and 2.5% and 97.5% percentiles, giving a 95% credible interval of the model fit to data. Residuals are weighted with respect to the data at each time point. We plot residuals

of the best-fit NLLS solutions and the median posterior predictive distribution. At each time point, three residuals are plotted for the median and the NLLS solution as three experimental replicates were performed (see subsection 4.4.3 for more details).

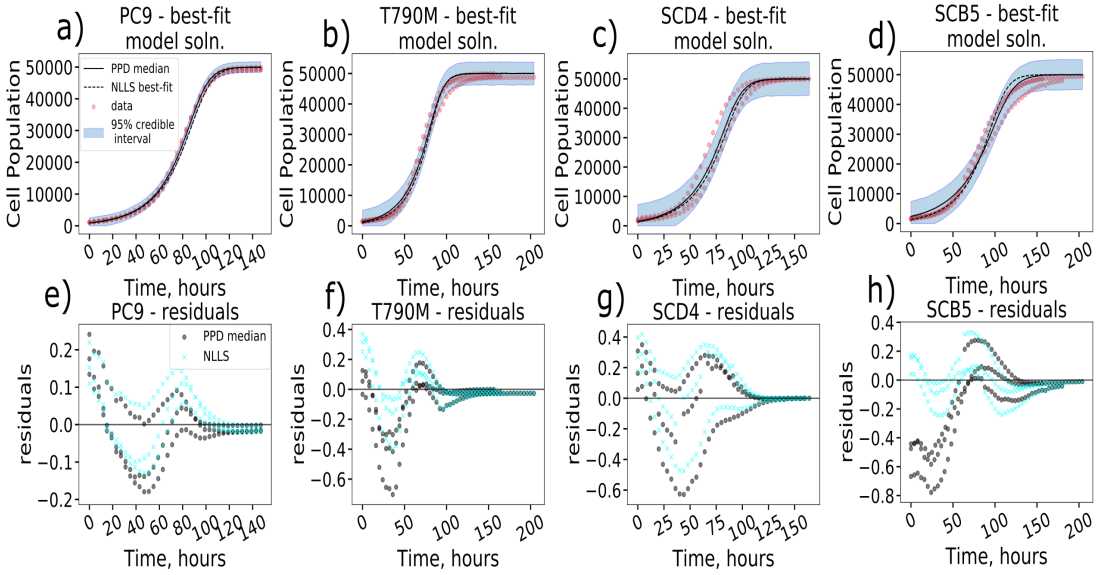


Figure 4.5.8: Richards growth equation model, with fixed β, fitted to the in vitro data. (a)-(d) Solutions of the Richards growth equation model fitted to the in vitro data. Black dashed line; best-fit model solutions found using non-linear least squares, black block line; median of posterior predictive distribution, blue shaded region; 95% credible interval of the model (between the 2.5th and 97.5th percentile of the posterior predictive distribution) found using MCMC. Red dots represent the data. (e)-(h) Corresponding residual plots to subplots (a)-(d). Cyan crosses represent residuals for NLLS. Black dots represent residuals for the median of the posterior predictive distribution. Black horizontal lines indicate residual values equal to zero.

In Figure 4.5.8, we observe similar dynamics for PC9, T790M and SCD4 cells. For the SCB5 cells, we observe a larger variation between the solutions of the median of the PPD and the solution found using NLLS. However, the NLLS solution lies within the 95% credible interval. The 95% credible intervals for SCD4 and SCB5 are larger than those of the PC9 cells whose interval is relatively narrow for all time points.

We plot the weighted residuals with respect to the data for the NLLS solution (cyan crosses) and the median PPD (black dots). By visual inspection, for $t < 20$ hours, we observe clusters of residuals above (PC9, T790M and SCD4) and below (SCB5) the x-axis, showing the model solution generated by NLLS and the median of the PPD deviate from the data for these times. For T790M and SCB5 cell lines, the residuals appear to

deviate from the data whereby, for a series of consecutive time points, residuals across all three experiments are either positive or negative for both the NLLS solution and the median PPD. This suggests that the residuals may not be randomly scattered above and below the x-axis. This could be a result of constraining the fit by fixing β in the Richards growth equation.

4.5.6 Practical identifiability of the model fit to in vivo data

In this section, we use the profile likelihood method to assess the practical identifiability of the Richards growth equation when fitted to the in vivo data. We assess the identifiability of the parameters r, K, β and y_0 using GenSSI2 where K is an unknown parameter to be estimated. Previously, for the in vitro data studies, K was fixed in our models of tumour growth. We find the model to be structurally locally identifiable where the parameters r, K and β are locally identifiable. Ranges of parameters considered in parameter estimation and in calculating profile likelihoods are given in Table 4.4.2. Profile likelihoods for the model parameters for a given tumour cell type (PC9, T790M and 1:1 SCD4:SCB5) are given in Figure 4.5.9.

In Figure 4.5.9, we observe that only the growth rate parameters, r (units of hours^{-1}), for tumours derived from a 1:1 mixture of SCD4 and SCB5 cells and PC9 cells are identifiable. The parameter β has a flat likelihood for all cell types, suggesting that the fit is totally insensitive to changes in β within the predefined range. The profiles for the carrying capacity, K, flatten out for increasing values of K, for all tumour types, with confidence intervals that are determined by the *a priori* parameter bounds given in Table 4.4.2. The same holds for the y_0 profiles. Based on these results, for simplicity, we fix $\beta = 1$, in which case, the Richards growth equation reduces to the logistic growth model (see Table 4.3.1).

In Figure 4.5.10, we present the best fit solutions of the logistic growth model (Richards growth equation with fixed $\beta = 1$) fitted to the in vivo tumour volume data. A summary of the estimated model parameter values and their standard errors is given in Table 4.5.9 and correlations between estimated parameters are given in Table 4.5.10.

In Table 4.5.9, T790M tumours are estimated to have the highest in vivo growth rate ($r = 0.063$ hour^{-1}) compared to PC9 and 1:1 SCD4:SCB5 tumours which have estimated

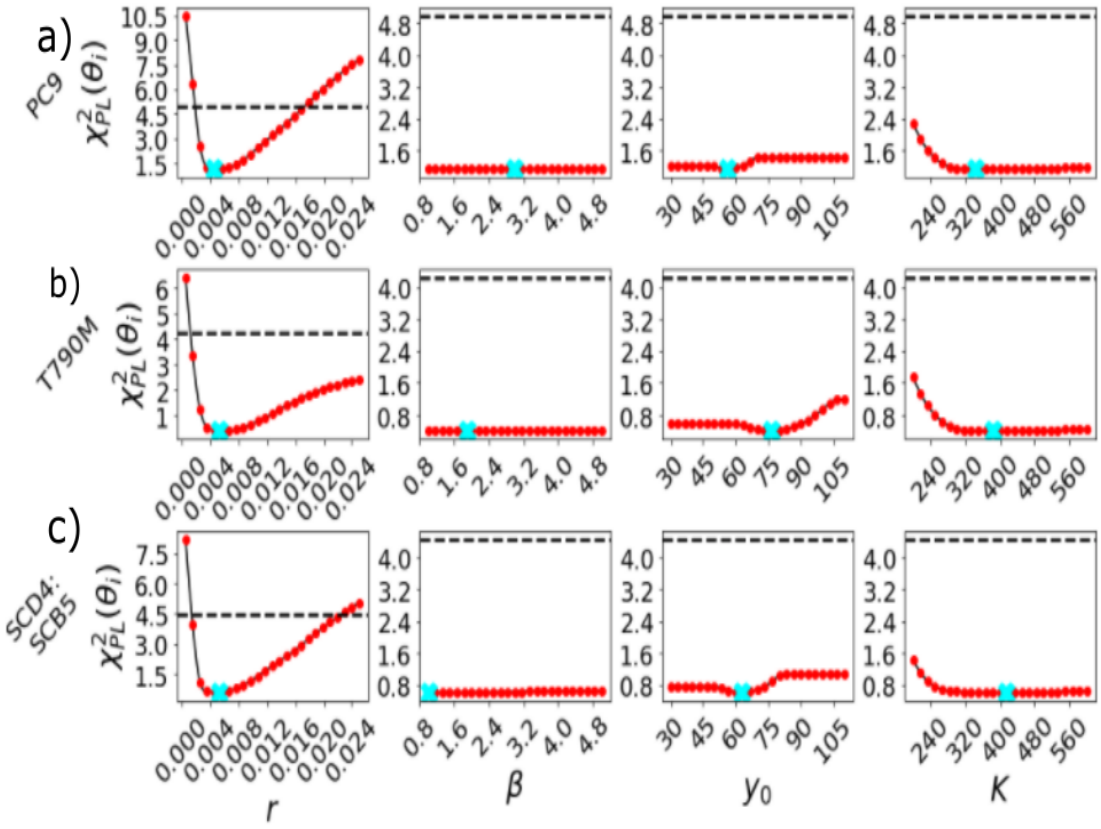


Figure 4.5.9: One-dimensional parameter profile likelihoods of the Richards growth equation model, with estimated β and carrying capacity, K. Horizontal black dashed lines represent the 95% confidence level, plotted at $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$. Red dots represent profile likelihood values. Black block lines represent the interpolated profile likelihood. Cyan crosses represent optimal estimated parameter values.

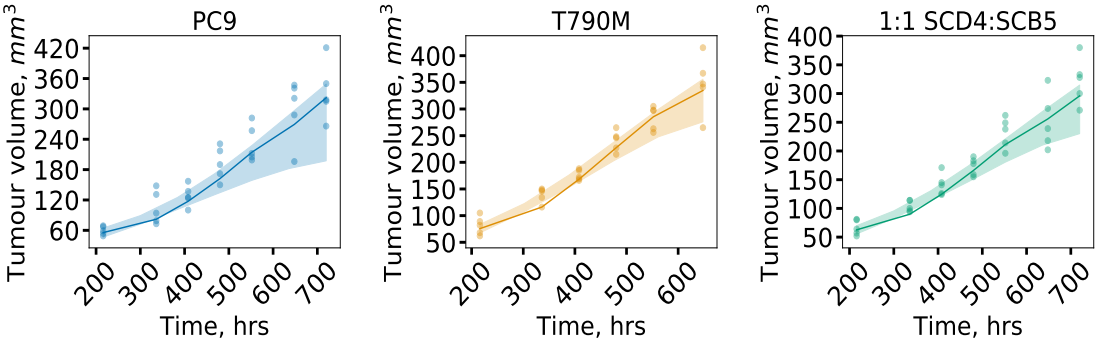


Figure 4.5.10: Best-fit model solutions of the Richards growth equation fitted to the in vivo data for each cell line, for fixed $\beta = 1$. Dots represent raw in vivo tumour volume data. Block lines represent the best fit solution for $\beta = 1$. Shaded regions represent 95% confidence bands, calculated by sampling from a multivariate normal distribution with model parameter estimates (see Table 4.5.9) as the means of the distribution. We employed each sample of parameter values (150000 in total) in the Richards growth equation. At each time point, we calculated the median and the 2.5th and 97.5th quantiles of the distribution of model solutions, yielding 95% confidence intervals of the model fit to the data.

Cell line	$\boldsymbol{r}$ Estimate ± std. err	$\boldsymbol{y_0}$ Estimate ± std. err	$\boldsymbol{K}$ Estimate ± std. err	$\chi^2(\hat{\boldsymbol{\theta}})$
PC9	0.0054 ± 0.0008	55.54 ± 4.54	490.41 ± 143	1.17
T790M	0.0063 ± 0.0008	75.57 ± 4.06	438.9 ± 66.2	0.41
1:1 SCD4:SCB5	0.0053 ± 0.0007	62.6 ± 3.67	413.01 ± 78.2	0.62

Table 4.5.9: Table of estimated in vivo model parameters (± standard error) for each tumour type. Estimated parameter values include the carrying capacity, K, growth rates r and the initial tumour volume, y_0. Std. err stands for the standard error.

Cell line		r	y_0	K
PC9	r	1.000	-0.726	-0.915
	y_0	-0.726	1.000	0.493
	K	-0.915	0.493	1.000
T790M	r	1.000	-0.644	-0.902
	y_0	-0.644	1.000	0.387
	K	-0.902	0.387	1.000
1:1 SCD4:SCB5	r	1.000	-0.694	-0.917
	y_0	-0.694	1.000	0.461
	K	-0.917	0.461	1.000

Table 4.5.10: Table of correlations between estimated parameters of the Richards growth equation model fitted to the in vivo tumour growth data for each tumour type.

growth rates $r = 0.0054$ hour^{-1} and $r = 0.0053$ hour^{-1}, respectively. Estimated values of K vary across the different tumour types with an estimated value of 490mm^3 for PC9 tumours, a value larger than 420mm^3 which is the largest recorded tumour size at 720 hours. Across all tumour types, we do not observe saturating growth dynamics. Therefore, we anticipate there will not be sensitivity to the carrying capacity because the carrying capacity is not being reached within the time period of these experiments.

In Table 4.5.10, the parameters r and K are shown to be strongly negatively correlated with correlations approximately equal to -0.9 for all tumour types. This suggests that, as the estimated value of r increases, the estimated value of K can decrease in order for a good data-model match to be maintained. Further, there is a strong negative correlation between estimates of y_0 and r with correlations ranging from -0.64 to -0.73 for all tumour types. Since, we do not observe sigmoid growth dynamics in the data and we have recorded a negative correlation with large magnitude between r and K, the model could be practically non-identifiable due to it containing too many parameters for the amount of data available to us. We also attribute lack of identifiability to

be a result of large variability in the experimental data. A response to this could be to fit an exponential model to the data whereby, the exponential model contains fewer parameters than the logistic model as only the parameters r and y_0 are estimated.

4.6 Discussion

When selecting which mathematical models to use to infer the biological properties of cell growth from experimental data, parameter identifiability analysis must be carried out. Such analysis allows the user to determine whether the available data are sufficient to obtain precise estimates of model parameters, for a given model. In this chapter, we have used a model selection procedure which incorporates both information criteria/goodness-of-fit metrics and parameter identifiability analysis. Our aim was to identify a model, with sufficient complexity, to describe the in vitro growth dynamics of homogeneous NSCLC cell populations and, then to compare the growth dynamics of different NSCLC cell lines.

We considered a suite of models, with the generalised-logistic model being the most complex (i.e. the largest numbers of model parameters), followed by the Richards growth equation and logistic growth models, both being subcases of the generalised-logistic model, and, finally, the Gompertz growth model.

During our initial stage of model selection, we fitted each model to the in vitro data set and selected models on the basis of the lowest Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics and the residual sum of squared errors. The generalised-logistic model gave the lowest residual sum of squared errors for all cell lines and, although this model has the highest level of complexity (i.e. largest number of parameters), it yielded the lowest AIC and BIC values. The Richards growth equation yielded the second lowest AIC and BIC values and residual sum of squared errors. Although both AIC and BIC statistics were used and gave similar results, we have employed MCMC in this chapter for uncertainty quantification and it therefore could be more appropriate to use BIC as a Bayesian argument was used by Schwarz to adopt the statistic [122]. We proceeded with these two models and assessed the identifiability of their parameters using the profile likelihood method. Although both models were found

to be structurally identifiable, they suffered practical identifiability issues for all in vitro datasets. We found 3 out of the 5 model parameters in the generalised-logistic model to be practically non-identifiable given the data. Conversely, the Richards growth equation yielded 1 practically non-identifiable model parameter out of 3 model parameters. Based on these results, we proceeded with the Richards growth equation. These results highlight that the AIC and BIC statistics are less informative than the profile-likelihood method and can be a misleading tool if used in isolation for model selection. Our findings corroborate similar results obtained by Simpson et al. who highlighted the benefits of the profile likelihood method over this information criterion [127].

The main advantage of the profile likelihood method is that it allows for the identification of practically non-identifiable model parameters. Further, by observing their profiles, hypotheses can be proposed about the cause of practical non-identifiability. For example, we observed a flat profile for increasing values of β. Using synthetic data sets with levels of noise ranging from 0% to 20%, we investigated whether non-identifiability of β was due to noise in the data or whether the dynamics of the Richards growth equation were responsible. We found that β suffered practical non-identifiability issues when fitted to the synthetic data, even with 0% noise. Our results highlight that, even in the case of noise-free data and structural identifiability, practical identifiability issues may still arise.

In this chapter, obtaining additional experimental data was not feasible. Therefore, we sought to resolve practical non-identifiability of the parameter, β, by holding it fixed. In so doing, we achieved practical identifiability of the model. A drawback of fixing a model parameter is the possible reduction in the relevance of the model to describe the biological system. By fixing β, however, the precision of our estimates of r and y_0 did not decrease (i.e. likelihood-based confidence intervals of approximately the same size obtained in both cases). We attribute this to the dynamics of the Richards growth equation being insensitive to changes in β if β is sufficiently large.

An alternative to fixing the practically non-identifiable parameter, β, could be to rescale the Richards growth equation model. A possible scaling could be $V = y/K$ and $\tau = rt$, giving $\frac{dV}{d\tau} = V(1 - V^\beta)$. However, this does not simplify the problem

or eliminate the power term which gives rise to practical non-identifiability issues. We wish to characterise the different cell lines and therefore require unique estimates of the growth rate r for each cell line which this scaling does not provide. Alternatively, we could augment the model and include an additional variable, as shown in the appendix of this chapter. We could then calculate the profile likelihoods for the new model which is linear in its parameters. This is subject to future work.

A question that was of particular interest to us was how the identifiability of our model differed between in vitro and in vivo settings. We applied the profile likelihood method to the Richards growth equation model, with free β and the in vivo data. The identifiability of the model decreased in the in vivo setting. A possible explanation for the inadequacy of our model could be due to the fact that it does not take into account the increased number of factors influencing tumour growth in vivo (e.g. hypoxia and vasculature), factors that are absent in vitro. Another, and more plausible explanation could be that, as we do not observe saturating growth dynamics in the in vivo data, practical non-identifiability of model parameters could be a result of the logistic model containing too many parameters for the amount of data available to us, therefore making it an inappropriate model. Further, we observe large variability in tumour volume data across replicates at each time point, making it difficult to obtain a good fit to the data and, therefore, making it very difficult to identify model parameters.

In this chapter, we have considered in vivo mouse data. We suspect that in vivo mouse outcomes will not be directly translatable to human outcomes. This is because the human immune system will impact the tumour growth dynamics whereas the data presented in this chapter is generated from immunosuppressed mice. Further, patients will have different genetic profiles and the extent to which adverse effects of the drugs affect the patients may vary and this could result in large variability across outcomes. However, mouse studies can still give insight into principles and mechanisms and thus inform understanding even though they do not directly translate to human studies. Preclinical studies can provide useful, evidence based insight, for understanding tumour dynamics in general to inform our understanding of human responses.

Regarding the estimated model parameters for in vitro and in vivo settings, we

conclude that the T790M cell populations have the highest growth rates followed by the PC9 cell populations. The slowest estimated growth rates in vivo and in vitro correspond to the fully drug-resistant cells to osimertinib and gefitinib, with the T790M and C797S mutations. We find growth rates in vivo are approximately an order of magnitude lower than in vitro which may be due to a larger proportion of the tumour made up of quiescent cells in vivo. Another possible explanation is of higher death rates in vivo.

In summary, we have developed a model to describe four NSCLC cell lines grown in vitro, in the absence of drug. We have assessed the identifiability of our model when fit to in vitro and in vivo data. In light of our findings, we advocate that more focus should be put on inferring model parameters from the data and that model selection using goodness-of-fit statistics and information criterion statistics alone should be done with caution.

We hope to have provided the reader with an informed understanding of the process of model selection and identifiability methodologies and, where the benefits and pitfalls lie in the process of obtaining an identifiable model parameter to describe a biological system. This foundation will be useful in the next chapter where we will extend our model to investigate the effects of the TKIs, gefitinib and osimertinib on the four NSCLC cell lines.

Appendix

4.A Structural identifiability analysis

In this section, we present the augmented state-space Richards growth equation and generalised-logistic models that are in the form of polynomials and rational functions.

A Richards growth equation

In the Richards growth equation model, defined in equation (4.3.2) in Table 4.3.1, by assuming

$$X = \left(\frac{y}{K}\right)^{\beta}, \tag{4.A.1}$$

we can rewrite the model as the following system of polynomial functions,

$$\dot{y} = ry\,(1 - X) \tag{4.A.2}$$

$$\dot{X} = r\beta X\,(1 - X) \tag{4.A.3}$$

where we assume $y(t)$ is the observed output of the model. We consider the initial conditions

$$[y_0, X_0], \tag{4.A.4}$$

where $y_0 > 0$ and $X_0 > 0$ are unknown constants.

B Generalised-logistic model

The generalised-logistic model, defined by equation (4.3.1) in Table 4.3.1, can be written as the following system of polynomial and rational functions,

$$\dot{y} = rUZ \tag{4.A.5}$$

$$\dot{U} = r\alpha\frac{U^2 Z}{y}, \tag{4.A.6}$$

$$\dot{X} = r\beta\frac{XUZ}{y}, \tag{4.A.7}$$

$$\dot{Z} = r\gamma\beta\frac{XUZ^2}{y(X-1)}, \tag{4.A.8}$$

where

$$U = y^{\alpha}, \tag{4.A.9}$$

$$X = \left(\frac{y}{K}\right)^{\beta}, \tag{4.A.10}$$

and

$$Z = \left(1 - \left(\frac{y}{K}\right)^{\beta}\right)^{\gamma} = (1-X)^{\gamma}. \tag{4.A.11}$$

The observed output of the model is $y(t)$ and we consider the initial conditions,

$$[y_0, U_0, X_0, Z_0], \tag{4.A.12}$$

where $y_0 > 0$, $U_0 > 0$, $X_0 > 0$ and $Z_0 > 0$ are unknown constants.

4.B Markov chain Monte Carlo results, diagnostic testing and identifiability analysis

In this appendix, we present results of applying MCMC to our reduced Richards growth equation with fixed β and the in vitro cell growth data. We also present a visual posterior predictive check for the Richards growth equation fit to all cell lines with fixed β. We use MCMC to construct posterior and joint posterior distributions of model parameters and to calculate the 95% credible intervals of the model when fit to the in vitro data. Results for each cell line are presented in Figure 4.5.8 in the main text. In what follows, we explain how we constructed our MCMC and present our priors. We then present the results of convergence tests which were used to ensure our MCMC has converged, and we refer to these tests to assess practical identifiability of the model parameters.

When constructing our MCMC, we use uniform priors with $400 < y_0 < 3000$ for the initial cell number and $10 < \sigma < 100000$ for the noise. For the growth rate, r, we employ a lower bound of 0.02 and the upper bound is cell line specific and fixed at the upper bound of the likelihood-based confidence interval, given in Table 4.5.7. Hence, we employ upper bounds of 0.062, 0.072, 0.06 and 0.092 for PC9, T790M, SCD4 and SCB5 cells, respectively. We run 10 independent MCMC chains, each consisting of 100,000

iterations for the PC9 and T790M datasets and 150,000 iterations for the SCD4 and SCB5 datasets. Each chain is initiated with initial parameter values sampled from a neighbourhood of 0.1ϵ about the optimal estimated parameter values (see Table 4.5.7) for $\epsilon \sim U(0, 1)$.

We run diagnostic tests to test for convergence of our MCMC and to assess the identifiability of model parameters (see section 4.4.6 in main text for details).

Diagnostic tests

To ensure convergence of our MCMC chains, we refer to the following diagnostic plots in Figures 4.B.1 and 4.B.2. Diagnostic tests include:

i plots of the autocorrelation function (ACF) for all MCMC samples where samples from each independent MCMC chain (10 independent Markov chains in total) are combined (see Figure 4.B.1);

ii trace plots for three MCMC chains for a given in vitro dataset (see Figure 4.B.2).

Figure 4.B.1 presents plots of the autocorrelation function (ACF), we observe, for all cell lines, rapidly decaying autocorrelation functions for all model parameters. This indicates independence of the samples and convergence of the Markov chain, therefore demonstrating the identifiability of this model [116].

Trace plots of 3 realisations of the MCMC chain are presented in Figure 4.B.2.

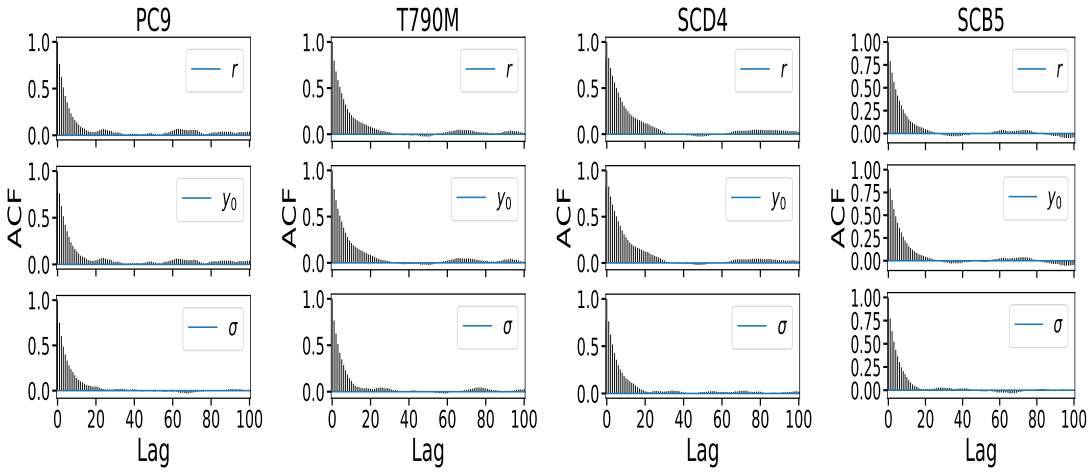


Figure 4.B.1: Autocorrelation function plots of combined Markov chain samples (10 independent Markov chains in total) for the Richards growth equation with fixed β. A thinning of 1/100 and a burn in of 5000 samples were used to increase the independence of the samples. Rapid decay of the autocorrelation function for all parameters and all datasets is observed, demonstrating the independence of the MCMC samples and convergence of the MCMC chains. Lag (x-axis) describes the distance between two samples in the MCMC sequence. The autocorrelation drops to a value close to zero within a lag of approximately 20 for all parameter and datasets, a value that is much smaller than the number of MCMC samples (up to 150,000 for SCD4 and SCB5 data sets), demonstrating the samples of the MCMC quickly become independent and are not correlated. These results suggest that the model is identifiable when β is fixed.

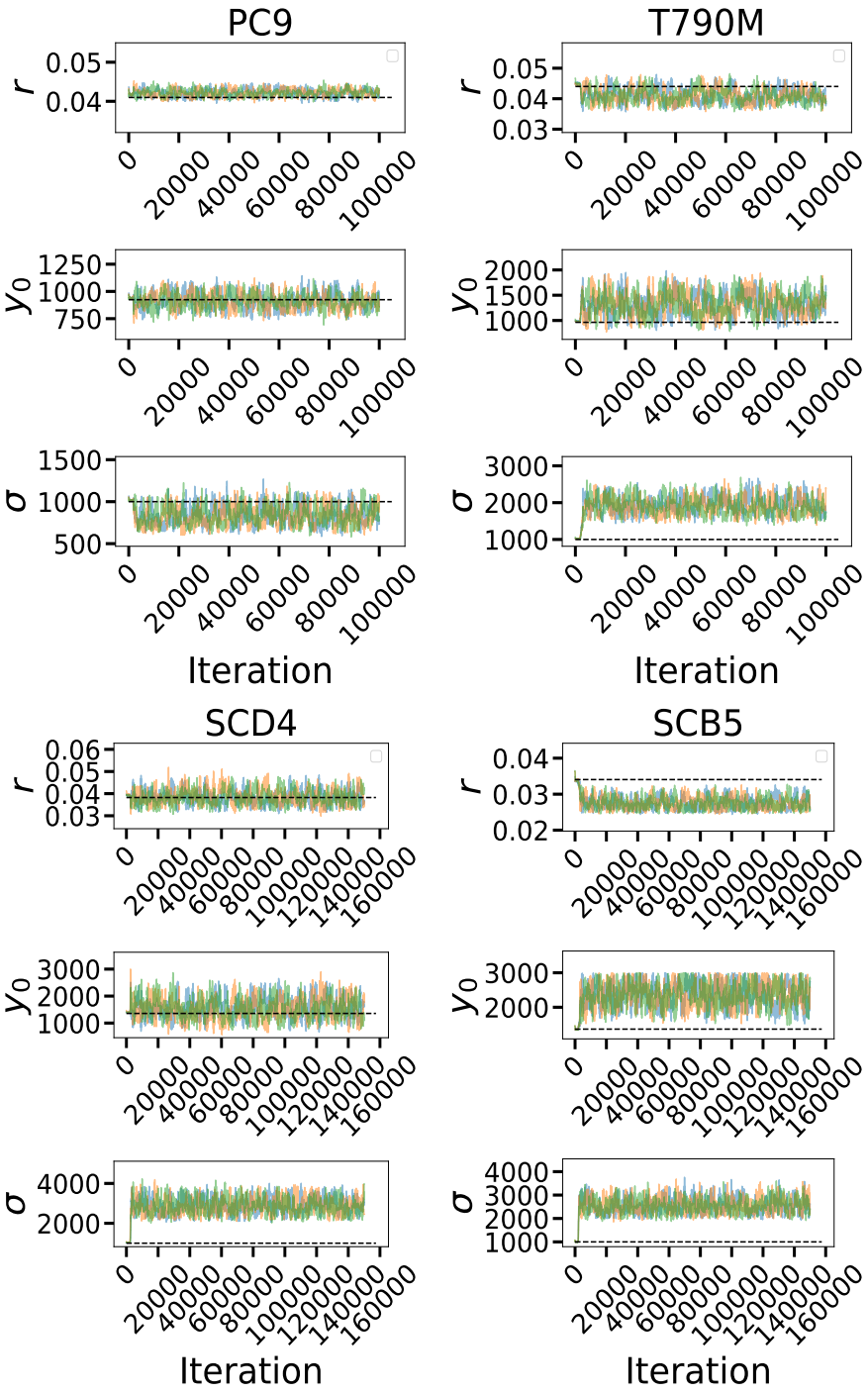


Figure 4.B.2: Trace plots of 3 randomly selected, independent Markov chains for each data set and for the Richards growth equation with fixed β. The Markov chains are of length 100,000 for PC9 and T790M data sets and of length 150,000 for SCD4 and SCB5 datasets. Each chain is initiated with initial parameter values which are sampled from a neighbourhood of 0.1ϵ about the optimal estimated parameter values, where $\epsilon \sim U(0,1)$. The Markov chains appear to converge for all model parameters and datasets. Key: colours represent different chains. These results suggest that the model is identifiable when β is fixed.

In Figure 4.B.2, rapid convergence is observed for all chains, regardless of their initial parameter values, the chains remain within the assigned parameter bounds having moved away from the initial parameter values. While the chain corresponding to y_0 for SCB5, appears to converge, it reaches the upper bound of the assigned parameter value, which suggests less certainty in this parameter.

In summary, by visual inspection, trace plots of MCMC chains were shown to converge for all parameters and all datasets and autocorrelation plots were shown to rapidly decrease to values close to zero, for all parameters and all datasets, demonstrating independence of samples and convergence of the Markov chain. These diagnostic plots indicate practical identifiability of the model parameters [116].

Posterior and joint posterior parameter distributions and posterior predictive checks

Practical identifiability of model parameters is also demonstrated by referring to the posterior distributions of model parameters in Figure 4.B.3. For all parameters and datasets, we observe unimodal proper posterior distributions that are approximately symmetric. Conversely, when performing MCMC on the Richards growth equation with free β, practical non-identifiability was diagnosed by non-converging trace plots, non-proper parameter posterior distributions, that presented broad ranges of values, and non-decaying autocorrelation plots (results not presented in this chapter due to limitations on thesis length).

Since our initial diagnostic inspection for convergence is satisfactory, we perform graphical posterior predictive checks (see section 4.4.6 for details) for each cell line. We present a visual posterior predictive check (PPC) with 30 randomly selected samples of the converged Markov chain where all samples across the 10 independent chains are combined. The results in Figure 4.B.4 provide visual confidence in the MCMC results and the suitability of the model.

In summary, we have demonstrated that the Richards growth equation, with fixed β, is practically identifiable. We have also shown the model is appropriate in describing the data across all cell lines by performing a posterior predictive check.

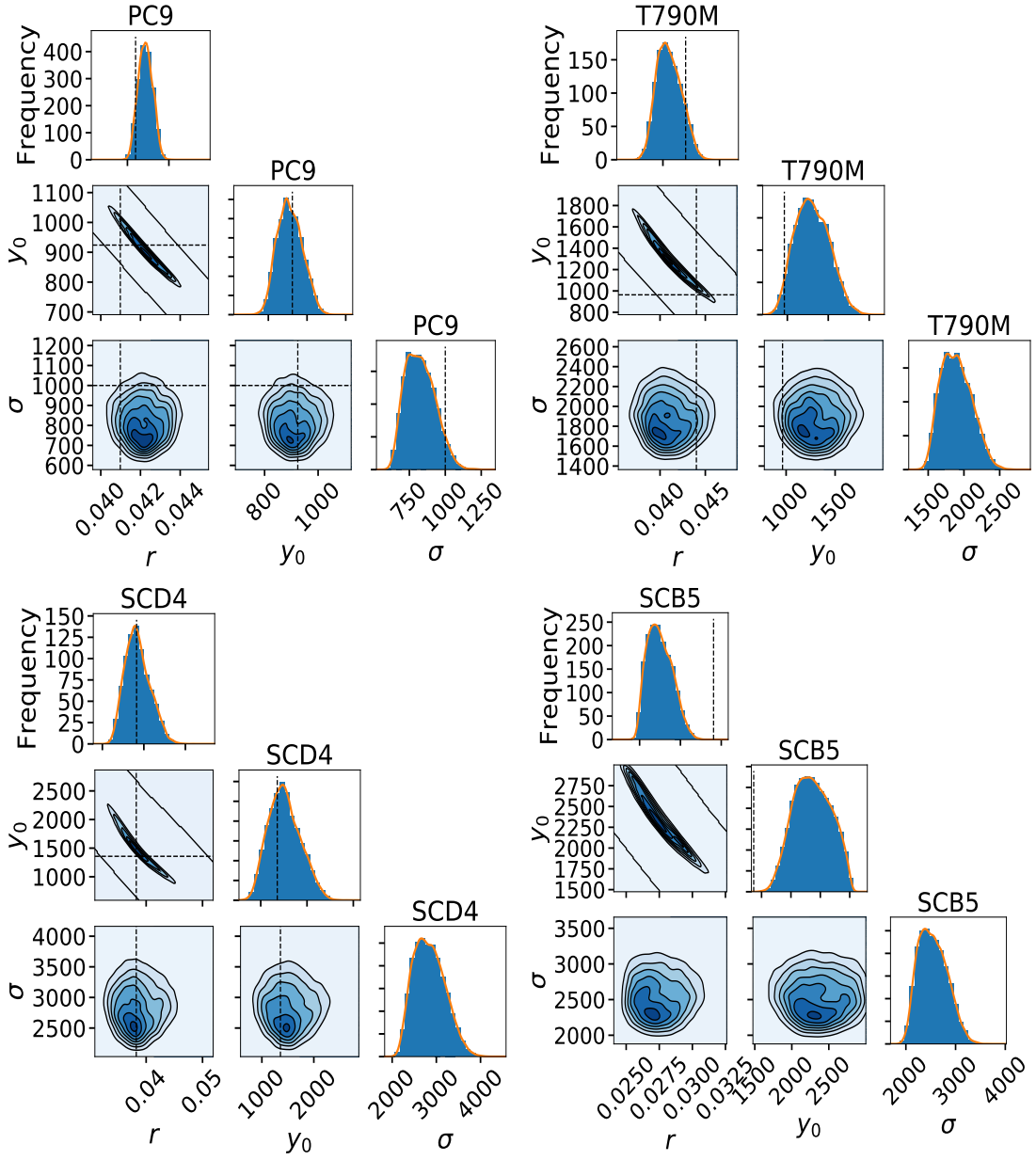


Figure 4.B.3: Posterior distributions (histograms on the diagonal) and joint posterior distributions (contour plots) of the Richards growth equation model parameters when β is fixed. For each in vitro dataset, 10 MCMC chains were run independently, with lengths 100,000 for the PC9 and T790M data sets and of lengths 150,000 for the SCD4 and SCB5 data sets. The independence of samples and the convergence of the MCMC chains were checked by the autocorrelation function and the trace plots. A Gelman-Rubin statistic of $\hat{R} \leq 1.01$ was calculated for all model parameters and all cell lines, confirming convergence of the MCMC chains. Black dashed lines indicate the optimal r and y_0 parameter values, estimated using non-linear least squares method and for σ, correspond to the initial value of the proposed noise level $\sigma = 1000$ in the prior parameter distribution, employed in our MCMC. Key: orange curves on diagonal plots; the kernel-density estimation for the histograms. For the contour plots: dark and light regions correspond to high and low densities, respectively.

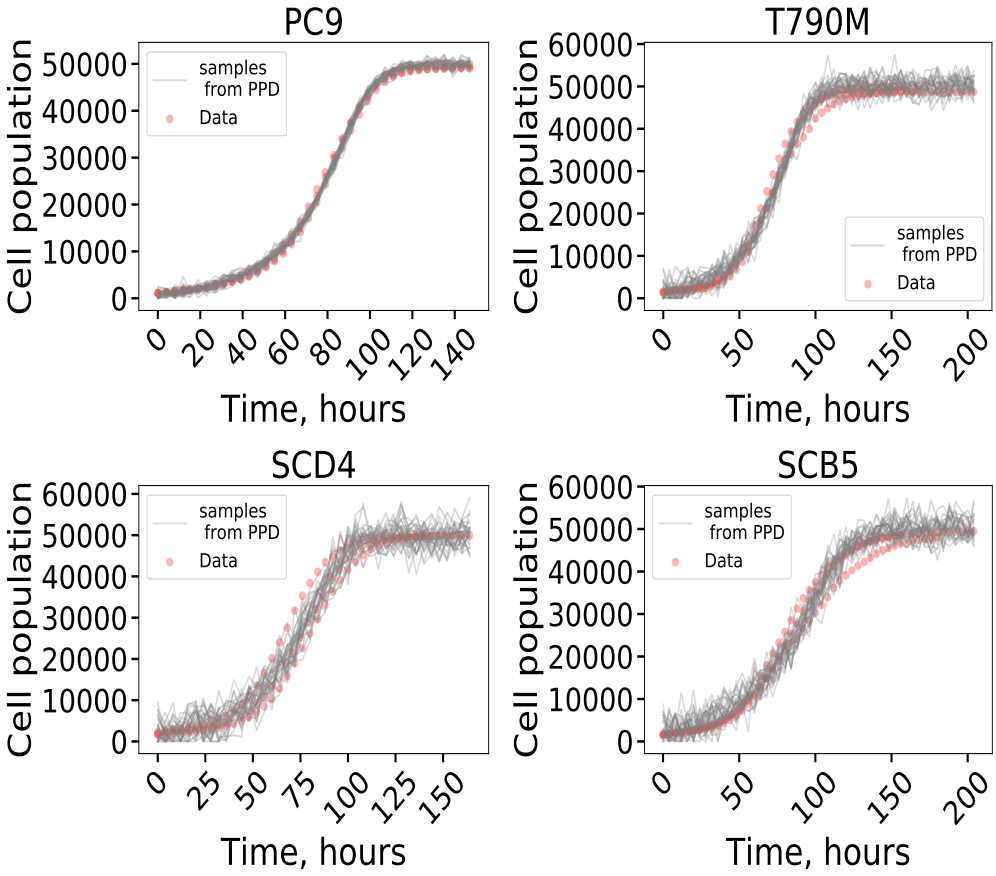


Figure 4.B.4: Visual posterior predictive check (PPC). 30 samples of the posterior predictive distribution are plotted, where random Gaussian noise is added at each time point with mean zero and variance equal to the σ value for a given sample of the Markov chain (see section 4.4 for details on how the PPC is performed). These results demonstrate the model is appropriate for the data for all datasets. Key: red dots; data, grey lines; model solutions sampled from the posterior predictive distribution, with noise.

Chapter 5

Modelling preclinical data of NSCLC response to TKIs

5.1 Introduction

In the previous chapter, we developed ODE models with sigmoid growth dynamics that describe in vitro growth of non-small cell lung cancer (NSCLC) cell lines. We performed model selection using parameter estimation, information criteria/goodness-of-fit metrics and identifiability analysis. We uncovered practical identifiability issues in our selected model using in vitro and synthetic data. By fixing non-identifiable model parameters, we obtained practical identifiability of the model (i.e. all estimated model parameters are identifiable given the in vitro data).

In this chapter, we extend the one-compartment model, developed in chapter 4, to account for treatment with a single tyrosine kinase inhibitor (TKI). The TKIs considered are the first and third-generation TKIs, gefitinib and osimertinib. We propose a two-compartment ODE model which distinguishes live and dead cells. In preclinical experiments, it is common to measure total tumour cell populations/volume over time. However, tumours may consist of subpopulations of cells which differ in their characteristics, for example, live and dead cells. Multi-compartment models can make the distinctions between cell types. When fitting these models to experimental data that describes total cell numbers only, structural and practical identifiability issues may arise. This is because the size of the unobserved cell subpopulations is unknown. Therefore, the ability to infer model parameters may not be possible when fitting such models to the data and, misleading trajectories of cell numbers may be predicted by a model which

consists of unobserved variables [36]. Further, using a practically non-identifiable model to infer drug-effect parameters in a clinical setting can have significant implications.

In this chapter, we apply a range of techniques to our two-compartment model which include, parameter estimation and identifiability analysis. By doing so, we identify the mechanisms of action of the different drugs on the different NSCLC cell lines. We use synthetic data to evaluate how experimental design (i.e. what variables are measured) affects certainty in parameter estimates and, to generate hypotheses about the likely mechanisms of action of the two TKIs on the NSCLC cells. Our approach is similar to that used in chapter 4 and is outlined in Figure 5.1.1. The work presented in this chapter follows a similar framework presented by Eisenberg et al. in [36].

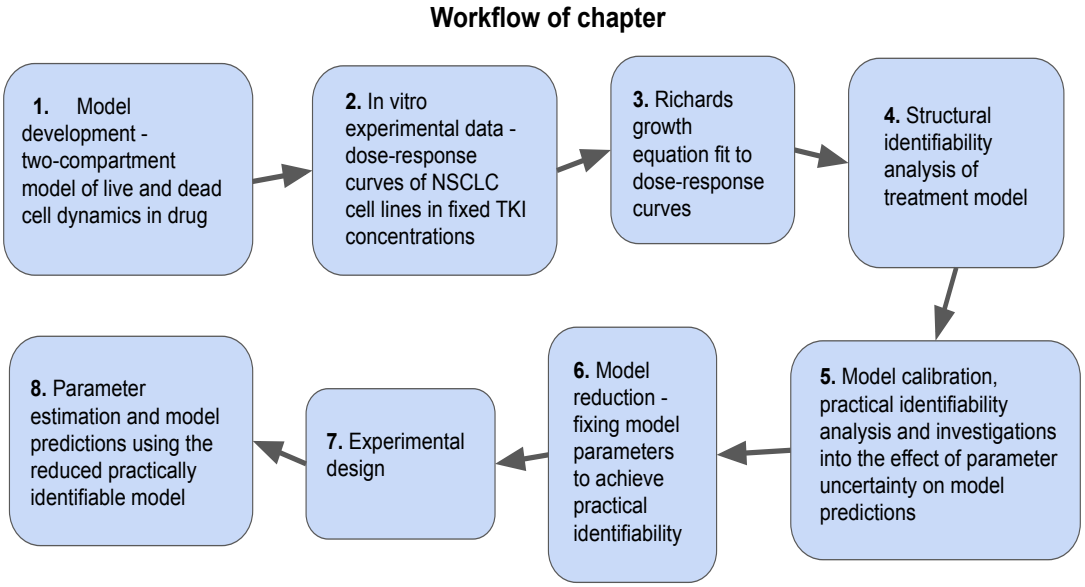


Figure 5.1.1: Schematic of the workflow in this chapter.

We summarise Figure 5.1.1:

1. We begin with model development. We develop a model that describes the response of a single NSCLC cell line to a single TKI.

2. We then describe the in vitro experimental data setup and, we introduce dose-response curves of four NSCLC cell lines when a single TKI is considered.

3. We fit the reduced Richards growth equation (see chapter 4) to individual dose-response curves. We show the reduced Richards growth equation describes the drug-resistant cellular dynamics well for all concentrations of the drug, demonstrating their resistance. Using our results, we suspect a cell kill drug-effect is taking place on the drug sensitive cells as the Richards growth equation does not fit the data well at higher concentrations of drug.

4. We perform structural identifiability analysis on our treatment model and find it to be structurally identifiable.

5. We assess the practical identifiability of our model and demonstrate that practical identifiability issues arise when the model is fit to the in vitro data. We identify practically non-identifiable parameters in the model and, by employing these parameters in the model, demonstrate that misleading model predictions can be made.

6. We then carry out model reduction through fixing non-identifiable model parameters to achieve a practically identifiable model.

7. We propose additional experiments that we hypothesise will increase the practical identifiability of the model and, we test this by generating synthetic data of these experiments.

8. We finish by estimating model parameters using our reduced practically identifiable model and we compare the drug-effects of different TKIs on different NSCLC cell lines.

In section 1.2.4 in chapter 1, we presented the Hill equation that is commonly used to describe dose-response curves and to determine the pharmacological effect of a drug on tumour cells [1]. In this chapter we generalise such models to consider TKI treatments whereby we introduce a model that incorporates cell kill due to drug using a Hill equation term. We also consider the inclusion of a dead cell population and we model tumour growth in the absence of drug using the Richards growth equation (see chapter 4). These

aspects in combination with the experimental data to which we fit the models represents the novel aspects of the modelling developed in this chapter.

The remainder of this chapter is structured as follows. In section 5.2, we present in vitro dose-response curves. Model development takes place in section 5.3. We introduce parameter bounds used for parameter identifiability analysis in section 5.4. We present results in section 5.5. The chapter finishes in section 5.6 where we give a discussion on our findings.

5.2 In vitro dose-response curves

In this section, we extend the in vitro data sets from chapter 4 by introducing the treatment data. In Figures 2.1(a)-(h), dose-response curves for the four cell lines (PC9, T790M, SCD4 and SCB5) to treatment with osimertinib and gefitinib are shown. Each data set corresponds to a particular cell line and a particular drug (for example, the PC9 gefitinib data set describes the set of dose-response curves of PC9 cells for a range of gefitinib concentrations; see Figure 5.2.1(a)).

The PC9 cell line is treatment naïve and is sensitive to osimertinib and gefitinib. The T790M cell line was established in vitro from the PC9 cell line by chronic exposure to vandetanib (tradename Caprelsa), a TKI applied in thyroid cancer [65]. The T790M mutation confers resistance to gefitinib. Using CRISPR, the T790M cell line was modified to produce two mutants, SCD4 and SCB5 with the T790M and C797S mutations. The C797S mutation confers resistance to osimertinib. Hence, SCD4 and SCB5 cell lines with the T790M and C797S mutations are resistant to both gefitinib and osimertinib.

5.2.1 In vitro experiment protocol

The same experiment protocol is applied as in chapter 4 (see section 4.2 in chapter 4). However, in step 2 of the experimental protocol, fixed doses of the TKIs were applied to the new plates before cells were seeded onto them. The concentration of drug is constant throughout these experiments.

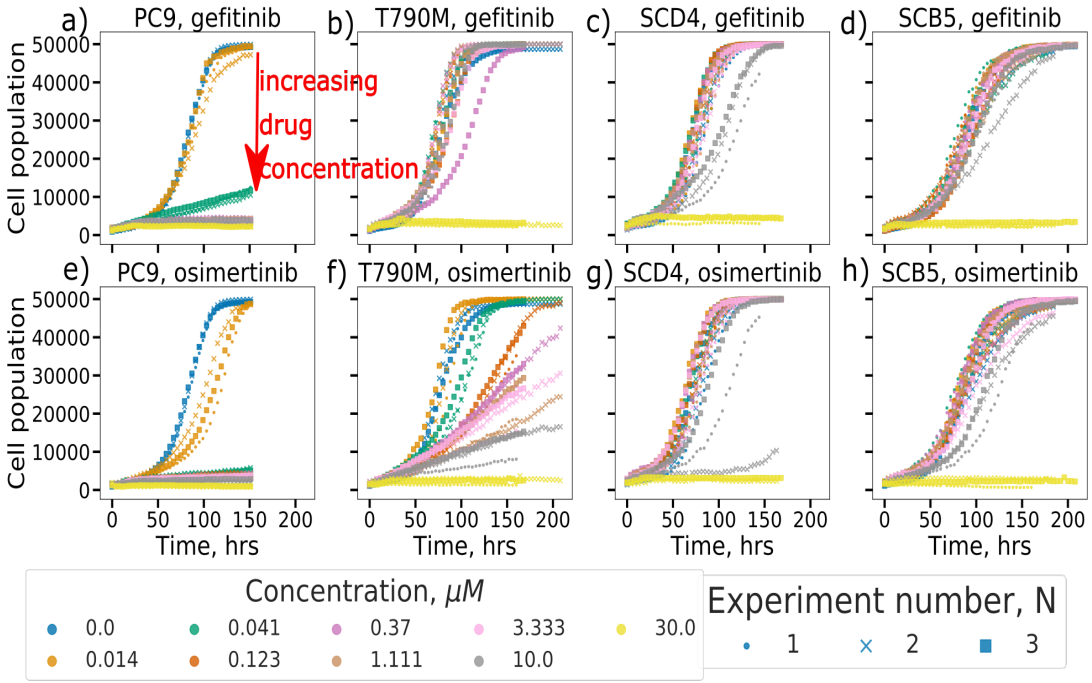


Figure 5.2.1: In vitro treatment data. (a)-(h) In vitro dose-response curves of the 4 cell lines; PC9, T790M, SCD4 and SCB5. Key: colours represent different drug concentrations in micromolar (μM). Shapes represent experiment repeats: circles; $N = 1$, crosses; $N = 2$, squares; $N = 3$. Drug concentration, c is fixed throughout the experiment.

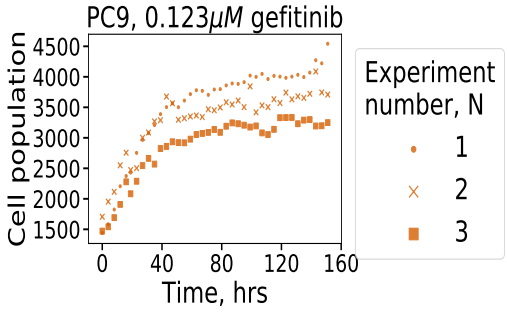


Figure 5.2.2: In vitro dose-response curves of PC9 cells in 0.123μM gefitinib. Key: shapes represent experiment repeats: circles; $N = 1$, crosses; $N = 2$, squares; $N = 3$.

5.2.2 Discussion on in vitro data sets

Let c denote the constant drug concentration in the experiments, with units of micromolar (μM). Figures 5.2.1(a) and (e) show that both gefitinib and osimertinib inhibit the growth of PC9 cells when administered at doses with $c > 0.041$ μM (light orange curves correspond to $c = 0.041$ μM). PC9 cells exposed to doses of osimertinib (see Figure 5.2.1(e)) with $c = 0.041$ μM (light orange curves), grow slower than those exposed

to the same dose of gefitinib (see Figure 5.2.1(a)), suggesting that osimertinib has a stronger effect on PC9 cells at this concentration.

Figure 5.2.2 shows the dose-response curves of PC9 cells in gefitinib 0.12μM. Similar dynamics are observed for PC9 cells in gefitinib $c > 0.041$ μM and osimertinib $c > 0.014$ μM (see Figures 5.2.1(a) and (e)) as in Figure 5.2.2.

SCD4 and SCB5 cells with the resistance-conferring mutations, T790M and C797S, exhibit the same saturating growth dynamics as their respective controls (blue curves) for $c < 30\mu$M (see Figure 5.2.1(c), (d), (g) and (h)) and hence, are resistant to both drugs. The same applies to T790M cells in gefitinib (see Figure 5.2.1(b)). For drug-resistant cells, at 30 μM gefitinib and osimertinib (yellow curves), the growth dynamics are similar to those described in Figure 5.2.2. This may be due to non-specific toxicity of the drug-resistant cells which causes total tumour cell arrest. For these reasons, we do not consider the data relating to gefitinib and osimertinib concentrations of 30 μM for the remainder of this chapter.

T790M cells are sensitive to osimertinib (see Figure 5.2.1(f)) yet present different dynamics to the drug sensitive PC9 cells (see Figure 5.2.1(e)). For $c > 0.014\mu$M, maximum inhibition of PC9 cells is observed whereas T790M cell growth is slowed. This suggests that osimertinib may have a lower efficacy for T790M cells. Alternatively, osimertinib may have different mechanisms of action when applied to T790M cells and PC9 cells.

We observe variability across experimental repeats for a given cell line and drug concentration. Possible sources of variability include variability in the cells themselves, fluctuations in culture media and variability in culture conditions. The process through which drug-resistant cell lines are derived (the T790M cell line is derived from the PC9 cell line through chronic exposure to the TKI, vandetanib and SCD4 and SCB5 cell lines were established by modifying the T790M cell line using CRISPR) may also contribute to increased heterogeneity across cells of a particular type. In Figures 5.2.1(b) and (g), extreme variability is observed in the following data sets: T790M in gefitinib 0.37 μM, repeat 3, SCD4 in osimertinib 10 μM, repeat 2. We therefore do not consider these experimental repeats for the remainder of this chapter. We focus on the PC9 gefitinib and osimertinib data sets and the T790M osimertinib data set as the drug sensitive data

sets, together with SCD5 and SCB5 data sets and the T790M gefitinib data set as the drug-resistant data sets.

5.3 Model development

We develop a two-compartment ODE model that distinguishes live and dead cells. The model describes the response of a single NSCLC cell line to a single TKI which is fixed at a constant drug concentration, c where c (μM) is the constant drug concentration in the experiments. Since the in vitro data sets record only total tumour cell numbers, when fitting our model to the in vitro data, we fit the total tumour cell population (live and dead cells combined).

The model variables are as follows:

- $y(t)$ denotes the number of live tumour cells, at time t,
- $z(t)$ denotes the number of dead tumour cells,
- $N(t)$ denotes the total population of tumour cells, $N(t) = y(t) + z(t)$.

The governing equations are as follows:

$$\overbrace{\frac{dy}{dt}}^{\text{live cells}} = \underbrace{ry\left(1 - \left(\frac{y+z}{K}\right)^{\beta_F}\right)}_{\text{Richards growth equation with competition between live and dead cells}} - \overbrace{h(c)y}^{\text{drug-induced cell kill}}, \tag{5.3.1}$$

$$\overbrace{\frac{dz}{dt}}^{\text{dead cells}} = h(c)y - \underbrace{kz}_{\text{degradation of dead cells}}, \tag{5.3.2}$$

where,

$$h(c) = \frac{\lambda c^m}{IC50^m + c^m}. \tag{5.3.3}$$

In equation (5.3.1), parameter r (in units of hour^{-1}) denotes the growth rate of the live cells. We assume live and dead cells compete for space in the reduced Richards growth equation with fixed β_F (see chapter 4) and K (units of cells) denotes the carrying

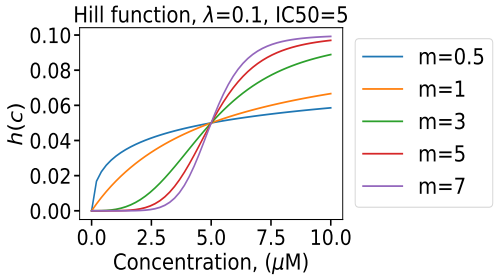


Figure 5.3.1: Hill equation dynamics for a range of values of the Hill coefficient, m (equation 5.3.3). Parameters: $\lambda = 0.1$, $IC50 = 5$.

capacity. We introduce a cell death term that describes drug-induced cell kill of the live cells. The rate at which the live cells are killed by the drug is proportional to the number of live cells, $y(t)$, and the rate of cell kill, $h(c)$ (in units of hour^{-1}), which is an increasing, saturating function of the drug concentration, c (units μM). We model $h(c)$ using a Hill equation (see equation (5.3.3)) which we provide more details of below. The cell kill term in equation (5.3.1) acts as a source of dead cells in equation (5.3.2). We assume further that dead cells degrade at constant rate, $k \geq 0$ (in units of hour^{-1}). In the absence of drug ($c = 0$), the growth of the live cells is described by the reduced Richards growth equation model (see chapter 4).

Equation (5.3.3) describes the Hill equation and we illustrate its dynamics in Figure 5.3.1. In equation (5.3.3), the parameter m is the Hill coefficient. It determines the steepness of the Hill equation (see Figure 5.3.1) and λ (in units of hour^{-1}) denotes the maximum rate of cell kill due to drug. The parameter $IC50$ (units μM) denotes the drug concentration required to achieve 50% maximum drug-effect on the cells. As $c \to \infty$, $h(c) \to \lambda$. As $IC50 \to 0$, the cell's sensitivity to drug increases and for $IC50 \to \infty$, cells are more resistant.

For a fixed drug concentration, the Hill equation (see equation (5.3.3)) can be described as a constant, $\tilde{h}$, where

$$\tilde{h} = \frac{\lambda c^m}{IC50^m + c^m}. \tag{5.3.4}$$

In order to determine the relationship between drug concentration and the rate of cell kill due to drug, we fitted our model, defined in equations (5.3.1)-(5.3.2) to individual dose-response curves and estimated values of y_0, r, k and $\tilde{h}$, for each drug concentration.

We found the estimated values of $\tilde{h}$ to follow a saturating behaviour for increasing values of c. In response to this, we fitted the Hill equation (5.3.3) to estimated $\tilde{h}$ values and found the Hill equation to describe the data well with values of $m \neq 1$ for the majority of data sets. With estimates of $m \neq 1$, we concluded that simpler models such as Michaelis-Menten were not suitable. Hence, we consider Hill equation dynamics when describing the cell kill effect of the drugs on the cells.

5.3.1 Initial conditions

We assume that initially the tumour comprises live cells only, so that,

$$y(t = 0) = y_0, z(t = 0) = 0, \tag{5.3.5}$$

where $y_0 > 0$.

We assume natural cell death is negligible during the course of the experiments and therefore assume dead cells do not occur naturally. Hence, at the beginning of the experiment, dead cells are not present within the total tumour population. For the control case, $c = 0$, dead cells are not present within the total tumour population for the duration of the experiment.

To summarise, our model of cell growth and response to treatment with a single TKI is defined by equations (5.3.1)-(5.3.3), with initial conditions prescribed in equation (5.3.5). A summary of the model parameters is given in Table 5.4.2.

5.4 Methods

In this section, we restate the estimated growth rates, r, from chapter 4 of the four NSCLC cell lines. We also present biologically motivated bounds of model parameters defined in equations (5.3.1)-(5.3.3) and we present a Monte Carlo Simulation Method, used to quantify uncertainty in model fits to the in vitro data. We use the same parameter estimation methods, goodness-of-fit metrics and identifiability analysis methodologies that we used in chapter 4. When performing structural identifiability analysis, we use a differential algebra approach which we implement using DAISY [8]. We implement a differential algebra approach to perform structural identifiability analysis on the model

for a single, fixed drug concentration, c as the model is a system ODEs in the form of rational and polynomial functions. To assess the structural identifiability of the model over multiple drug concentrations, we employ a scaling method presented by Castro and de Boer [18].

We use the extended Fourier Amplitude Test (eFAST) to perform global parameter sensitivity analysis on the model (see chapter 2 for details of eFAST). When fitting our model to a dose-response curve, we pool the data from the three experimental replicates for each cell line and each drug concentration.

5.4.1 Parameter bounds

In Table 5.4.1, we state estimated values and standard errors of the growth rates, r, for the four NSCLC cell lines. These values were obtained in chapter 4 by fitting the reduced Richards equation, with fixed β, to the control in vitro data sets.

	PC9	**T790M**	**SCD4**	**SCB5**
r	0.0410±0.0004	0.0440±0.0007	0.0382±0.0010	0.0341±0.0005

Table 5.4.1: Estimated values (± standard error) of the growth rate, r, for each cell line in the absence of drug. Estimates were found by fitting the reduced Richards growth equation, with fixed $\beta = 3.315$, to the control in vitro data sets in chapter 4.

In Table 5.4.2, we provide descriptions of model parameters defined in equations (5.3.1)-(5.3.3), and the bounds employed for parameter estimation and practical identifiability analysis.

In what follows, we motivate the parameter bounds stated in Table 5.4.2.

Parameter bounds on the growth rate, r (hour^{-1}), are based on the likelihood-based confidence intervals for PC9 cells, calculated in chapter 4. We take the conservative bounds of $0.015 < r < 0.06$.

Initial cell numbers in the in vitro data (see Figures 5.2.1(a)-(h)) range from 1000 to 2250 across experiments. We take the conservative bounds, $100 < y_0 < 2500$.

Bounds on the half-maximal drug concentration, $IC50$ (μM), are taken from in vitro experimental studies on EGFR-mutant NSCLC cell lines, with recorded values of $IC50 = 0.023\mu$M and $IC50 = 0.008\mu$M, for gefitinib and osimertinib, respectively [29]. We take the conservative bounds $0.005 < IC50 < 0.2$.

Parameter	Description	Units	Parameter range
r	Growth rate	hour^{-1}	[0.015, 0.06]
k	Dead cell degradation rate	hour^{-1}	[0.00001, 0.15]
y_0	Initial live cell number	cells	[100, 2500]
λ	Maximum rate of cell death due to drug	hour^{-1}	[0.01, 0.15]
$IC50$	Half-maximal drug concentration	μM	[0.005, 0.2]
m	Hill coefficient	dimensionless	[0.5, 5]
K	Carrying capacity	cells	fixed; 50000 cells
β_F	Competition strength parameter	dimensionless	fixed; 3.315

Table 5.4.2: Summary of estimated and fixed parameters for the in vitro treatment model defined by equations (5.3.1)-(5.3.3). Values of fixed parameters and parameter bounds considered for estimated model parameters are given along with their units. Except for the carrying capacity, K which is fixed at 50,000 cells and the competition strength parameter β_F which is also fixed ($\beta_F = 3.315$), all model parameters are estimated by fitting to the in vitro data sets. Bounds on the parameters are used for parameter estimation and the calculation of profile likelihoods (see subsection 5.5.4.2).

Bounds on the Hill coefficient, m, can also be inferred from in vitro experimental data [137]. Vainstein et al. fitted the Hill equation to dose-response curves describing chronic myeloid leukaemia (CML) BCR-ABL mutant cells in the BCR-ABL TKI, nilotinib and found estimated slopes of the curve (m) in the range $0.94 < m < 5.64$. The authors found that lower values of m were found to correspond to cells with drug resistance mechanisms. We take the bounds, $0.5 < m < 5$.

Bounds on the maximum rate of drug-induced cell kill, λ, are taken to be $0.01 < \lambda < 0.15$. This range of values is chosen so that, at low values of λ (e.g. $\lambda = 0.01$ hour^{-1}), the maximum rate of cell death is slower than the growth rate. By contrast, when $\lambda = 0.15$ hour^{-1}, the rate of cell kill due to drug is larger than the rate of cell growth.

Bounds on the rate of dead cell degradation, k (units hour^{-1})are taken to be, $0.00001 < k < 0.15$. This range is chosen so that, at low values of k, (e.g. $k = 0.00001$ hour^{-1}), degradation of dead cells is negligible during the experiment. Conversely, for high values of k (e.g. $k = 0.15$ hour^{-1}), dead cells degrade much quicker than the growth rate of live cells.

5.4.2 Monte Carlo Simulations for calculating 95% confidence bounds of the model fit to data

We use Monte Carlo Simulations to generate 95% confidence bands of the fits of our practically identifiable model to the in vitro data. For these simulations, we use values of model parameters that lie within their likelihood-based confidence intervals, calculated using the profile likelihood method. We note that, in the previous chapter, we used MCMC to quantify the uncertainty in model fits.

1. For practically identifiable model parameters of which finite likelihood-based confidence bounds are yielded, we define a uniform distribution with bounds equal to those of the likelihood-based confidence intervals.

2. For each model parameter, we randomly sample from the uniform distribution, defined in step 1. Let θ^* denote the sample of model parameters.

3. We employ the set of parameters, θ^*, in our model and calculate the residual sum of squared errors, $\chi^2(\theta^*)$, by fitting the model to the pooled dose-response curves for a given cell line and drug (we note that all model parameters are fixed and hence, no parameters are estimated when doing so). If $\chi^2(\theta^*) - \chi^2(\hat{\boldsymbol{\theta}}) < icdf(\chi^2_{df}, \alpha)$, we accept θ^*,

   else: we return to Step 2.

   Recall, $\chi^2(\hat{\boldsymbol{\theta}})$ denotes the goodness-of-fit and $icdf(\chi^2_{df}, \alpha)$ is the inverse cumulative distribution function of the χ^2 distribution with df degrees of freedom at the 95% confidence level, where df is equal to the number of model parameters.

4. Using our accepted θ^*, we simulate the model for all drug concentrations and, for each drug concentration, we store solutions of the total, live and dead cell numbers. We repeat then return to step 1 until a total number of 5000 iterations is completed.

5. Using the sets of stored model solutions, for a given drug concentration, at each time point, we calculate the maximum and minimum values of the distribution of total cell numbers. We do the same for the distribution of live cell numbers and

similarly for the dead cell numbers. The maximum and minimum values of the total, live and dead cells for all time points are the upper and lower limits of our 95% confidence region of the model fit to the data.

5.5 Results

In this section, we present results in line with the workflow presented in Figure 5.1.1.

5.5.1 Richards growth equation fit to individual dose-response curves

We aim to determine how well the reduced Richards growth equation model (see equation (4.3.2) in Table 4.3.1 in chapter 4), with fixed $\beta = \beta_F$ ($\beta_F = 3.315$), describes the dynamics of NSCLC TKI-sensitive and resistant cell lines in a single fixed concentration of drug. In so doing, we investigate how the TKIs affect the dynamics of the cells compared to the control.

We begin by fitting the reduced Richards growth equation to each dose-response curve describing drug sensitive PC9 cells and drug-resistant SCD4 cells in a fixed gefitinib concentration (see Figures 5.2.1(a) and (c)). We estimate values of parameters r and y_0 for each drug concentration and each cell line. By doing so, we determine how the drug affects a cell's growth rate, r. In Figures 5.5.1 and 5.5.2, we present solutions when the model is fitted to the data for a range of values of c. In Figure 5.5.3, we plot the estimated values of r for all values of c and the corresponding goodness-of-fit, denoted by $\chi^2(\hat{\boldsymbol{\theta}})$.

For PC9 cells in gefitinib, by visual inspection, Figures 5.5.1(a)-(b) show the reduced Richards growth equation with fixed β gives a good description of the cellular dynamics for $c \leq 0.014\mu$M. However, for $c > 0.041\mu$M, poorly fitting solutions of the model are shown (Figures 5.5.1(c)-(d)). Although not presented here, poor fits were also observed for larger values of the drug concentration, $c > 0.123\mu$M, for PC9 cells in gefitinib. In Figure 5.5.3(a), for the PC9 gefitinib data set, when $c = 0.014\mu$M, a lower value of $r = 0.025$ hour^{-1} is estimated compared to the untreated case ($c = 0\mu$M) where $r = 0.041$ hour^{-1} (see Table 5.4.1). These results could suggest that gefitinib slows the growth rate of the PC9 cells at low drug concentrations and that, at higher concentrations

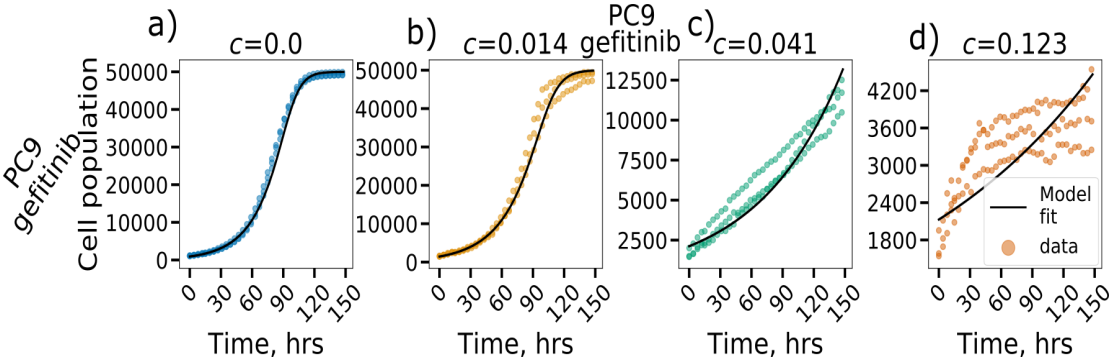


Figure 5.5.1: Best-fit solutions of the reduced Richards growth equation, with fixed $\beta = \beta_F$ (see equation (4.3.2) in Table 4.3.1 in chapter 4), fitted to drug sensitive PC9 cells in gefitinib. For each drug concentration, c, parameters r and y_0 are estimated. Key: coloured dots; data, black lines; model solution describing $y(t)$ over time. Values of c are indicated in the subplot titles.

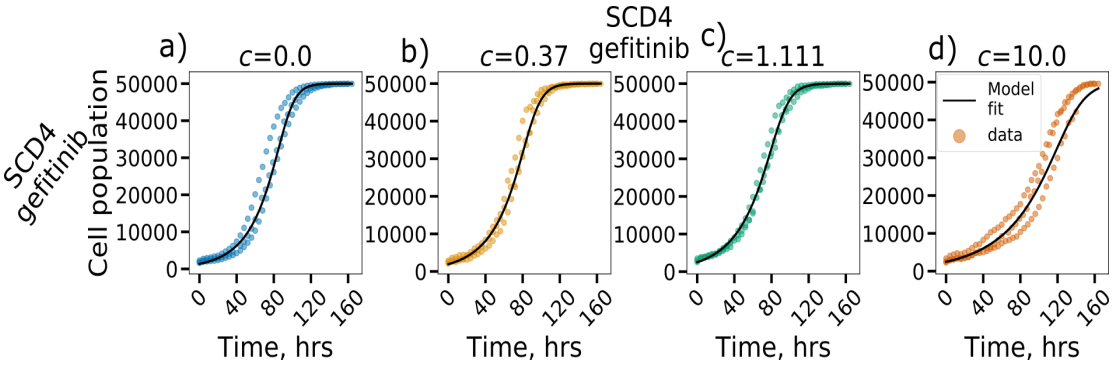


Figure 5.5.2: Best-fit solutions when the reduced Richards growth equation, with fixed $\beta = \beta_F$ (see equation (4.3.2) in Table 4.3.1 in chapter 4), is fitted to drug sensitive SCD4 cells in gefitinib. Figure details the same as in Figure 5.5.1.

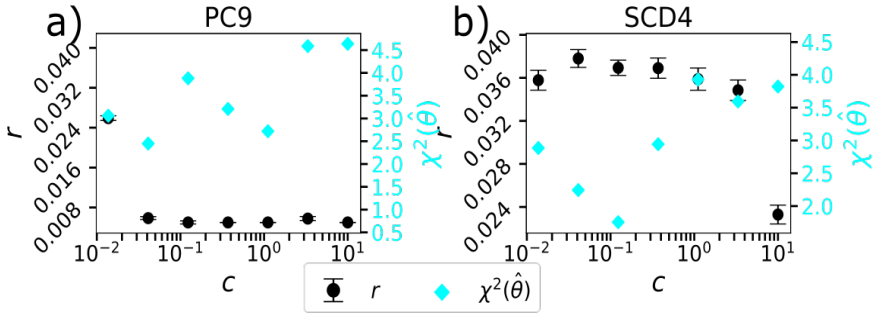


Figure 5.5.3: Estimated values of r (left hand y-axis) and corresponding minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ (right hand y-axis) plotted with respect to c for PC9 and SCD4 gefitinib data sets when fitting the reduced Richards growth equation, with fixed β, to individual dose-response curves. Key: black dots; estimated values of r and bars represent the standard error of each estimate, cyan diamonds; $\chi^2(\hat{\boldsymbol{\theta}})$ values.

of drug, a drug-induced cell kill mechanism is acting. Conversely, the drug may act by inducing a cell kill mechanism alone. We discuss these alternatives in later sections

where we propose experiments that may be used to elucidate our understanding further of the mechanisms of drug. Further, we use the model to generate synthetic data that describe these experiments.

For the drug-resistant SCD4 cells in gefitinib, by visual inspection, Figures 5.5.2(a)-(d) shows that the reduced Richards growth equation fits the in vitro data for concentrations ranging from 0 to 10 μM.

In Figure 5.5.3(b), we plot estimated values of r for SCD4 cells in gefitinib. The values of r are similar ($0.035 < r < 0.038$ for $0.014 < c < 3.33$) to those in the untreated case where $r \approx 0.038$ hour^{-1} (not plotted but given in Table 5.4.1). A large reduction in the estimated value of r is observed for $c = 10\mu$M (estimated value of $r = 0.022$ hour^{-1} at $c = 10\mu$M). Hence, increasing doses of gefitinib in the range $c < 10\mu$M, have a small effect on the growth rate of the SCD4 cells and we observe that these cells are resistant as estimated values of r do not vary considerably from the untreated case.

Similar results to those presented in Figure 5.5.3(a) are observed for the PC9 osimertinib data set (see Figure 5.A.1(a) in the appendix) and similar results to those in Figure 5.5.3(b) are observed for the resistant data sets (both SCB5 data sets and the T790M gefitinib data set, see Figure 5.A.1 in the appendix). For the T790M cells in osimertinib, the Richards growth equation gives a good description of the the dose-response curves at higher concentrations ($c \leq 0.123\mu$M) than the PC9 cells and, a decrease in estimated r values is observed (see Figure 5.A.1(c) in the appendix). This suggests osimertinib has a weaker effect on T790M cells than PC9 cells. We investigate this further in the following sections.

For the remainder of this chapter, we focus on the drug sensitive data sets (PC9 gefitinib, PC9 osimertinib and T790M osimertinib) and we explore in more detail the mechanisms by which the drugs act on the different cell lines. We do not carry out further analysis on the drug-resistant data sets (SCD4 and SCB5 data sets and the T790M gefitinib data set). We propose future work in identifying the mechanisms by which drug resistance mutations, such as those of the drug-resistant cells presented in this chapter, reduce treatment efficacy. For example, do such mutations prevent the action

of drugs that slow the growth of the cells, or do they evade treatment by preventing a cell kill mechanism of the drug.

5.5.2 Steady-state analysis

For a fixed drug concentration, c where the rate of cell kill is denoted by the constant, $\tilde{h}$, the steady states of the model are $(0, 0)$ and $(y^*,\ z^*)$ where,

$$y^* + z^* = K\left(1 - \frac{\tilde{h}}{r}\right)^{\frac{1}{\beta_F}}, \tag{5.5.1}$$

with

$$y^* = \frac{K\left(1 - \frac{\tilde{h}}{r}\right)^{\frac{1}{\beta_F}}}{\left(1 + \frac{\tilde{h}}{k}\right)}, \tag{5.5.2}$$

and

$$z^* = \frac{\tilde{h}y^*}{k}. \tag{5.5.3}$$

Therefore, referring to the steady state of the dead cells, z^*, the composition of the tumour at steady state depends on the rate of cell kill, $\tilde{h}$ and the rate of dead cell degradation, k. For $\tilde{h} < k$, the tumour is made up of more live cells than dead cells at steady state and for $\tilde{h} > k$, the tumour is made up of more dead cells than live cells at steady state. If $\tilde{h} = k$, the tumour is made up of live and dead cells with a 1:1 ratio at steady state. The total tumour population size at steady state depends on the parameters $\tilde{h}$ and r. If $\tilde{h} = 0$, the total tumour population size at steady state is equal to K. In order for the steady state to be biologically feasible, we require $r \geq \tilde{h}$. By carrying out this steady state analysis, we have found constraints on parameters that are necessary for biologically feasible solutions and we have determined how tumour composition at steady state will vary depending on the model parameter values.

5.5.3 Structural identifiability

We perform structural identifiability analysis on the treatment model (defined in equations (5.3.1)-(5.3.3)) for a single, fixed drug concentration, c, using a differential algebra approach and we implement this approach using DAISY, a software package used to determine the structural identifiability of ODE models with polynomial and rational

functions [8]. For multiple drug concentrations, we perform structural identifiability analysis on the model using a scaling method presented by Castro and de Boer [18].

We assess structural identifiability from time-course data on total tumour cell numbers (i.e., $N(t) = y(t) + z(t)$) in response to, (i) a single fixed drug concentration, c (i.e. analysis on a data set that describes a single dose-response curve of a cell line in a fixed concentration of drug); (ii) multiple (≥ 2) drug concentrations (i.e. analysis on a data set consisting of pooled dose-response curves where more than two drug concentrations are considered).

For (i), we note that for any single dose-response curve, the drug concentration, c is constant. Hence, the Hill equation (equation 5.3.3) can be represented by a constant, $\tilde{h}$, where,

$$\tilde{h} = \frac{\lambda c^m}{IC50^m + c^m}. \tag{5.5.4}$$

Therefore, the set of model parameters for a single dose-response curve, of which we apply structural identifiability analysis, are $\{r,\ \tilde{h},\ k,\ y_0\}$. For (i), DAISY confirmed global structural identifiability of all parameters $\{r,\ \tilde{h},\ k,\ y_0\}$ for any single dose-response time-course of the total tumour population.

In order to demonstrate structural identifiability of the model for multiple drug concentrations, we need to show that the parameters, $\{r, \lambda, IC50, m, k\}$ are identifiable for all drug concentrations, c_j. To do this, we implement a simple scaling method presented by Castro and de Boer [18]. For ease of understanding we use I in place of $IC50$.

Recall $N(t) = y(t) + z(t)$ denotes the observed variable describing the total tumour population over time. Hence, we rewrite the treatment model (defined in equations (5.3.1)-(5.3.3)) as the following system of equations,

$$\frac{dy}{dt} = ry\left(1 - \left(\frac{N}{K}\right)^{\beta_F}\right) - \frac{\lambda c^m y}{c^m + I^m}, \tag{5.5.5}$$

$$\frac{dN}{dt} = ry\left(1 - \left(\frac{N}{K}\right)^{\beta_F}\right) - kN + ky, \tag{5.5.6}$$

where $N(t)$ is observed, $y(0) = y_0$ and $z(0) = 0$, where y_0 is a known, fixed parameter.

Identifiability equations

Consider the scalings,

$$\begin{cases} \lambda \to u_\lambda \lambda, \\ I \to u_I I, \\ m \to u_m m, \\ r \to u_r r, \\ k \to u_k k, \\ y \to u_y y, \\ N \to N, \\ c \to c, \\ K \to K, \\ b_F \to b_F. \end{cases} \tag{5.5.7}$$

Since the parameters, β_F, K and c are known, $u_c = 1$, $u_{\beta_F} = 1$ and $u_K = 1$. Further, $u_N = 1$ as N is observed.

We can define the linear independent functions:

$$f_{11} = ry\left(1 - \left(\frac{N}{K}\right)^{\beta_F}\right),\ f_{12} = \frac{\lambda c^m y}{c^m + I^m}, \tag{5.5.8}$$

$$f_{21} = f_{11},\ f_{22} = kN,\ f_{23} = ky. \tag{5.5.9}$$

From equation (14) in [18],

$$ry\left(1 - \left(\frac{N}{K}\right)^{\beta_F}\right) = u_r r u_y y\left(1 - \left(\frac{N}{K}\right)^{\beta_F}\right) \tag{5.5.10}$$

$$\frac{\lambda c^m y}{c^m + I^m} = \frac{u_\lambda \lambda c^{u_m m} u_y y}{c^{u_m m} + (u_I I)^{u_m m}} \tag{5.5.11}$$

$$kN = u_k k N \tag{5.5.12}$$

$$ky = u_k u_y k y \tag{5.5.13}$$

From equation (5.5.12), $u_k = 1$ and therefore, from equation (5.5.13), $u_y = 1$. For equation (5.5.10), $u_r u_y = u_r = 1$. Consider equation (5.5.11) and multiply both sides by $(c^m + I^m)(c^{u_m m} + (u_I I)^{u_m m})$ to obtain,

$$\lambda c^m y(c^{u_m m} + (u_I I)^{u_m m}) = u_\lambda \lambda c^{u_m m} u_y y(c^m + I^m), \tag{5.5.14}$$

giving,

$$c^{(u_m+1)m} + (u_I I)^{u_m m} c^m = u_\lambda u_y c^{(u_m+1)m} + u_\lambda u_y I^m c^{u_m m}. \tag{5.5.15}$$

By comparing the coefficients of $c^{(u_m+1)m}$ and $c^{u_m m}$, we find $u_\lambda u_y = u_\lambda = 1$ (given $u_y = 1$ from above) and $u_m = u_I = 1$ to be a solution. To check this is the only solution, assume $u_m \neq 1$. By comparing coefficients in equation (5.5.15), we get $u_\lambda = 1$ and

$$(u_I I)^{u_m m} c^m = I^m c^{u_m m}. \tag{5.5.16}$$

where $u_m = u_I = u_\lambda = 1$ is the only solution.

Hence, the identifiability equations are

$$\begin{cases} u_\lambda = 1, \\ u_I = 1, \\ u_m = 1, \\ u_r = 1, \\ u_k = 1, \\ u_y = 1, \\ u_N = 1, \\ u_c = 1, \\ u_K = 1, \\ u_{b_F} = 1. \end{cases} \tag{5.5.17}$$

Since all scalings are unity, the model is structurally identifiable. We note that, in the case where live ($y(t)$) and dead ($z(t)$) tumour cell populations are observed, the model is structurally identifiable, although we do not present the workings here.

5.5.4 Parameter estimation and practical identifiability

5.5.4.1 Parameter Estimation

We carry out an initial estimation of the model parameters, prior to practical identifiability analysis, by fitting our model to the pooled dose-response curves for a range of fixed drug-concentrations, c, for a given cell line and drug. When we fit the model to the pooled dose-response curves, we estimate parameters r, $IC50$, λ, m and k. Before performing a fit on the pooled data set, we estimate and fix the initial live cell number, y_0, for each dose-response curve in the pooled data set. We do this by first fitting the model to the individual dose-response curves and estimating values of r, y_0, k and $\tilde{h}$, where $\tilde{h}$ is the constant rate of cell kill due to drug (see equation 5.5.4 in section 5.5.3). We estimate and fix values of y_0 for individual dose-response curves first because different dose-response curves have different initial cell numbers. We summarise the estimated values of y_0 in Table 5.A.1 in the appendix. In Table 5.A.1 in the appendix, for each

data set, there is a wide range of estimated values of y_0 across the different drug concentrations. Variation in the estimated values of y_0 are as a result of the variation in the initial cell numbers in the in vitro data which occurs due to the fact that, for each drug concentration and experimental repeat, a different experiment was carried out where a different initial cell number was seeded at the start of the experiment. Differences in y_0 values will impact the estimated values of the growth rate, r, and other model parameters as we have shown in section 4.5.5 in chapter 4 by calculating correlations between estimated parameters. This variation in the estimated values of y_0 across the different drug concentrations will increase the level of uncertainty in estimated model parameters when fitting the model to pooled dose-response curves.

Model fits for PC9 cells in gefitinib (pooled dose-response curves) are shown in Figure 5.5.4. We plot the total, live and dead cell dynamics. However, when performing the fits, we fitted the total tumour population to the data only. We summarise optimal estimated parameter values in Table 5.5.1 where we also present optimal parameters for PC9 and T790M osimertinib data sets. Model fits for these data sets are plotted in Figures 5.A.2 and 5.A.3, in the appendix. Correlations between estimated parameters are given in Tables (5.5.2)-(5.5.4) for all data sets.

Parameter	PC9 gefitinib	PC9 osimertinib	T790M osimertinib
r	0.041±0.0003	0.041±0.0004	0.043±0.0006
$IC50$	0.026±0.0004	0.015±0.0002	0.033±0.0034
m	2.23±0.05	2.59±0.09	0.50±0.0012
λ	0.041±0.0004	0.041±0.0005	0.037±0.0009
k	0.032±0.0016	0.035±0.0024	0.020±0.0033
$\chi^2(\hat{\boldsymbol{\theta}})$	11.45	25.45	53.20

Table 5.5.1: Estimated parameter values (± standard error) and $\chi^2(\hat{\boldsymbol{\theta}})$ values of the treatment model fitted to pooled dose-response curves of the drug-sensitive data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

Figure 5.5.4 shows the model fits the PC9 gefitinib data well for all concentrations of drug except for $c = 0.041\mu$M. For $c > 0.123\mu$M, live cell numbers are predicted to decrease and dead cell numbers to increase for increasing drug concentration. For $c = 0.041\mu$M, the model predicts slow, nonlinear growth of the cell population whereas the data shows the cells to increase linearly with time. In section 5.5.1, at lower values

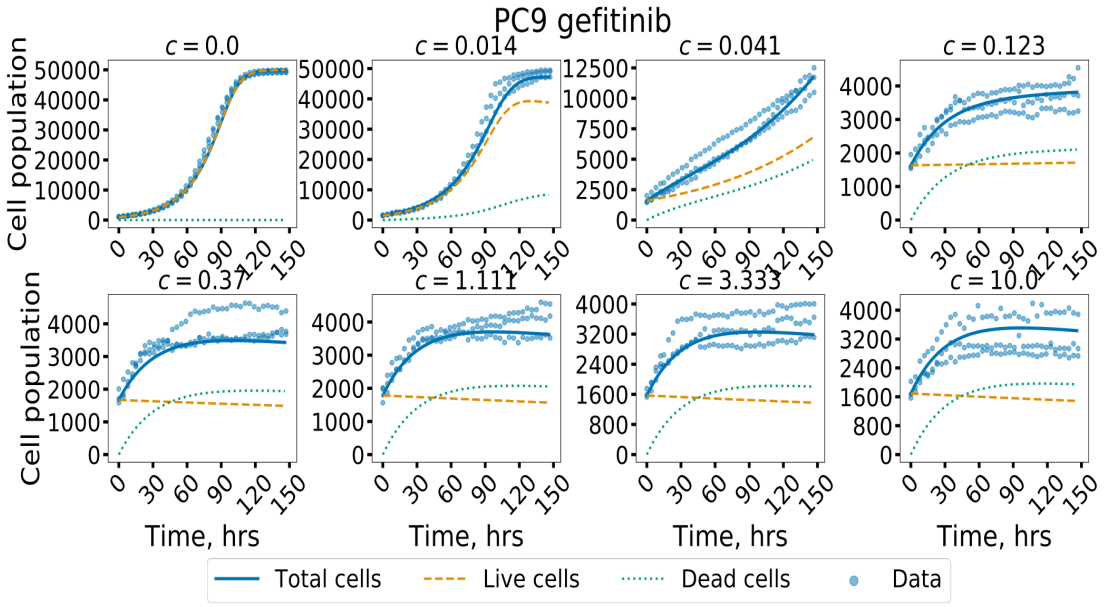


Figure 5.5.4: Model fits for the PC9 gefitinib data set. The model is fitted to pooled dose-response curves for all drug concentrations. Key: blue dots; data, blue block lines; best-fit solution of model describing the total population over time, $N(t)$, orange dashed lines; best-fit solution describing live cell numbers, $y(t)$; green dotted lines; best-fit solution describing dead cell population, $z(t)$. The drug concentrations, c are indicated in each subplot title.

PC9 gefitinib		r	$IC50$	m	λ	k
	r	1.00	-0.69	-0.46	0.57	0.28
	$IC50$	-0.69	1.00	0.62	-0.25	-0.34
	m	-0.46	0.62	1.00	-0.39	-0.04
	λ	0.57	-0.25	-0.39	1.00	-0.60
	k	0.28	-0.34	-0.04	-0.60	1.00

Table 5.5.2: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

PC9 osimertinib		r	$IC50$	m	λ	k
	r	1.00	-0.65	-0.13	0.63	0.29
	$IC50$	-0.65	1.00	0.02	-0.28	-0.29
	m	-0.13	0.02	1.00	-0.21	0.02
	λ	0.63	-0.28	-0.21	1.00	-0.52
	k	0.29	-0.29	0.02	-0.52	1.00

Table 5.5.3: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 osimertinib data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

T790M osimertinib		r	$IC50$	m	λ	k
	r	1.00	-0.70	-0.23	0.54	0.36
	$IC50$	-0.70	1.00	0.20	-0.27	-0.21
	m	-0.23	0.20	1.00	-0.67	0.05
	λ	0.54	-0.27	-0.67	1.00	-0.41
	k	0.36	-0.21	0.05	-0.41	1.00

Table 5.5.4: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the T790M osimertinib data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

of c, when fitting the Richards growth equation to individual dose-response curves, we estimated slower growth of the drug-sensitive cells whereas, at higher values of c, we speculated that a different mechanism of action is taking place. Based on these results, we speculated that a dual mechanism of action could be taking place (i.e. slowing of growth at low concentrations and cell kill at high concentrations). A dual action of TKIs has been observed in the preclinical data study [150]. The cellular dynamics at $c = 0.041\mu$M could be as a result of these mechanisms acting simultaneously. We do not however consider a model whereby the drug affects the growth rate and also induces a cell kill effect. Investigations into a model with said factors is subject to future work.

In Figure 5.5.4, for $c \geq 0.123\mu$M, the qualitative dynamics of the live and dead cell numbers are the same, i.e. live cell numbers remain approximately constant and dead cell numbers increase and saturate. Similar results were observed for the PC9 osimertinib data set (see Figure 5.A.2 in the appendix). Visual inspection of Figure 5.A.3 in the appendix shows the model does not fit the T790M osimertinib data well at low drug concentrations.

In Tables (5.5.2)-(5.5.4), for all data sets, parameters are shown to be correlated which could suggest practical non-identifiability of the model. For all data sets, we find r and λ to be positively correlated with values ranging from 0.54 to 0.63 across the different data sets. This suggests that a higher estimated value of λ (higher rate of cell kill) can maintain a good data-model match when cells are estimated to divide at a quicker rate (higher value of r). We also observe a negative correlation between parameters λ and k with values ranging from -0.6 to -0.41 across the different data sets. This makes sense as, in order to maintain a good data-model match, if a higher value

of λ is estimated (coupled with a higher estimated value of r), then a lower rate of clearance of the dead cells can maintain a good match of the model to the total cell population data and our model could predict a cell population made up of dead cells at long times. These results could suggest that the set of parameters (r, λ, k) is a locally identifiable set of parameters whereby different sets of values of these parameters can give rise to the same total tumour cell population dynamics. Similarly, we observe strong negative correlations between parameters, r and $IC50$ with values ranging from -0.7 to -0.65 across the different data sets. This demonstrates that, when cells are estimated to divide at a higher rate, a lower value of $IC50$ can be estimated. This lower value of $IC50$ ensures sensitivity of the cells to the drug and hence, more cell kill takes place at lower drug concentrations. We also observe a negative correlation between k and $IC50$ for all data sets with values ranging from -0.34 to -0.21 across the different data sets. Again, if more cell kill is estimated to take place (lower values of $IC50$ are estimated) then the tumour population will be made up of predominantly dead cells and in order to maintain a good data-model match, a slower rate of dead cell clearance may be estimated.

Interestingly, the correlation between $IC50$ and m varies greatly across the different data sets. We observe effectively zero correlation (0.02) for PC9 cells in osimertinib and a strong positive correlation (0.62) for PC9 cells in gefitinib. This is interesting as these two data sets present similar cellular dynamics. The difference in these correlations may be as a result of remaining model parameters being correlated with m and $IC50$.

Next, we assess our certainty in these model predictions and parameter estimates by performing practical identifiability analysis on the model, using the profile likelihood method.

5.5.4.2 Practical identifiability

In section 5.5.3, we showed that the model is structurally identifiable for pooled dose-response curves. This result is encouraging as any uncertainty in estimated model parameter values can be attributed to the data alone.

In this section, we use the profile likelihood method to assess the practical identifiability of the treatment model parameters when the model is fitted to pooled dose-response curves for a given cell line and drug (see Table 5.4.2 for a summary of the estimated

model parameters and bounds on the parameter values employed). Profile likelihoods for drug-sensitive data sets are plotted in Figure 5.5.5.

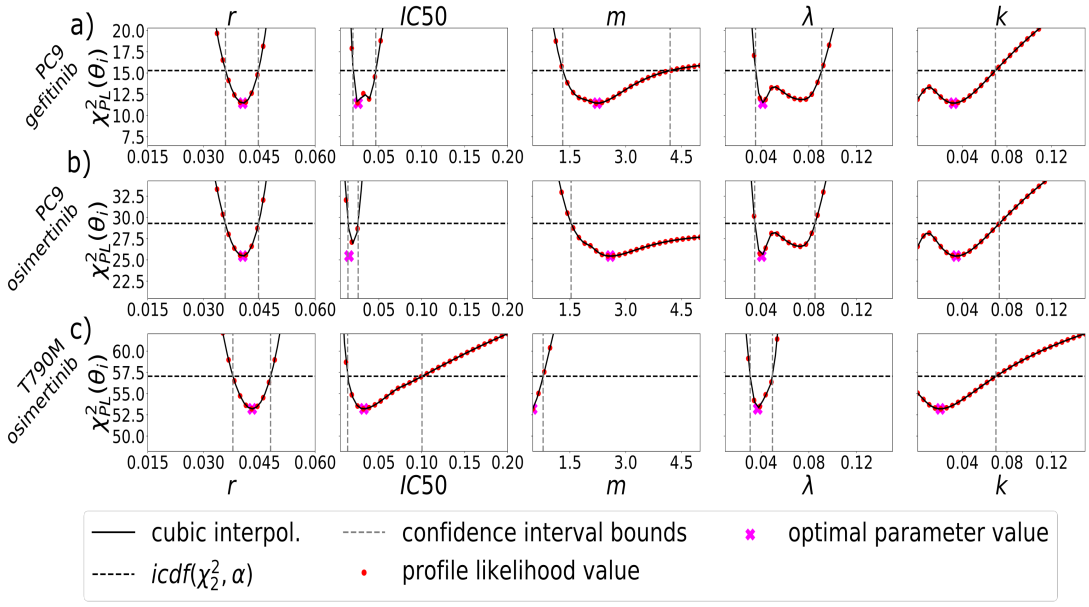


Figure 5.5.5: One-dimensional parameter profile likelihoods for treatment model parameters for each drug-sensitive data set. Key: magenta crosses; optimal parameter values (see Table 5.5.1), black dashed horizontal lines; 95% confidence level calculated as $icdf(\chi^2_1, \alpha)$ and plotted at the value $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$, grey vertical dashed lines; lower and upper likelihood-based confidence interval bounds, red dots; calculated profile likelihood values, black block line; cubic interpolation of the profile likelihood values. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

In Figure 5.5.5, profile likelihoods reveal lack of practical identifiability of the model parameters for the drug-sensitive data sets. Hence, the model is practically non-identifiable given the data which consists of total cell numbers only.

For the PC9 data sets, parameters r, $IC50$ and λ are practically identifiable; their profiles yield finite upper and lower confidence interval bounds. For all data sets, the parameter k is practically non-identifiable; its profiles yield finite upper likelihood-based confidence bounds only. For the T790M osimertinib and the PC9 osimertinib datasets, the parameter m is shown to be practically non-identifiable; profiles yield a finite lower bound only for PC9 cells in osimertinib and a finite upper bound only for T790M cells in osimertinib. For PC9 cells in geftinib, a finite upper confidence bound for m is yielded yet, we observe a large confidence interval ranging from approximately 1.3 to 4.2, suggesting large uncertainty in the estimated value of m for this data set. The

profile likelihoods for PC9 gefitinib and osimertinib data sets are similar. This is to be expected as the response of PC9 cells to both drugs is similar (see Figure 5.2.1).

For T790M cells, the parameter $IC50$ is practically identifiable yet, yields a large confidence interval with bounds at approximately 0.014 and 0.1. We attribute this to larger variability across experimental repeats in the T790M osimertinib data set than in the PC9 data sets (see Figure 5.2.1). For T790M cells, k is practically non-identifiable. This may be because dead cells degrade on a slower timescale than those on which cell growth (r^{-1}) and drug-induced cell kill (λ^{-1}) take place so that degradation of dead cells is negligible ($k \sim 0$) on the time span of the experiment (up to 200 hours). We investigate the impact of varying k on estimated model parameters in sections 5.5.6 and 5.5.7.1.

For PC9 cells in osimertinib, the profile for m yielded a minima with values that remained below the threshold for increasing values of m. Conversely, for T790M cells, the upper bound of the confidence interval for the m profile is low ($m_{\text{T790M}} \approx 0.77$) compared to the range of profiled values of m for the PC9 cells in osimertinib that lie below the 95% confidence threshold. In the profile of m for the T790M osimertinib data set, we observe a quick increase in the profile likelihood values as m increases. We attribute this to the dynamics of the Hill equation (see Figure 5.3.1). The Hill equation dynamics are relatively insensitive to changes in m for $m > 2$. Yet, they are sensitive to changes in m at lower values, $m < 1$. Therefore, the dynamics of the cells, predicted by the model, may vary drastically for different values of m which take low values ($m < 1$). Yet, at high values of $m > 2$, the predicted behaviour of the cells may not differ markedly, which, in turn, would result in little change to the goodness-of-fit.

The profiles for λ for the PC9 gefitinib and osimertinib data sets have two minima below the 95% confidence threshold. The minima occur at $\lambda \approx 0.04$ hour^{-1} and $\lambda \approx 0.073$ hour^{-1} for both data sets. We attribute this to the multiple minima observed in the profile of k (the minima in the profile of k occur at $k \approx 10^{-5}$ hour^{-1} for both sets and at $k \approx 0.031$ hour^{-1} for PC9 cells in gefitinib and at $k \approx 0.036$ hour^{-1} for PC9 cells in osimertinib). We conclude that this is a result of k and λ being negatively correlated with correlations, -0.6 and -0.52 for the PC9 gefitinib and PC9 osimertinib data sets,

respectively (see Tables 5.5.2 and 5.5.3, respectively). Hence, higher estimated values of k may occur as lower values of λ are estimated when fitting the model to the pooled dose-response curves. The two minima in the profile for λ for the PC9 cells which occur close together could suggest λ is locally identifiable around each minima. In future work, it would be interesting to investigate the structural identifiability of the model for a single, fixed drug concentration and for multiple drug concentrations using different methods as opposed to DAISY and Castro and de Boer's scaling method which are employed in this chapter. In so doing, information on local identifiability of model parameters may be given. In what follows, we investigate the cellular dynamics at the two minima in the k profile in the following subsection and examine their corresponding model predictions.

5.5.5 Implications of predicting unobserved variable trajectories

Certainty in values of clinically relevant drug-effect parameters when fitting models to data from preclinical studies is important. In this section, we investigate the uncertainty in model predictions when using a practically non-identifiable model (i.e. contains practical non-identifiable parameters).

In Figure 5.5.6, for PC9 cells in gefitinib, we plot trajectories of the total, live and dead cells when fixing two profiled parameter values of k ($k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1}) that yield similar $\chi^2(\hat{\boldsymbol{\theta}})$ values ($\chi^2(\hat{\boldsymbol{\theta}}) = 11.93$ for $k = 10^{-5}$ hour^{-1} and $\chi^2(\hat{\boldsymbol{\theta}}) = 11.46$ for $k = 0.031$ hour^{-1}). To generate these plots, for each value of k, we fix k and estimate all remaining model parameters by fitting the model to the pooled dose-response curves. We summarise the estimated values of the remaining parameters, found by performing these fits, in Table 5.5.5. In Figure 5.5.6, for simplicity, we only plot solutions of cell numbers over time for the drug concentration, $c = 0.123\mu$M. Correlations between estimated parameters are given in Tables (5.5.6)-(5.5.7).

In Table 5.5.5, we observe noticeable differences in the estimated values of $IC50$ and λ that lie within their finite 95% confidence regions (see profile likelihoods in Figure 5.5.5), as k varies. This is unsurprising as we have shown in Table 5.5.2 in the previous subsection that k and λ are negatively correlated which could suggest that these parameters are a locally identifiable pair of parameters when the total tumour population is the observed output of the in vitro experiments.

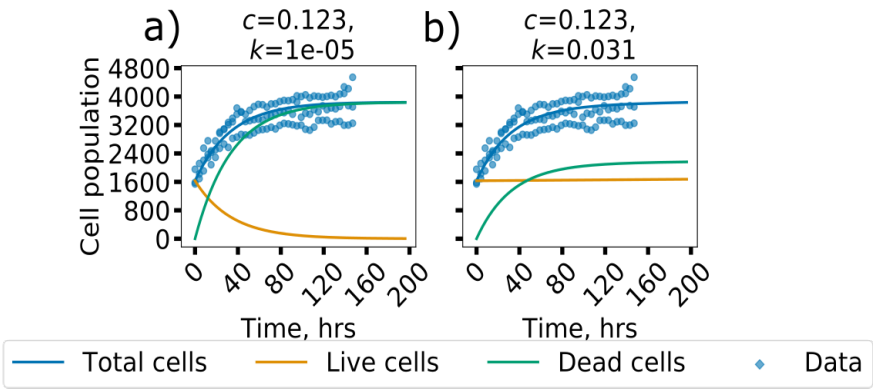


Figure 5.5.6: Model predictions of total ($N(t)$), live ($y(t)$) and dead ($z(t)$) cell populations describing PC9 cells in $c = 0.123\mu$M gefitinib for two profiled values of k, $k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1} which are fixed. For each value of k, the model is fitted to the pooled dose-response curves and the other model parameters are estimated (estimates of these parameters are given in Table 5.5.5). For fixed $k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1}, similar values of $\chi^2(\hat{\boldsymbol{\theta}})$ were calculated ($\chi^2(\hat{\boldsymbol{\theta}}) = 11.93$ for $k = 10^{-5}$ hour^{-1} and $\chi^2(\hat{\boldsymbol{\theta}}) = 11.46$ for $k = 0.031$ hour^{-1}). Key: blue dots; data, blue lines; best-fit solution of the total tumour population, orange; live cells and green; dead cells. Parameters: see Table 5.5.5.

Parameter	$k = 10^{-5}$	$k = 0.031$
r	0.040±0.0003	0.040±0.0002
$IC50$	0.038± 0.0005	0.026 ± 0.0003
m	1.92 ± 0.0501	2.24±0.0493
λ	0.077±0.0008	0.042± 0.0003
$\chi^2(\hat{\boldsymbol{\theta}})$	11.93	11.46

Table 5.5.5: Estimated values (± standard error) of model parameter when profiled values of k ($k = 10^{-5}$ hour^{-1} and $k = 0.031$ hour^{-1}) are fixed. For each fixed value of k, the model is fitted to the pooled dose-response curves and the other model parameters are estimated. These parameters are employed in the model and to generate model solutions in Figure 5.5.6. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

$k = 10^{-5}$		r	$IC50$	m	λ
	r	1.00	-0.27	-0.53	0.75
	$IC50$	-0.27	1.00	-0.10	0.22
	m	-0.53	-0.10	1.00	-0.59
	λ	0.75	0.22	-0.59	1.00

Table 5.5.6: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set for fixed $k = 10^{-5}$ hour^{-1}. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

$k = 0.31$		r	$IC50$	m	λ
	r	1.00	-0.66	-0.48	0.96
	$IC50$	-0.66	1.00	0.64	-0.60
	m	-0.48	0.64	1.00	-0.52
	λ	0.96	-0.60	-0.52	1.00

Table 5.5.7: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set for fixed $k = 0.31$ hour^{-1}. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

By fixing $k = 10^{-5}$ hour^{-1}, we assume negligible degradation of dead cells is observed over the course of the experiment. Unsurprisingly, a higher value of λ is estimated, in order to maintain a good data-model match whereby, live cells must be removed quicker as the dead cells will remain in the tumour population throughout the course of the experiment. Conversely, by assuming a higher fixed value of $k = 0.031$ hour^{-1}, dead cells degrade at a quicker rate and therefore, in order to maintain a good data-model match, a lower value of λ is estimated to ensure more live cells make up the total tumour population throughout the course of the experiment.

The parameter estimates, given in Table 5.5.5, contribute to similar total tumour population dynamics yet, markedly different live and dead cell trajectories, seen in Figure 5.5.6; for $k = 0.031$ hour^{-1} (Figure 5.5.6(b)), the dead cells degrade more rapidly and the relative proportions of live and dead cells at long times ($t \geq 80$ hours) is approximately 50%. By contrast, when $k = 10^{-5}$ hour^{-1} (Figure 5.5.6(a)), the degradation of dead cells in the experiment is negligible and the ratio of live to dead cells is effectively zero at long times ($t \geq 80$ hours) where we observe a rapid decrease in the number of live cells and an increase in the number of dead cells. As time evolves, the dead cells make up the total tumour cell population. The similar dynamics of the total tumour population

observed in Figure 5.5.6 and the fact that parameters k and λ are correlated support the fact k and λ could be a locally identifiable pair.

In Tables (5.5.6)-(5.5.7), we observe large differences in correlations associated with the parameter $IC50$ when k is fixed at low and high values. Interestingly, in Table (5.5.7), for fixed $k = 0.031$ hour^{-1}, parameters $IC50$ and m are strongly positively correlated (0.64) whereas, for fixed $k = 10^{-5}$ hour^{-1}, in Table (5.5.6), we observe a weak negative correlation (-0.1). These results demonstrate that practically identifiable parameter groupings that may be apparent from parameter correlations could differ depending on the values of the remaining model parameters.

5.5.6 Relationships between estimated model parameters using profile likelihoods

In what follows, we investigate the relationship between the profiled values of a practically non-identifiable model parameter and the estimated values of the remaining practically identifiable model parameters, along its profile. We also investigate the relationship between the profiled values of a parameter that yields a large finite likelihood-based confidence interval (for example, $IC50$ for the T790M osimertinib dataset) and the estimated values of the remaining practically identifiable model parameters, along its profile. We do so by using the calculated profile likelihoods (see Figure 5.5.5 in section 5.5.4.2) where, for each fixed value of a parameter along its profile, we record the estimated values of the remaining model parameters.

In Figure 5.5.7, we plot the values of the estimated practically identifiable model parameters that correspond to the profiled values of m and k for the PC9 osimertinib data set. We also plot the estimated values of model parameters along the profile of $IC50$ for the T790M osimertinib data set.

Referring to Figure 5.5.7(b), for the PC9 osimertinib data set, we observe that, for increasing values of k along its profile, there is a noticeable decrease in the estimated values of λ and $IC50$. These results demonstrate that estimates of the clinically relevant parameter, $IC50$, can vary greatly when the model suffers practical identifiability issues. Similarly, in Figure 5.5.7(c) for the T790M osimertinib data set, values of parameters r

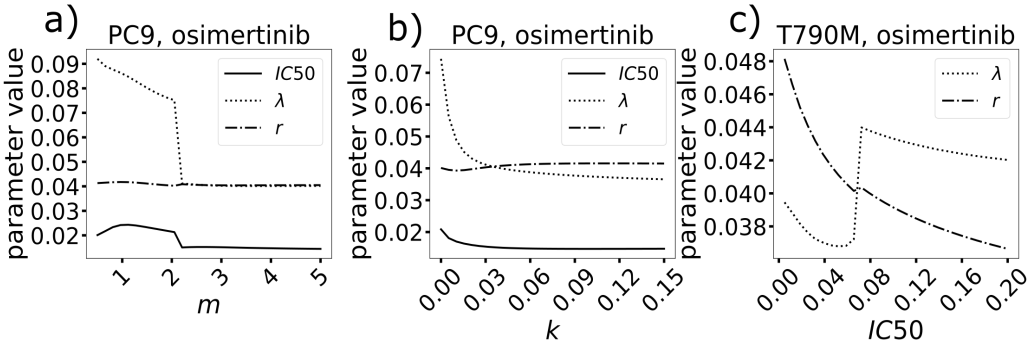


Figure 5.5.7: Estimated values of practically identifiable model parameters along profiles of practically non-identifiable model parameters, calculated using profile likelihoods (see Figure 5.5.5 in section 5.5.4.2). The x-axis represents values of the profiled practically non-identifiable model parameter. For each profiled parameter value, we record the estimated values of the practically identifiable model parameters from the profile likelihoods which we plot on the y-axis. (a) Estimated values of practically identifiable parameters, r, $IC50$ and λ along the m profile for PC9 cells in osimertinib, (b) Values of r, $IC50$ and λ along the k profile for PC9 cells in osimertinib, (c) values of r and λ along the $IC50$ profile for T790M cells in osimertinib. Key: block lines; estimated values of $IC50$, dotted lines; λ; dashed-dotted lines; r.

and λ change markedly as the profiled value of $IC50$, which yields large finite likelihood-based confidence intervals, increases. We note that profiled values of $0.014 < IC50 < 0.1$ for this data set yield likelihood values within the 95% confidence region. However, for changes in values of $IC50$ within this range ($0.014 < IC50 < 0.1$), we observe noticeable changes in the estimated values of r and λ. Our results presented here have demonstrated the importance of carrying out identifiability analysis when estimating model parameters.

5.5.7 Model reduction

There are two ways in which we might increase the practical identifiability of our model and increase the precision in the estimated parameters (by decreasing the width of likelihood-based confidence intervals); (i) model reduction (i.e. grouping sets of practically non-identifiable parameters and/or by fixing a non-identifiable parameter) [36, 22] and/or, (ii) improved experimental design (i.e. collecting additional data that will enable non-identifiable model parameters to be estimated) [75]. Obtaining additional data has not been feasible within the duration of this DPhil. Therefore, we proceed with (i) and look to fix model parameters, m and k to achieve practical identifiability of the

model when it is fitted to the PC9 data sets. We also aim to increase the practical identifiability of the model when it is fitted to the T790M osimertinib data set. For the PC9 gefitinib data set, the parameter m yields finite, albeit large, likelihood-based confidence intervals for the PC9 gefitinib dataset (see Figure 5.5.5). We seek to fix m for this data set as large uncertainty in the estimated value of m could lead to uncertainty in the other model parameters.

Our decision to fix parameters m and k is supported by the global parameter sensitivity analysis performed on the model (results of which are presented in Figure 5.A.4 in the appendix), where parameters m and k were identified as not being influential parameters (see the appendix for more details of eFAST performed on the model).

Henceforth, we fix $k = 10^{-5}$ hour^{-1} so that, the rate at which dead cells degrade is negligible during the course of the experiment compared to the rates of cell growth (r) and the maximum rate of drug-induced cell death (λ).

5.5.7.1 Estimating and fixing values of m

With fixed $k = 10^{-5}$ hour^{-1}, we perform the following steps to determine fixed values of m;

1. We fitted the model to each dose-response curve for each fixed value of the drug concentration, c, and, we estimate $\tilde{h}$, the constant rate of cell kill due to the drug (see equation 5.5.4) and y_0, the initial number of live cells, whilst holding the growth rate, r fixed at values estimated in the absence of drug (see Table 5.4.1).

2. We fitted the Hill equation defined by equation (5.3.3) to the estimated values of $\tilde{h}$ and estimate values of $IC50$ and λ for the case where; (a) m is also estimated and, (b) m is fixed using a value in $\{0.5, 1, 2, 3, 4\}$.

By fixing the growth rate, r, we ensure that the rate of cell kill due to the drug is representative of the drug-induced cell kill mechanism only and that the estimated values of $\tilde{h}$ are not influenced by changes in the estimated growth rate.

In Figure 5.5.8, we present estimated values of $\tilde{h}$ (blue diamonds) and the corresponding best-fit solutions of the Hill equation (lines) for PC9 cells exposed to gefitinib. We

present best-fit solutions for case (a), where m is estimated (solutions are plotted using black dashed-dotted line) and others for (b) in which m is held fixed ($m \in \{0.5, 1, 2, 3, 4\}$; coloured block lines). The optimal parameter values for each fit are given in the Figure legend. Results for the PC9 and T790M osimertinib data sets are given in the appendix.

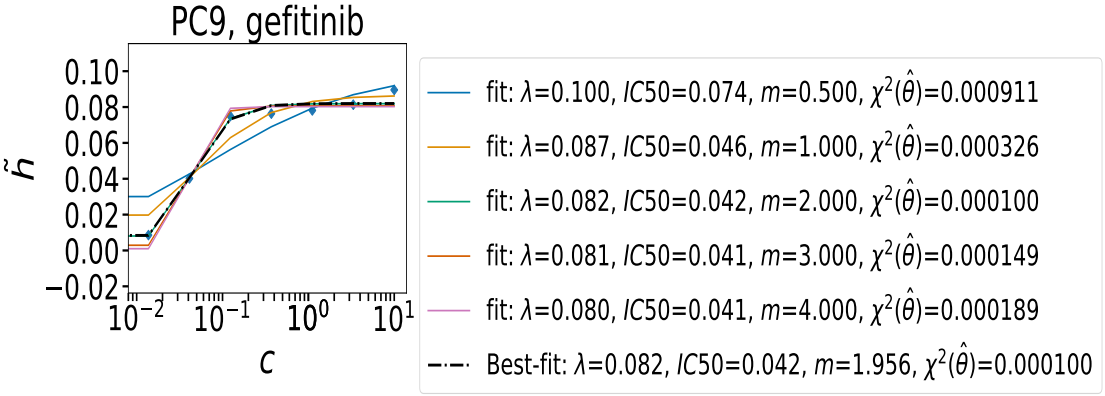


Figure 5.5.8: Best-fit solutions when the Hill equation (equation (5.3.3)) is fitted to estimated values of $\tilde{h}$ ($\tilde{h}$ is estimated by applying step 1 which we describe at the beginning of this section) for the PC9 gefitinib data set. Key: coloured block lines; m fixed with values in $\{0.5, 1, 2, 3, 4\}$, black dashed-dotted lines; m is estimated. We also provide, in the Figure legend, for a given fit, the estimated value of λ and $IC50$ along with the estimated or fixed value of m and the goodness-of-fit value, $\chi^2(\hat{\boldsymbol{\theta}})$.

In Figure 5.5.8, in the Figure legend, we observe that the estimated values of $IC50$ and λ are the same when m is fixed at $m = 2$ and when m is estimated (the estimated value m is $m = 1.956$). Goodness-of-fit values, $\chi^2(\hat{\boldsymbol{\theta}})$ do not differ markedly for $m \geq 2$. Therefore, we fix $m = 2$ for the remainder of this chapter when describing PC9 cells in gefitinib. We fix $m = 2$ for the PC9 osimertinib data set for the same reasons (see Figure 5.A.5 in the Appendix). For the T790M osimertinib data set, we find that the best-fit estimate of $m = 0.78$ yields values of $\chi^2(\hat{\boldsymbol{\theta}})$ similar to those when m is fixed at $m = 1$. For simplicity, henceforth, we fix $m = 1$ for the T790M osimertinib data sets. In Table 5.5.8 we summarise the estimated parameters of the Hill equation along with the fixed values of m, for drug-sensitive data sets.

In section 5.5.5, we showed how different values of the dead cell degradation rate, k, affect the dynamics of the unobserved tumour variables. We now investigate this further, focussing in particular on the effect fixed values of k have on the estimated values of the Hill equation parameters, m, $IC50$ and λ. To do this, we apply steps 1 and 2 from

Parameter	Data sets		
	PC9 gefitinib	PC9 osimertinib	T790M osimertinib
λ	0.082±0.002	0.080±0.006	0.041±0.001
$IC50$	0.042±0.004	0.023±0.008	0.048±0.010
m	2 (fixed)	2 (fixed)	1 (fixed)
$\chi^2(\hat{\boldsymbol{\theta}})$	1e-4	9.7e-4	4.1e-5

Table 5.5.8: Estimated parameter values (± standard error) of Hill equation parameters, λ, $IC50$ and m for drug-sensitive data sets. Estimates were calculated by performing steps 1 and 2 in subsection 5.5.7. The units of parameters are, $IC50$ (μM), m (dim.) and λ (hour^{-1}). Dim. stands for dimensionless.

subsection 5.5.7 where we fix k to a value $k \in \{0.0001,\ 0.001,\ 0.01,\ 0.1\}$. In Figure 5.5.9, we show how the estimated values of λ and $IC50$ change as k varies. We note that negligible changes in estimated values of m were recorded for these changes in k (results not presented).

In Figure 5.5.9, we also present the mean of the $\chi^2(\hat{\boldsymbol{\theta}})$ values, where each $\chi^2(\hat{\boldsymbol{\theta}})$ value is calculated by performing step 1 from subsection 5.5.7, for a given dose-response curve with a fixed value of k.

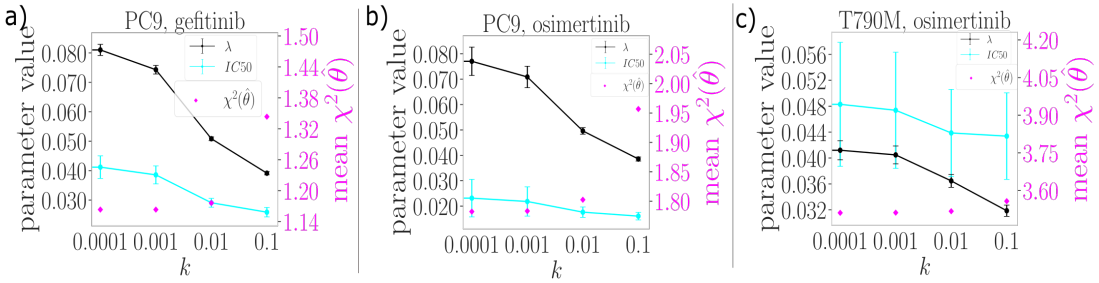


Figure 5.5.9: Right hand y-axis: estimated values of λ and $IC50$ for varying k, obtained using steps 1 and 2 from subsection 5.5.7 for fixed values of $k \in \{0.0001,\ 0.001,\ 0.01,\ 0.1\}$. Key: black dots (joined with black lines); estimated values of λ with bars representing standard errors of the estimates, cyan (joined with cyan lines); estimated values of $IC50$ with standard errors as bars. Left hand y-axis: the mean of the $\chi^2(\hat{\boldsymbol{\theta}})$ values, where each $\chi^2(\hat{\boldsymbol{\theta}})$ value is calculated by performing step 1 from section 5.5.7, for a given dose-response curve with a fixed value of k. Key: pink dots; mean $\chi^2(\hat{\boldsymbol{\theta}})$ value.

.

The results presented in Figure 5.5.9 show how the estimated values of λ and $IC50$ decrease as k increases. We conclude that, as k increases and dead cells degrade more quickly, a slower (maximum) rate of cell kill, λ, is needed to maintain a good fit to the data. Indeed, the results for the PC9 gefitinib (Figure 5.5.9(a)) data set show similar

values of the mean $\chi^2(\hat{\boldsymbol{\theta}})$ values for $k \leq 0.01$ hour^{-1} and yet, the estimated value of λ decreases by approximately 40% as k increases from 0.0001 hour^{-1} to 0.01 hour^{-1}. The same can be said for the PC9 osimertinib data set (Figure 5.5.9(b)). Smaller changes in values of λ are observed in the T790M osimertinib data set for increasing values of k (Figure 5.5.9(c)).

To summarise, in this section we have carried out model reduction on our treatment model by fixing the model parameters, m and k. Further, we have investigated the implications of doing so on the estimated values of the drug-effect parameters, $IC50$ and λ. Henceforth, we proceed with the reduced treatment model where m and k are fixed at values of $k = 10^{-5}$ hour^{-1} and values of m given in Table 5.5.8. Next, we assess the practical identifiability of the reduced model.

5.5.8 Practical identifiability analysis and parameter estimation of the reduced treatment model

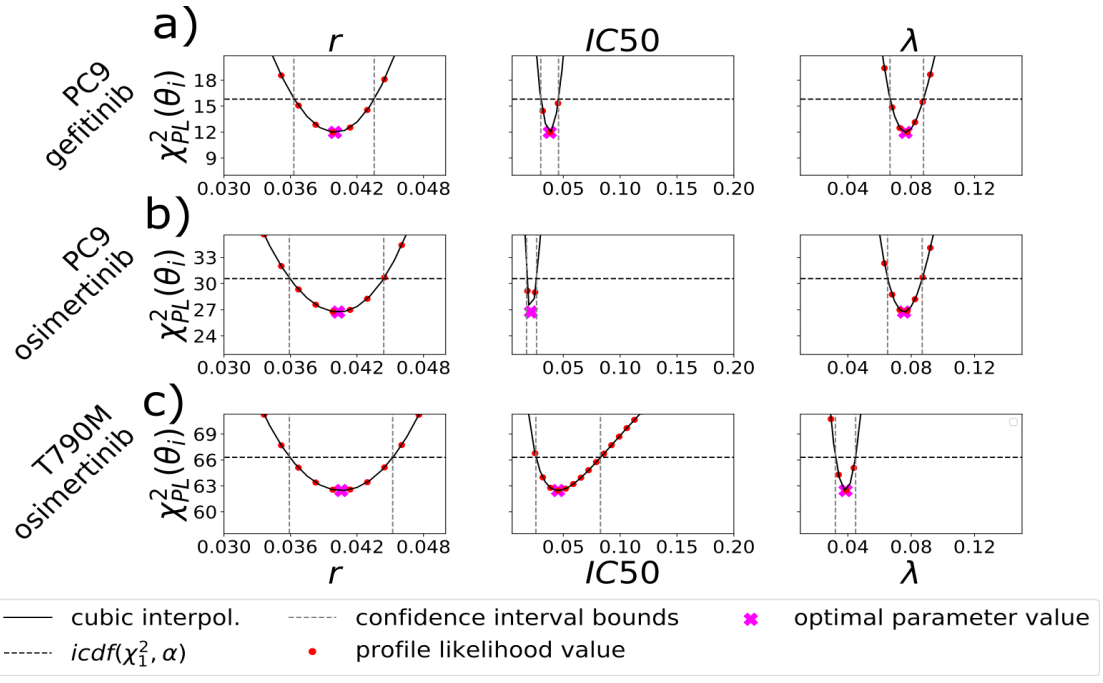


Figure 5.5.10: One-dimensional profile likelihoods for parameters r, $IC50$ and λ when m and k are fixed with $k = 10^{-5}$ hour^{-1} and values of m fixed at values given in Table 5.5.8. Key: horizontal black dashed lines; 95% confidence level plotted at $\chi^2(\hat{\boldsymbol{\theta}}) + icdf(\chi^2_1, \alpha)$, vertical grey dashed lines; likelihood-based confidence interval bounds, cyan crosses; optimal parameter values, red dots; profile likelihood values, block black lines; cubic interpolation of the profile likelihood values. The units of parameters are, r (hour^{-1}), $IC50$ (μM) and λ (hour^{-1}).

Figure 5.5.10 shows the profile likelihoods for parameters r, $IC50$ and λ when m and k are fixed. The profiles have unique minima and finite likelihood-based confidence intervals for all parameters and data sets. Hence, by fixing m and k we achieve a practically identifiable model in which all model parameters can be identified given the data and we obtain narrower finite likelihood-based confidence intervals for drug-effect parameters λ and $IC50$. In Figure 5.5.10(c), for T790M cells, a large confidence interval of $IC50$ is calculated when the interval is narrower compared to Figure 5.5.5 where all the model parameters are estimated. This demonstrates that, by fixing m and k, the precision in the estimated value of $IC50$ for this data set has increased. The likelihood-based confidence interval for the T790M cells is larger than the PC9 datasets and we attribute this to large variability in the data set relative to the PC9 data sets.

We present fits of the model to the PC9 gefitinib data set in Figure 5.5.11 (fits of the model to PC9 osimertinib and T790M osimertinib data sets are given in Figures 5.A.6 and 5.A.7 in the appendix, respectively). We note that the model is fitted to the total number of tumour cells only. In Figure 5.5.11, we plot the best-fit model solutions of the number of live and dead cells along with the 95% confidence regions (see Methods section for details). We observe good fits to the total cell population data for all cell lines.

Parameter	**PC9 gefitinib**	**PC9 osimertinib**	**T790M osimertinib**
r	0.040±0.0002	0.040±0.0004	0.041±0.0005
$IC50$	0.038±0.0004	0.021±0.0004	0.045±0.0029
m	2.0 (fixed)	2.0 (fixed)	1.0 (fixed)
λ	0.076±0.0006	0.075±0.0009	0.039±0.0007
k	10^{-5} (fixed)	10^{-5} (fixed)	10^{-5} (fixed)
$\chi^2(\hat{\boldsymbol{\theta}})$	11.97	26.73	62.47

Table 5.5.9: Estimated values of model parameter (± standard error) and $\chi^2(\hat{\boldsymbol{\theta}})$ values found by fitting the reduced model, with fixed m and k, to the pooled dose-response curves for each drug-sensitive data set. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

In Table 5.5.9, we present estimates of the parameters when the treatment model is fitted to the PC9 gefitinib and osimertinib data sets and the T790M osimertinib data set. In Table 5.5.9, the parameter estimates are close in value for the PC9 data sets,

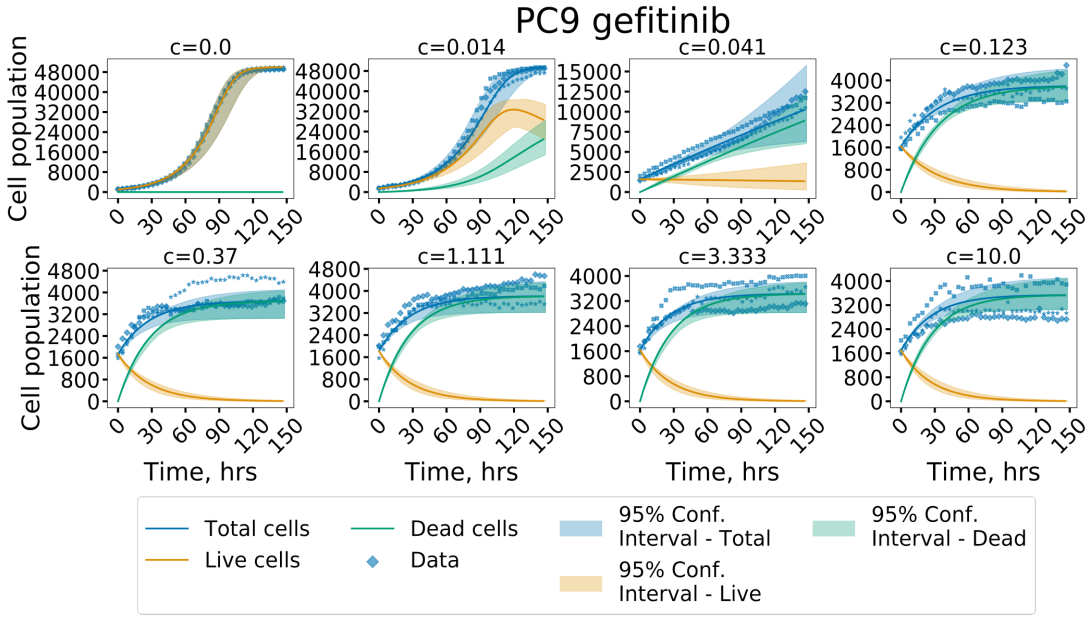


Figure 5.5.11: Best-fit model solutions when the model is fitted to pooled dose-response curves of PC9 cells in gefitinib. Key: blue lines; total tumour cell population, orange lines; live cells, green lines; dead cells. Shaded regions indicate the 95% confidence bound of the model solution for a given cell line, calculated using the Monte Carlo Simulations Method with 5000 simulations (see Methods section of this chapter). Parameters: see Table 5.5.9.

PC9 gefitinib		r	$IC50$	λ
	r	1.00	-0.38	0.65
	$IC50$	-0.38	1.00	0.20
	λ	0.65	0.20	1.00

Table 5.5.10: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 gefitinib data set for fixed $k = 10^{-5}$ hour^{-1} and $m = 2$. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

PC9 osimertinib		r	$IC50$	λ
	r	1.00	-0.54	0.75
	$IC50$	-0.54	1.00	-0.10
	λ	0.75	-0.10	1.00

Table 5.5.11: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the PC9 osimertinib data set for fixed $k = 10^{-5}$ hour^{-1} and $m = 2$. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

although the PC9 cells in osimertinib have lower estimated values of $IC50$ than those for gefitinib, indicating higher sensitivity of these cells to osimertinib than gefitinib.

T790M osimertinib		r	$IC50$	λ
	r	1.00	-0.64	0.93
	$IC50$	-0.64	1.00	-0.45
	λ	0.93	-0.45	1.00

Table 5.5.12: Correlations between estimated parameters of the treatment model fitted to pooled dose-response curves of the T790M osimertinib data set for fixed $k = 10^{-5}$ hour^{-1} and $m = 1$. The units of parameters are, r (hour^{-1}), $IC50$ (μM), m (dim.), λ (hour^{-1}) and k (hour^{-1}). Dim. stands for dimensionless.

We compare estimates of parameters for the PC9 osimertinib data set and the T790M osimertinib data set (Table 5.5.9). We observe lower estimated values of the maximum rate of drug-induced cell kill for T790M cells ($\lambda = 0.039$ hour^{-1}) than that of the PC9 cells exposed to osimertinib ($\lambda = 0.075$ hour^{-1}). However, a larger value of $IC50$ ($IC50 = 0.045\mu$M) is estimated for T790M cells compared with the PC9 cells ($IC50 = 0.021\mu$M). These results suggest that T790M cells are less sensitive to osimertinib than PC9 cells whereby, a lower maximum effect of the drug on these cells is achieved and more drug is required to reach a 50% drug-effect than the PC9 cells. Further, lower values of m correspond to the T790M cells which again, suggests a weaker effect of osimertinib on these cells. We also note that, for T790M cells in osimertinib, we observe markedly different estimates of the value of $IC50$ when k is fixed ($IC50 = 0.045\mu$M in Table 5.5.9) as opposed to when k is also estimated ($IC50 = 0.033\mu$M in Table 5.5.1). This is a concern as inaccurate estimates of the $IC50$ parameter can lead to misleading outcomes when predicting the tumour cell trajectories in different drug concentrations. To conclude, further understanding of the rate at which dead cell degradation takes place would be beneficial.

In Tables (5.5.10)-(5.5.12) we present correlations between estimated model parameters. For all data sets, $IC50$ is more negatively correlated with r and, λ is more positively correlated with r compared to when the full model is fitted to the data and no model parameters are fixed (see Tables (5.5.2)-(5.5.4)). We suspect that this stronger correlation is as a result of fewer parameters being estimated. When more parameters are estimated, their values will depend on the values of all other parameters. However, when fewer parameters are estimated, more constraint on their estimated values is incurred.

5.5.9 Experimental design

In section 5.5.4.2, we found our full treatment model (i.e. all parameters are estimated) to be practically non-identifiable (see Figure 5.5.5). We propose that experiments which measure both live and dead cell numbers may resolve practical non-identifiability issues of our model. In this section, we test this hypothesis by generating synthetic data describing how live and dead cell numbers change over time across a range of drug concentrations. Such experiments can be conducted using live/dead cell staining.

To generate synthetic data, we employ an approach presented by Eisenberg et al. [36]; we generate the data by running model simulations using the best-fit parameter estimates for PC9 cells in gefitinib (see Table 5.5.1) and the same fixed drug doses as the in vitro experiments. We employ the same values of y_0 for the PC9 gefitinib data set as those presented in Table (5.A.1). We record live, dead and total cell numbers at the same times as the in vitro data (every 4 hours, for a period of 156 hours). The resulting synthetic data are presented in Figure 5.5.12.

We perform global fits of the full treatment model (i.e. all model parameters are estimated) to the pooled synthetic dose-response curves. We note that the full treatment model is structurally identifiable for a single dose-response curve, and for data sets consisting of multiple drug concentrations, when both live and dead cell numbers are observed (see section 5.5.3). We calculate profile likelihoods for model parameters and present the resulting profiles in Figure 5.5.13. Since we are fitting the model to data points which, in some cases, are identically zero (dead cell numbers are equal to zero in the control case, with $c = 0$, and at the start of every experiment), we do not weight our residuals with respect to the data (see Methods section in chapter 4 for details on parameter estimation).

Figure 5.5.13 shows that all model parameters are practically identifiable, yielding profiles that cross the threshold (horizontal dashed lines) twice. In this way, we confirm that given separate data streams on live and dead cell numbers, all model parameters are practically identifiable. We note that, as we have not weighted our residuals when performing our fits to the synthetic data sets, the profile likelihood values are large and,

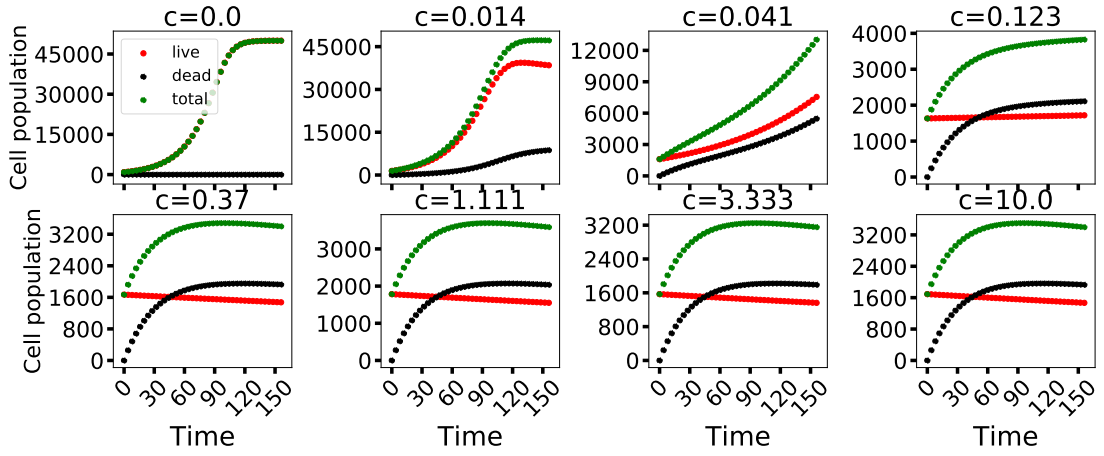


Figure 5.5.12: Synthetic data describing live ($y(t)$), dead ($z(t)$) and total ($N(t)$) cell numbers over time for a range of drug concentrations. Synthetic data are generated from the full treatment model using the best-fit parameter estimates of PC9 cells in gefitinib (see Table 5.5.1). Key: green; total cell numbers, red; live, black; dead.

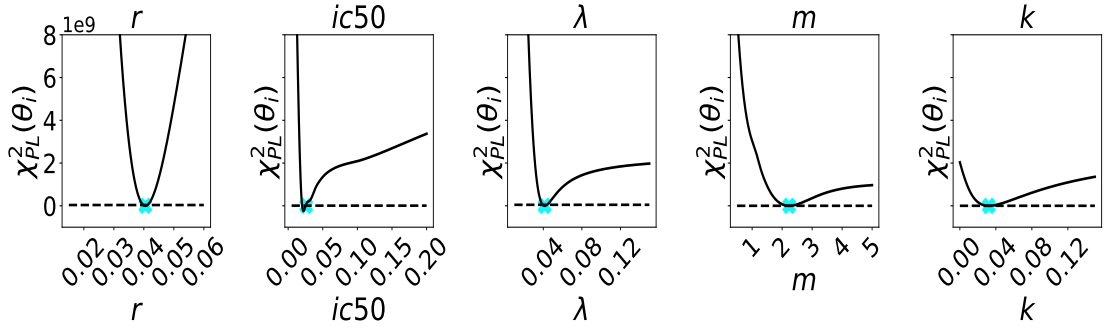


Figure 5.5.13: Profile likelihoods for the treatment model parameters when fitted to the pooled synthetic data sets consisting of live and dead cell numbers. Key: cyan crosses; optimal parameter estimates, black horizontal dashed lines; 95% confidence threshold calculated as the $min(\chi^2_{PL}(\theta_i)) + icdf(\chi^2, \alpha)$, black block lines; cubic interpolation of the profile likelihood.

therefore, observing the intersection between the threshold and the profile in Figure 5.5.13 is not possible.

To summarise, we have suggested experiments that we hypothesise will increase our ability to infer the underlying mechanisms of TKIs on NSCLC cell lines.

5.6 Discussion

Mathematical models, when coupled with experimental data, are valuable tools for inferring the mechanisms by which drugs act on cancer cells. These models can also be used to suggest experiments that further our understanding of such mechanisms. These

experiments allow for the testing of hypotheses that are based on modelling observations. In this chapter, we implemented a framework that has allowed us to do just this. The aim of this chapter was to infer the effect of different TKIs on different NSCLC cell lines.

We extended a model that we developed in chapter 4 which described the growth of four NSCLC cell lines. The model developed in this chapter focusses on the dynamics of live and dead cell populations and describes the response of four NSCLC cell lines to TKIs. We used a Hill equation to model drug-induced cell kill. We applied parameter estimation and structural and practical identifiability approaches to our model to assess the identifiability of model parameters given the in vitro data. Further, we used this analysis to propose experiments that we hypothesised would increase our understanding of how TKIs affect NSCLC cell lines.

First, we investigated how the TKIs affect the NSCLC cell dynamics compared to the dynamics of the cells in the absence of drug. We did so by fitting the reduced Richards growth equation, with fixed β, from chapter 4 to dose-response curves that describe the response of drug-sensitive and drug-resistant cell lines to a single TKI. We found the model to fit the data well for the drug-resistant cells. For the drug-sensitive cells, good fits were observed at low drug concentrations where estimated values of the growth rate, r, were found to decrease as drug concentrations increased. At high concentrations however, the fits were less good, suggesting a different mechanism of drug action takes place. Based on these results, we speculated that the drug may have dual action on the cells. The AKT inhibitor, AZD5363, has been shown to exhibit both anti-proliferative and drug-induced cell kill effects when used to treat breast cancer xenografts [150].

We then performed identifiability analysis on the treatment model. Although we found our model to be structurally identifiable, practical identifiability issues arose when we fitted the model to pooled dose-response curves. In particular, our profile likelihood analysis revealed large uncertainty in the rate of dead cell degradation, k and the Hill coefficient, m. We showed that vastly different predictions of unobserved model variables can be made when employing different values of the practically non-identifiable model parameter, k, in our model. By doing so, we demonstrated the importance of

investigating the identifiability of model parameters, given the data, when the purpose of the model is to make predictions.

From our practical identifiability results, we also identified causes for practical non-identifiability of the Hill coefficient, m. We attributed this to the dynamics of the Hill equation whereby, cell dynamics are more sensitive to changes in m values when m takes low values, as opposed changes in values of m when m is large. This was demonstrated in our practical identifiability analysis in which we found, for the PC9 data sets, the profile of m to remain below the threshold for increasing values of m. Conversely, for the T790M osimertinib data set, lower estimated values of m contributed to a better goodness-of-fit and the profile of m increased markedly for small changes in values of m in the range $m < 1$.

Since additional data generation was not feasible, we proceeded with model reduction and fixed non-identifiable model parameters, m and k, which, we found, using sensitivity analysis, to not be influential on the total tumour cell dynamics in fixed drug-concentration. Through this model reduction, we achieved practical identifiability of the model. We highlighted however, that careful consideration of the estimated values of the dead cell degradation rate should take place when considering dose-response experiments that present total cell numbers only, as we demonstrated that changes in this parameter greatly affect the estimated values of clinically relevant parameters such as the half maximal drug concentration, $IC50$.

An alternative to fixing parameters is to reparameterise the model where non-dimensionalisation of the model could take place. However, similar to chapter 4, we seek to estimate unique drug-effect parameters so as to characterise the effect of the different drugs on the different cell lines and rescaling the model does not simplify the problem. In this chapter, in response to our model being practically non-identifiable, we proposed experiments that record live and dead cell numbers separately and we hypothesised that such experiments would increase the practical identifiability of our model. We simulated the outputs of these experiments using synthetic data consisting of live and dead cells. We showed such experiments improved the certainty in estimates of our model parameter and practical identifiability was achieved.

In this chapter, we did not consider pharmacokinetic models. In order to introduce pharmacokinetics into the model, we would require information about whether the drug concentration varies over the course of the in vitro experiments. By including pharmacokinetics into the model without this information, we would introduce unknown parameters that could increase the non-identifiability of the model. In future work and with the knowledge of how drug concentrations vary over the course of the in vitro experiments, it would be interesting to consider pharmacokinetics.

To summarise, we have used theoretical approaches to infer the underlying mechanisms of action of TKIs in NSCLC cell lines. Further, we have suggested experiments to further elucidate our understanding of such mechanisms.

Appendix

5.A Supplementary results

Fits of the reduced Richards growth equation to dose-response curves

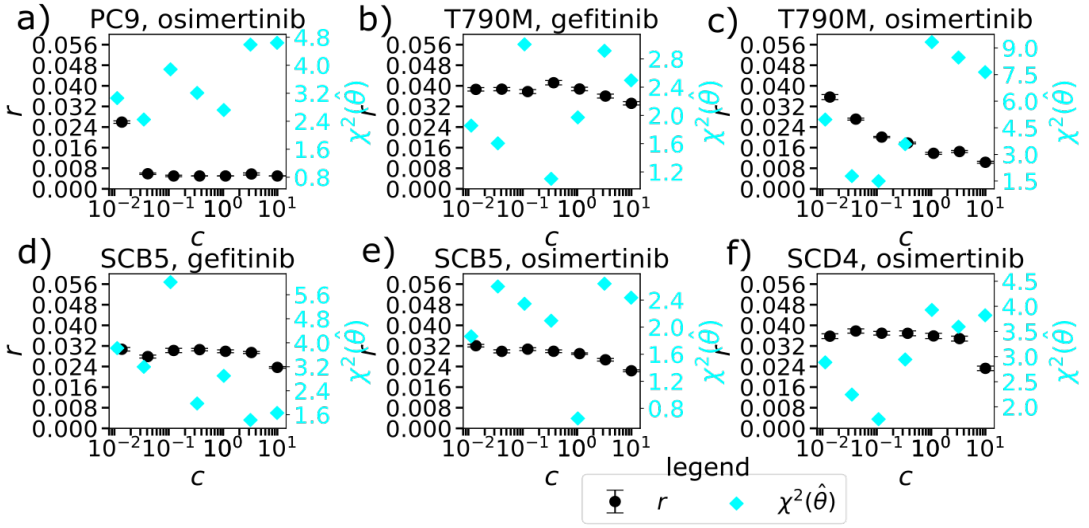


Figure 5.A.1: Estimated values of r (left hand y-axis) and corresponding minimum residual sum of squared errors, $\chi^2(\hat{\boldsymbol{\theta}})$ (right hand y-axis) plotted, with respect to c for the dose-response curves in the PC9 osimertinib data set (a), T790M data sets (b)-(c), the SCB5 data sets (d)-(e) and the SCD4 osimertinib data set (f). Values of r and $\chi^2(\hat{\boldsymbol{\theta}})$ are calculated by fitting the reduced Richards growth equation, with fixed β, to individual dose-response curves. Key: black dots; estimated values of r and bars represent the standard error of each estimate, cyan diamonds; $\chi^2(\hat{\boldsymbol{\theta}})$ values.

Estimated initial live cell numbers

Conc. μM	**PC9**		**T790M**
	gefitinib	osimertinib	osimertinib
0	(924.6±18.3)	(925.0±19.2)	(955.5±38.2)
0.0137	(1446.7±26.8)	(1514.7±55.2)	(1488.8±75.7)
0.0412	(1602.4±81.1)	(1560.8±69.6)	(1967.0±52.2)
0.124	(1631.0±77.1)	(1346.9±82.3)	(1933.3±63.1)
0.370	(1669.0±79.3)	(1505.1±65.1)	(2080.2±38.2)
1.111	(1783.3±67.7)	(1488.5±67.0)	(1442.9±114.3)
3.333	(1568.9±85.4)	(1168.6±110.4)	(1685.5±133.5)
10	(1690.6±111.0)	(1384.5±78.7)	(1895.4±199.6)

Table 5.A.1: Estimated values (± standard error) of initial cell numbers, y_0, for the in vitro PC9 data sets and the T790M osimertinib data set. Values of y_0 are estimated by fitting the treatment model to single dose-response curves with a fixed drug concentration. Values of y_0 are estimated and fixed at their estimated values prior to fitting the model to pooled dose-response curves.

Parameter Estimation

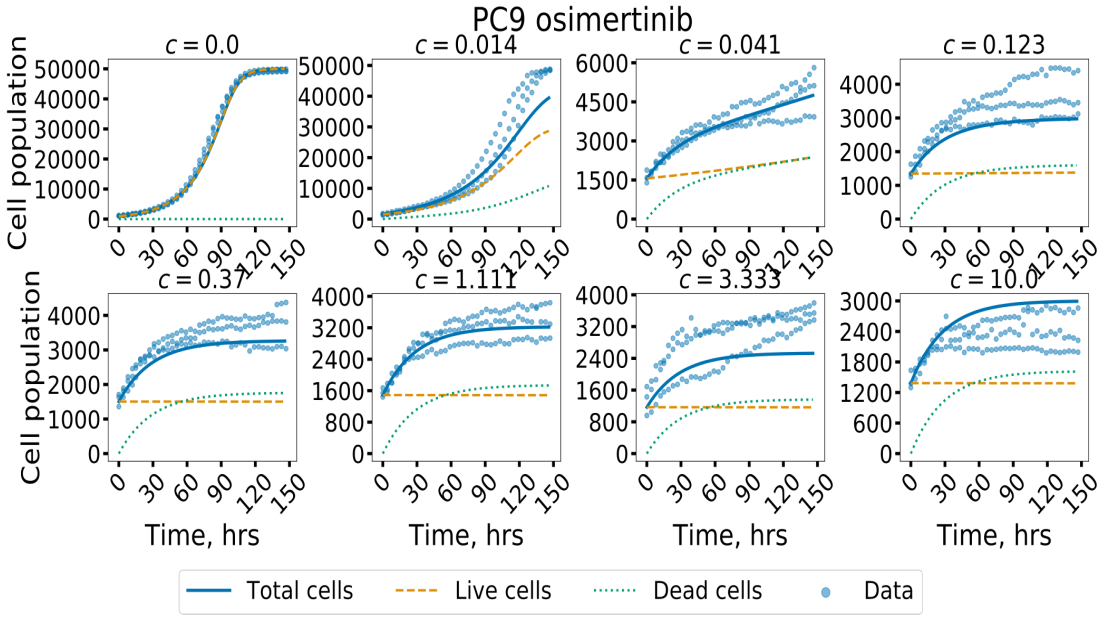


Figure 5.A.2: Model fits for the PC9 osimertinib data set. The model is fit to pooled dose-response curves for all drug concentrations. Key: blue dots; data, blue block lines; best-fit solution of model describing the total population over time, $N(t)$, orange dashed lines; best-fit solution describing live cell numbers, $y(t)$; green dotted lines; best-fit solution describing dead cell population, $z(t)$. The drug concentration, c are indicated in each subplot title.

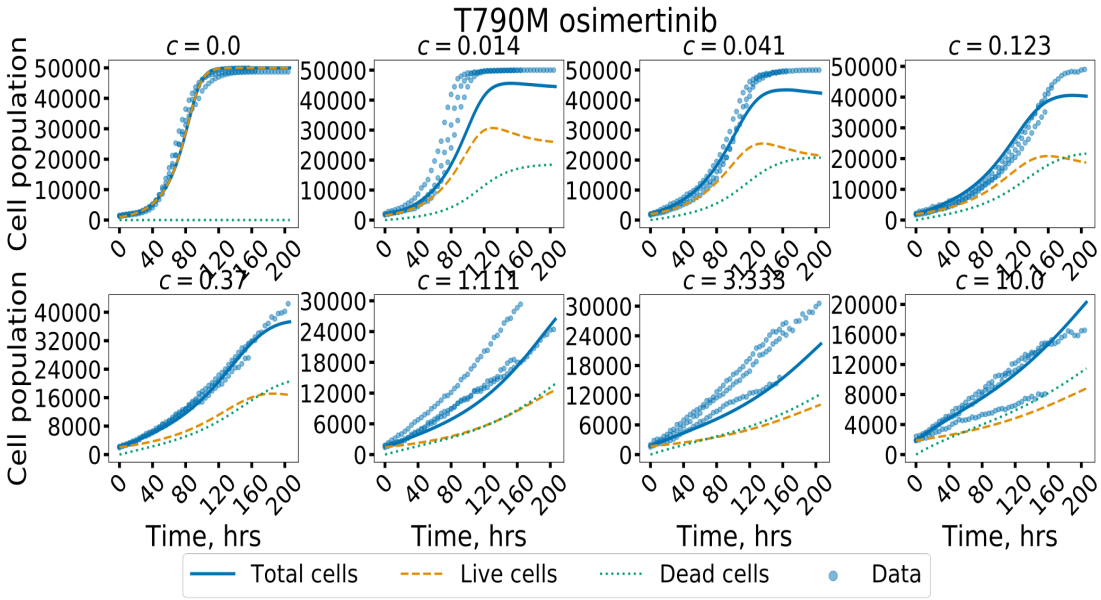


Figure 5.A.3: Model fits for the T790M osimertinib data set. The model is fit to pooled dose-response curves for all drug concentrations. Key: see Figure 5.A.2 for plotting details.

Sensitivity analysis

We employ the global sensitivity analysis technique, eFAST to our treatment model to calculate first and total-order sensitivity indices S_i and S_{Ti}, respectively, of the model parameters at fixed times ($t = 10$, $t = 60$ and $t = 120$ hours) and for fixed drug concentrations ($c \in \{0.01, 0.1, 1, 3\}$). Ranges of parameter values employed in eFAST are given in Table 5.A.2. Bar plots showing first and total-order sensitivity indices for all model parameters ranges are presented in Figure 5.A.4.

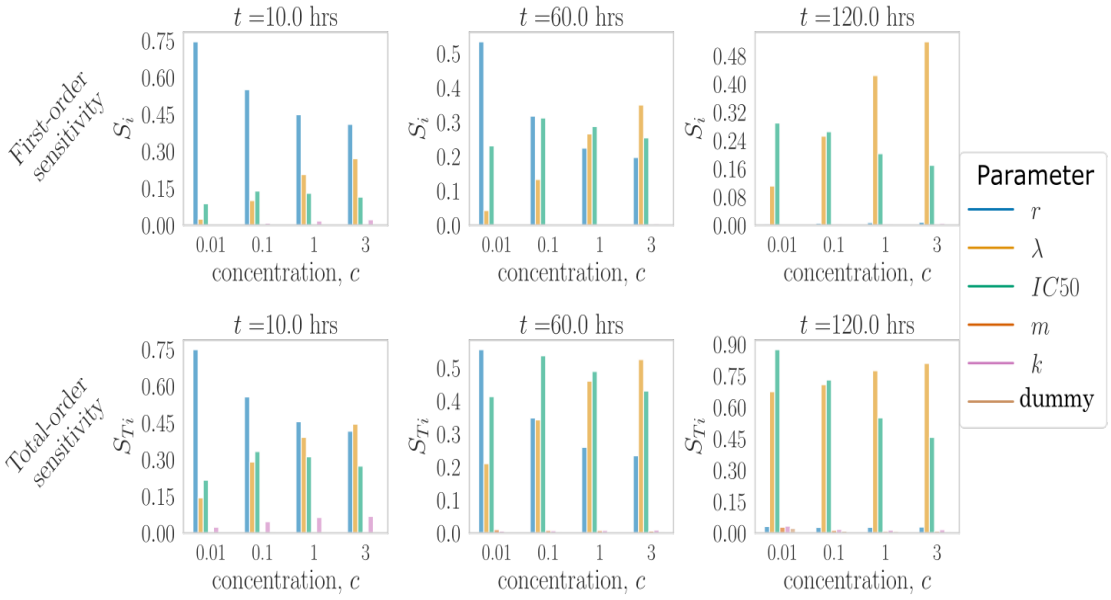


Figure 5.A.4: Bar plots of first and total-order sensitivity indices of model parameters recorded at fixed times ($t = 10$, $t = 60$ and $t = 120$ hours) for fixed drug concentrations $c \in \{0.01, 0.1, 1, 3\}$. Different colours represent parameters (see Figure legend). eFAST control parameters: number of samples per resample; 4000, number of resamples; 8. Convergence testing was carried out to determine these control parameter values although not presented (see chapter 2 for details of eFAST).

Parameter	**Description**	**Units**	**Parameter range**
r	Growth rate	hour^{-1}	[0.035, 0.048]
k	Dead cell degradation rate	hour^{-1}	[0.0001, 0.1]
λ	Maximum rate of cell kill due to drug	hour^{-1}	[0.001, 0.1]
$IC50$	Half-maximal drug concentration	μM	[0.002, 200]
m	Hill coefficient	dimensionless	[1, 6]
dummy	dummy parameter to test for convergence of eFAST	dimensionless	[0,10]
K	Carrying capacity	cells	fixed at 50000 cells
β_F	Competition strength parameter	dimensionless	fixed at 3.315
y_0	Initial live cell number	cells	fixed at 2000 cells

Table 5.A.2: Bounds for treatment model parameters defined in equations (5.3.1)-(5.3.3) when performing global sensitivity analysis of the treatment model using eFAST.

First and total-order sensitivity indices in Figure 5.A.4 show λ and $IC50$ to be the two most influential parameters at time, $t = 120$ hours. For this time, we observe a

switch from $IC50$ to λ as the most influential parameter for increasing values of c. The same can be said for $t = 60$ hours for concentrations $c > 0.01\mu$M. At early times, ($t = 10$ hours), the growth rate, r is the most influential parameter for $c \leq 1\mu$M, followed by $IC50$ and λ. Parameters m and k are shown to have little influence on the total tumour population. In the following section, we look to fix values of m and we examine the impact of varying fixed values of the dead cell degradation rate, k on estimated model parameter values.

Fits of the Hill equation to estimated rates of drug-induced cell kill

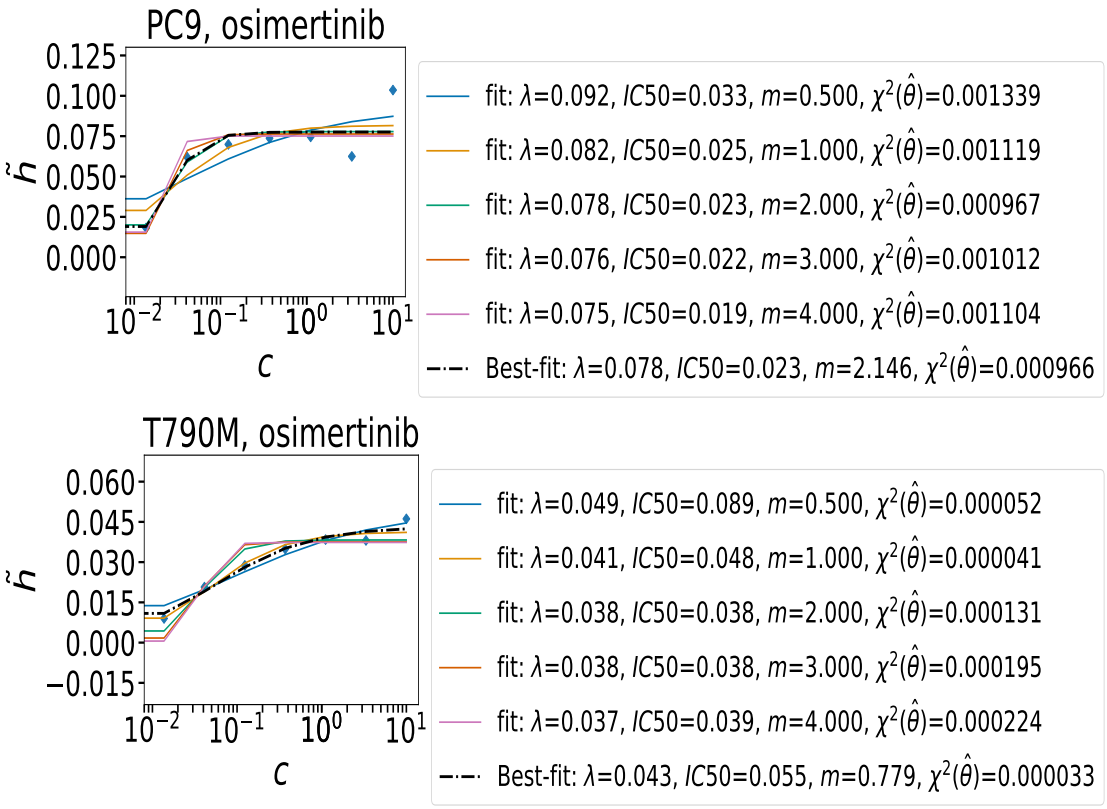


Figure 5.A.5: Best-fit solutions when the Hill equation (equation (5.3.3)) is fitted to estimated values of $\tilde{h}$ ($\tilde{h}$ is estimated by applying step 1 in subsection 5.5.7) for the PC9 and T790M osimertinib data sets. Key: coloured block lines; m fixed with values in $\{0.5, 1, 2, 3, 4\}$, black dashed-dotted lines; m is estimated. We also provide, in the Figure legend, for a given fit, the estimated value of λ and $IC50$ along with the estimated or fixed value of m and the goodness-of-fit value, $\chi^2(\hat{\boldsymbol{\theta}})$.

Fits of treatment model to data with fixed values of m and k

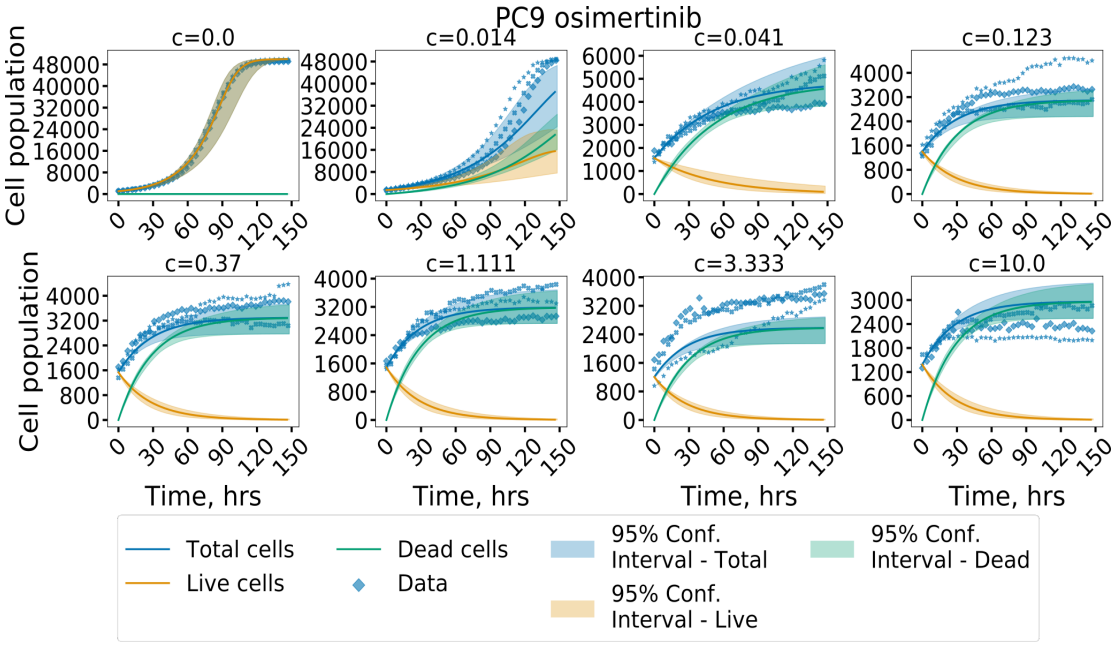


Figure 5.A.6: Best-fit model solutions when the model is fitted to pooled dose-response curves of PC9 cells in osimertinib. Key: blue lines; total tumour cell population, orange lines; live cells, green lines; dead cells. Shaded regions indicate the 95% confidence bound of the model solution for a given cell line, calculated using the Monte Carlo Simulations Method with 5000 simulations (see Methods section of this chapter). Parameters: see Table 5.5.9.

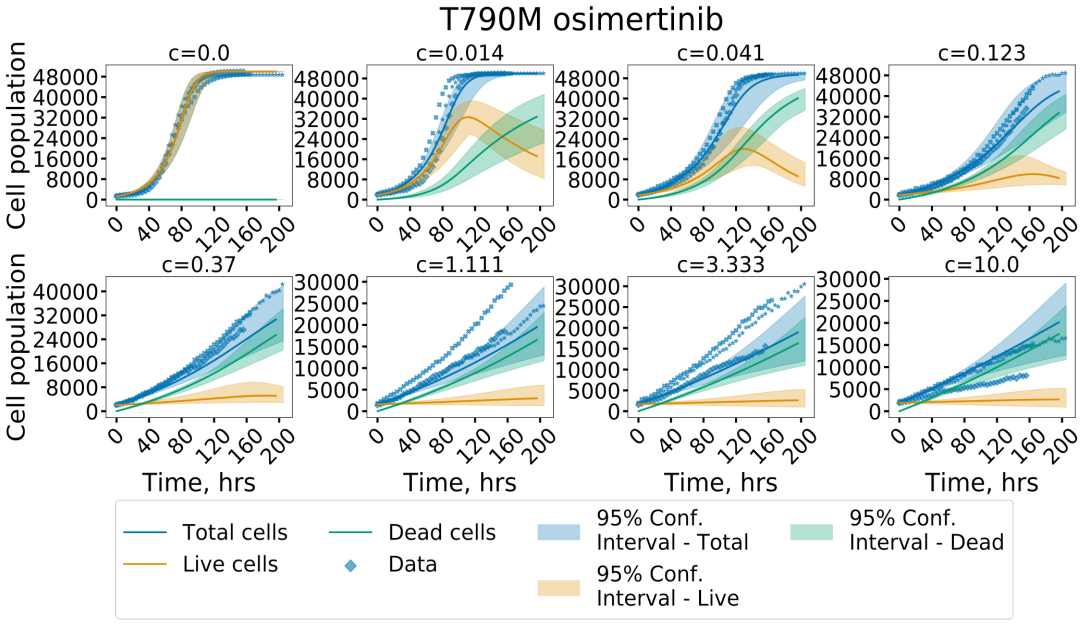


Figure 5.A.7: Best-fit model solutions when the model is fit to pooled dose-response curves of T790M cells in osimertinib. Key: plotting details the same as in Figure 5.A.6.

Chapter 6

Conclusions and Future Work

In this thesis we developed a series of ordinary differential equations (ODEs) and stochastic models to attempt to describe tumour growth, the emergence of drug resistance and, response to tyrosine kinase inhibitor (TKI) treatment in non-small cell lung cancer (NSCLC). We have drawn conclusions regarding the biological mechanisms that drive the evolution to resistant tumours and response to treatment. Further, we have demonstrated the impact of stochasticity on these conclusions. We have combined ODE models with preclinical data, which allowed us to draw conclusions on the growth properties of different NSCLC cell lines. We have also investigated the underlying mechanisms of the action for TKIs on NSCLC cells and have suggested strategies for improving experimental design. In this chapter, we summarise our findings, place our work in the context of current research and present future directions.

6.1 Research summary and contextualisation

In chapter 2 we developed a spatially averaged model of tumour growth consisting of four NSCLC cell populations which differ in their resistance to two TKIs. We incorporated factors in the model known to influence tumour cell heterogeneity and drug resistance including, mutations and competition for resources. A recent in vitro study demonstrated the effects of cellular interactions between drug-sensitive and drug resistant NSCLC cell lines to alectinib, a TKI, on their respective growth rates [72]. We generalised these findings in our model by way of population-dependent cell birth rates. We demonstrated markedly different trajectories of cell populations within the tumour

when different types of cellular interactions were assumed. We then analysed the sensitivity of resistance in a tumour to model parameters and found four key factors influence resistance levels: (i) the probability of mutation per cell division, (ii) the relative growth rate of resistant cells to sensitive cells, (iii) the effect sensitive cells have on the growth rates of resistant cells and, (iv) the cell turnover rate. We employed a range of values of these four factors into our model and confirmed the results of the sensitivity analysis by predicting proportions of pre-existing resistance at diagnosis that span multiple orders of magnitude. Our results are relevant as pre-existing resistant proportions of TKI resistant NSCLC cells have been identified in newly diagnosed patients [24, 55, 81, 145]. Using droplet digital polymerase chain reaction (PCR), Wantanabe et al. found approximately 80% of EGFR-mutant NSCLC patients (373 patients in total) to harbour the T790M mutation with allele frequencies that range between 0.001% and 0.1% at diagnosis [145]. Our work has highlighted possible factors that contribute to the variation in pre-existing resistance levels at diagnosis.

Our work sheds light on tumour-specific factors that may contribute to heterogeneity across patients' tumours at diagnosis. Our results could be useful for guiding experiments that characterise these factors when taking biopsies from tumours. Our model lends itself to some useful simplifications that can be applied in an in vitro setting; the model can be used to guide preclinical studies of co-culture experiments by suggesting hypotheses, based on mathematical observations, of tumour cell trajectories and treatment response when cell-cell interactions are considered.

In the latter parts of chapter 2, when then extended our model to account for treatment. We considered two distinct protocols, each applying two TKIs. Using our model, we investigated the effects of key factors, determined using our sensitivity analysis, on a tumour's response to treatment. We showed that tumours with similar responses to treatment, observed in the clinic (i.e. progression times), can present with noticeably different underlying tumour compositions, which we hypothesised will influence the efficacy of second-line treatment.

Targeted therapies, such as TKIs, have been at the forefront of personalised medicine in NSCLC. Through genetic profiling of patient tumours, the characterisation of alter-

ations in lung cancer cells, such as mutations in the epidermal growth factor receptor (EGFR) gene, guide decision making on the appropriate choice of TKI treatment. However, the need for more biomarkers to predict the benefit of TKI treatment for a given patient is evident. This work highlights that, not only is there a benefit to genetic profiling of patient tumours at progression, which, in turn could be used for guiding second-line treatment strategies, but also a benefit to, the characterisation of tumour cell properties, such as cell growth rates and cell turnover rates [149, 119]. Furthermore, combination therapies are an increasingly popular approach in the pharmaceutical industry to tackle drug resistance. They are highly relevant with respect to TKIs as genetic profiling of NSCLC tumours unveils multiple populations of cells with different resistance conferring mutations. Models such as ours, with appropriate simplification and parametrisation, could be used in combination with experimental studies to test treatment strategies involving multiple TKI drugs.

In chapter 3 we introduced stochasticity into our model. In contrast to chapter 2, the details of population-dependent birth rates, involved in tumour growth trajectories, were not considered for parsimony. In the field of mathematical oncology, stochastic modelling of tumour growth and treatment response presents difficulties, for example, when modelling tumours that contain clinically relevant numbers of cells. Thus, in chapter 3, we implemented a τ-leap approximation method with adaptive step size calculation to accommodate the relatively large cell numbers [17]. A novel aspect of this work is that such techniques allowed us to model tumours with realistic tumour sizes and physiologically relevant parameters. Key questions were how changes in;

i the propensity for cells to mutate and,

ii the relative growth rate of resistant cells to sensitive cells,

influence the variation in pre-existing resistance proportions and tumour response to treatment.

We showed there was a large deviation from the ODE model predictions and, considerable variation across stochastic repeats for tumours with lower propensities to mutate,

especially when mutations give rise to resistant cells with higher growth rates than the sensitive cell population.

We then considered first and second-line treatment and examined the impact of the probability of mutation on variation in treatment response. Again, we demonstrated noticeable deviation from the ODE model and large variation in tumour response to combination TKI treatment for low probabilities of mutation. This work contributes to the qualitative understanding of the emergence of drug resistance in tumours and, factors influencing treatment response.

The use of stochastic models to describe the response of NSCLC tumours to TKI treatment has shown promise in recent years. Chmielecki et al. derived a stochastic model from in vitro studies of NSCLC cell lines and, used the model to make predictions on dosing schedules applying EGFR TKIs. In this study, smaller tumour cell numbers ($\approx 10^6$ cells) were considered than predicted numbers in patient tumours at diagnosis [24, 34]. Results showed promise in mouse models. However, when applied in patients, optimal dosing strategies developed in the preclinical studies did not delay drug resistance [24, 151]. The authors postulated that, lower peak plasma concentrations were achieved in these patients than in preclinical studies [151]. Based on our findings in chapter 3, we speculate that stochasticity is an additional source of treatment failure in this study, given the considerable predicted variation in treatment response for large tumours.

Our work highlights that, even for large tumour sizes (10^8 cells at diagnosis), the effects of stochasticity on tumour cell dynamics and treatment response should be considered. From this work, we propose, when modelling the emergence of drug resistance and treatment response using ODEs, stochastic models should be also considered. However stochastic modelling frameworks are the computationally expensive to implement and issues often occur when parametrising and validating these models.

Our collaboration with modellers at AstraZeneca played a key role in guiding the development of our ODE models presented in chapters 4 and 5. These models describe the growth of NSCLC cell lines and their response to treatment with a single TKI. Through this collaboration, we were able to access preclinical data sets. The key goals of chapters

4 and 5 were to develop mathematical models with appropriate complexity, given the available data and, to use these models to infer the underlying biological mechanisms behind the following: (i) the growth properties of different NSCLC cells and, (ii) the effect of different TKI drugs on different NSCLC cells. We did so by using mathematical modelling approaches, including parameter estimation and practical identifiability methodologies and, we used the well-established, Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics for model selection [2]. Parameter estimation is widely used in the pharmaceutical industry to estimate clinically relevant drug-effect parameters. However, drawbacks include, noisy experimental data used to establish models and, discerning the appropriate level of model complexity to ensure reliable identification of model parameters. Hence, there is increasing acknowledgement of the need for parameter identifiability analysis to infer the biological mechanisms and its representation within models. This important step in model selection and development increases the predictive power of models used for the design of treatment schedules, dose regimens and combination therapies.

In chapter 4, we presented a framework for model selection from a set of feasible tumour growth models that describe the in vitro growth of NSCLC cell lines. The framework combined information criteria/goodness-of-fit metrics and parameter identifiability analysis. We demonstrated that models with the highest levels of complexity (i.e. largest number of model parameters), which were, the generalised-logistic model and the Richards growth equation, performed best on the basis of the visual inspection of the model fit to the data, had the lowest residual sum of squared errors and had the lowest AIC and BIC statistics. The latter was of particular interest as lower dimensional models may outcompete more detailed mechanistic models on the basis of the AIC and BIC statistics, though this did not occur in the studies here.

We then proceeded with identifiability analysis and found the two most complex models suffered practical identifiability issues, with the most complex yielding worse results. Our work has highlighted that information criteria and goodness-of-fit metrics alone are not sufficient to identify the most appropriate models for inferring the underlying biological mechanisms of tumour cell growth properties. This conclusion has

also been reached by others recently [127]. We proceeded with the Richards growth equation, which yielded the fewest practically non-identifiable model parameters. In addition, by fitting this model to synthetic data sets, we found practical identifiability issues to arise even for low level noise. We attribute such difficulties to be in part due to the Richards growth equation dynamics. We achieved practical identifiability by fixing the non-identifiable model parameter, β, that describes the strength of competition between cells in the Richards growth equation. In future work, we could assess practical identifiability of the Richards growth equation when written in the augmented form presented in the appendix of chapter 4 whereby, through adding an additional variable in the model, the Richards growth equation is linear in its parameters. If we obtain practical identifiability, there will be no need for fixing β, which could avoid the prospect of bias in our estimates of the remaining model parameters. Although, we have found in chapter 4 that estimated values of the growth rate r that contribute to a low residual sum of squared errors do not differ markedly for increasing fixed values of β and therefore, this bias may be small.

We ended our model selection procedure by quantifying the uncertainty in the reduced Richards growth equation when fitted to the in vitro data, using Markov chain Monte Carlo (MCMC) Bayesian inference. Via assessment of diagnostic tests, performed to ensure convergence of our MCMC, we demonstrated the model is practically identifiable. Our results corroborated the results on practical identifiability, found using profile likelihoods. We then implemented a final parameter estimation of the reduced Richards growth equation model parameters using the in vitro data sets and we performed final model checks by carrying out a visual inspection of the residuals over time. Our residual plots showed that specific time points were consistently over/underestimated whereby the number of tumour cells predicted by the model was higher/lower than the data.

Monoculture 2D experiments are useful for studies of tumour cell biology involving cell growth, migration, and treatment response. In vitro studies provide a controlled environment in which cellular dynamics can be measured with relatively few factors influencing tumour cell trajectories. They can also provide data with relatively low noise from which mathematical models can be developed and model parameters inferred. Con-

versely, xenograft mouse models have factors such as the vasculature and the stroma, which influence the tumour growth dynamics and contribute to larger amounts of noise. A final study carried out in chapter 4 involved the assessment of the practical identifiability of the Richards growth equation when fit to coarse-grained, noisy, in vivo tumour volume data. We found there to be a larger number of non-identifiable model parameters in vivo than in vitro, which we attributed to larger variability across experimental repeats in vivo. This highlights issues when developing models from in vitro data to describe in vivo tumour cell dynamics, even when using models with simple structures.

The model presented in chapter 5 is an extension of our practically identifiable model, developed in chapter 4 and includes live and dead cell compartments. The model incorporates the effects of TKI treatment via a cell kill term, which we modelled using a Hill equation, a model that is commonly used in the pharmaceutical industry to describe dose-response curves [1, 137]. We showed that the model is structurally identifiable from a data set consisting of multiple dose-response curves for a range of fixed drug concentrations. However, we demonstrated the model was practically non-identifiable with the rate of dead cell degradation and the Hill coefficient yielding one-sided profiles, when using the profile likelihood method. With respect to the Hill coefficient, we attributed its non-identifiability to the structure of the Hill equation. We demonstrated that misleading biological predictions could be made using our model, given its practically non-identifiable model parameters. Further, a key finding of our work is that dead cell degradation rates in vitro substantially impact model predictions. Thus, we suggest one should carefully consider the dynamics of dead cell populations when inferring clinically relevant drug-effect parameters, in vitro.

We have proposed further experiments involving live/dead cell staining. We demonstrated, using synthetic data that, additional observables (i.e. live and dead cell numbers) result in the increased precision in estimated model parameters, yielding narrower finite confidence-interval bounds. This work may find use in guiding rational hypothesis generation for investigation in preclinical studies, that, in turn, may inform tumour evolution and resistance in a clinical context.

The work in chapter 2 is successful in that we reached our aims and objectives

that are summarised in the introduction chapter of this thesis. We have gained a deep understanding of which tumour-specific factors influence drug resistance and treatment response. Our analysis has paved the way for future investigations and has contributed to the knowledge of drug resistance and its impact on treatment failure. In chapter 3, we investigated the role of stochasticity on variation in drug resistance levels and treatment response and we reached our original aims and objectives by uncovering tumour-specific factors that, when coupled with stochasticity, could play a key role in variation in tumour composition and contribute to diverse treatment outcomes. Again, this work is novel and contributes to the knowledge of factors influencing drug resistance and treatment failure. The work in chapter 3 could potentially highlight factors contributing to patient-patient variability seen in the clinic. We worked with data in chapters 4 and 5. In chapter 4, we aimed to determine a model that describes the in vitro control data such that we can uniquely determine parameter estimates and characterise the different cell lines based on their growth rate properties. We were unable to determine a model that could do this. However, we carried out analysis of the most suitable model and proceeded with model reduction that enabled us to uniquely estimate model parameters and characterise the different cell lines. In chapter 5, we aimed to uniquely estimate drug-effect parameters for specific cell lines and TKIs. By doing so, we aimed to determine how different TKIs affect different NSCLC cell lines. Given the data available to us, we were unable to do this due to practically non-identifiability issues as a result of lack of sufficient data. However, in response to this, we designed experiments and, using synthetic data studies, demonstrated how data, generated from these experiments, could allow us to uniquely estimate the parameters.

6.2 Future directions

In this section, we summarise how models presented in this thesis can be extended and improved so as to increase their utility in drug-development and treatment scheduling studies.

The model in chapter 2 could be extended to account for a range of different tumour types with similar cellular structures. The usefulness of this model lies in the fact that

any type of cellular alteration (e.g. mutation), drug type and, combination treatment schedule can be considered. Further, back mutations could be included to represent increased sensitivity of cells when a drug is removed, capturing an acquired resistance element. In particular, this model presents a framework for dose regimen and, alternating therapy studies. There are ongoing clinical studies investigating combination TKI treatment [119]. An interesting extension of this model could be the inclusion of a drug-tolerant cell population in place of a partially resistant phenotype. Drug-tolerant cells were originally detected by Sharma et al. who showed them to have reversible resistance when in the absence of drug and more than a 100-fold reduced drug sensitivity in general [123]. Hata et al. later demonstrated the role of these drug-tolerant cells as an alternative pathway to resistance to TKI treatment, which gave rise to EGFR mutant cells with the T790M+ mutations [55].

In chapter 3, we investigate the differences in pre-existing resistance levels at diagnosis and treatment response predicted using ODE and stochastic models. Typically, when modelling tumours with clinically relevant cell numbers, ODEs are applied, as researchers generally believe that at this size, the impacts of stochasticity are negligible. However, we showed that this did not hold for tumours with varying subpopulation sizes, driven by the probability of mutation. We believe this could be a contributing factor to the failure of combination TKI treatment. Therefore, we propose that further studies should be carried out on the effects of stochasticity on tumour response, in particular, its impact on tumour composition variation during alternating treatment.

In chapters 4 and 5, we presented models of in vitro tumour growth and response to TKIs. Supposing we had the additional experimental data that we suggested in chapter 5 (consisting of live and dead cells), we could potentially infer drug-effect parameters of the TKIs acting on NSCLC cell lines. Then, we could potentially use these models to test optimal dose regimens and treatment schedules applying single TKIs. An extension would be to mix the cells and suggest experiments, to acquire additional data on the growth and treatment of cells in co-cultures. We could then carry out the same modelling framework, as that in chapter 5. In so doing, we would aim to achieve a practically identifiable model to test combination TKI treatment strategies.

References

[1] A. V. Hill. The Possible Effects of The Aggregation of The Molecules of Haemoglobin on its Dissociation Curves. *The Journal of Physiology Vol. 40* (1910), pp. iv–vii.

[2] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 6 (dec 1974), 716–723.

[3] Altrock, P. M., Liu, L. L., and Michor, F. The mathematics of cancer: Integrating quantitative models. *Nature Reviews Cancer 15*, 12 (nov 2015), 730–745.

[4] Álvarez-Arenas, A., Podolski-Renic, A., Belmonte-Beitia, J., Pesic, M., and Calvo, G. F. Interplay of Darwinian Selection, Lamarckian Induction and Microvesicle Transfer on Drug Resistance in Cancer. *Scientific Reports 9*, 1 (dec 2019), 9332.

[5] Bacevic, K., Noble, R., Soffar, A., Wael Ammar, O., Boszonyik, B., Prieto, S., Vincent, C., Hochberg, M. E., Krasinska, L., and Fisher, D. Spatial competition constrains resistance to targeted cancer therapy. *Nature Communications 8*, 1 (dec 2017), 1995.

[6] Beerenwinkel, N., Schwarz, R. F., Gerstung, M., and Markowetz, F. Cancer evolution: Mathematical models and computational inference. *Systematic Biology 64*, 1 (2015), e1–e25.

[7] Bellman, R., and Åström, K. J. On structural identifiability. *Mathematical Biosciences 7*, 3-4 (1970), 329–339.

[8] Bellu, G., Saccomani, M. P., Audoly, S., and D'Angiò, L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine 88*, 1 (2007), 52–61.

[9] Bethune, G., Bethune, D., Ridgway, N., and Xu, Z. Epidermal growth factor receptor (EGFR) in lung cancer: An overview and update. *Journal of Thoracic Disease 2*, 1 (2010), 48–51.

[10] Bloom, H. J., Richardson, W. W., and Harries, E. J. Natural history of untreated breast cancer (1805-1933). *British Medical Journal 2*, 5299 (jul 1962), 213–221.

[11] Bonadonna, G., Zambetti, M., and Valagussa, P. Sequential or Alternating Doxorubicin and CMF Regimens in Breast Cancer With More Than Three Positive Nodes: Ten-Year Results. *JAMA: The Journal of the American Medical Association 273*, 7 (feb 1995), 542–547.

[12] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians 68*, 6 (nov 2018), 394–424.

[13] Browning, A. P., Warne, D. J., Burrage, K., Baker, R. E., and Simpson, M. J. Identifiability analysis for stochastic differential equation models in systems biology: Identifiability analysis for stochastic differential equation models in systems biology. *Journal of the Royal Society Interface 17*, 173 (dec 2020), 20200652.

[14] Burrell, R. A., and Swanton, C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular Oncology 8*, 6 (oct 2014), 1095–1111.

[15] Byrne, H. M. Mathematical biomedicine and modeling avascular tumor growth. In *Mathematics and Life Sciences*, no. 12. 2012, p. 41.

[16] Cancer Research UK. Lung cancer survival. *URL: https://www.cancerresearchuk.org/about-cancer/lung-cancer/survival* (2022).

[17] Cao, Y., Gillespie, D. T., and Petzold, L. R. Efficient step size selection for the tau-leaping simulation method. *Journal of Chemical Physics 124*, 4 (jan 2006), 044109.

[18] Castro, M., and de Boer, R. J. Testing structural identifiability by a simple scaling method. *PLoS Computational Biology 16*, 11 (nov 2020), e1008248.

[19] Chang, G. C., Hsu, S. L., Tsai, J. R., Liang, F. P., Lin, S. Y., Sheu, G. T., and Chen, C. Y. Molecular mechanisms of ZD1839-induced G1-cell cycle arrest and apoptosis in human lung adenocarcinoma A549 cells. *Biochemical Pharmacology 68*, 7 (oct 2004), 1453–1464.

[20] Chang, G. C., Yu, C. T. R., Tsai, C. H., Tsai, J. R., Chen, J. C., Wu, C. C., Wu, W. J., and Hsu, S. L. An epidermal growth factor inhibitor, Gefitinib, induces apoptosis through a p53-dependent upregulation of pro-apoptotic molecules and downregulation of anti-apoptotic molecules in human lung adenocarcinoma A549 cells. *European Journal of Pharmacology 600*, 1-3 (dec 2008), 37–44.

[21] Chang, Y. S., Tu, S. J., Chen, Y. C., Liu, T. Y., Lee, Y. T., Yen, J. C., Fang, H. Y., and Chang, J. G. Mutation profile of non-small cell lung cancer revealed by next generation sequencing. *Respiratory Research 22*, 1 (dec 2021), 1–10.

[22] Chis, O. T., Banga, J. R., and Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE 6*, 11 (nov 2011), e27755.

[23] Chiş, O., Banga, J. R., and Balsa-Canto, E. GenSSI: A software toolbox for structural identifiability analysis of biological models. *Bioinformatics 27*, 18 (sep 2011), 2610–2611.

[24] Chmielecki, J., Foo, J., Oxnard, G. R., Hutchinson, K., Ohashi, K., Somwar, R., Wang, L., Amato, K. R., Arcila, M., Sos, M. L., Socci, N. D., Viale, A., De Stanchina, E., Ginsberg, M. S., Thomas, R. K., Kris, M. G., Inoue, A., Ladanyi, M., Miller, V. A., Michor, F., and Pao, W. Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Science Translational Medicine 3*, 90 (jul 2011), 59.

[25] Cho, H., Lewis, A. L., Storey, K. M., Jennings, R., Shtylla, B., Reynolds, A. M., and Byrne, H. M. A Framework for Performing Data-Driven Modeling of Tumor Growth with Radiotherapy Treatment. In *Association for Women in Mathematics Series*, vol. 22. Springer, Cham, 2021, pp. 179–216.

[26] Coldman, A. J., and Goldie, J. H. A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Mathematical Biosciences 65*, 2 (1983), 291–307.

[27] Coldman, A. J., and Goldie, J. H. A stochastic model for the origin and treatment of tumors containing drug-resistant cells. *Bulletin of mathematical biology 48*, 3-4 (1986), 279–292.

[28] Collis, J., Connor, A. J., Paczkowski, M., Kannan, P., Pitt-Francis, J., Byrne, H. M., and Hubbard, M. E. Bayesian Calibration, Validation and Uncertainty Quantification for Predictive Modelling of Tumour Growth: A Tutorial. *Bulletin of Mathematical Biology 79*, 4 (mar 2017), 939–974.

[29] Cross, D. A., Ashton, S. E., Ghiorghiu, S., Eberlein, C., Nebhan, C. A., Spitzler, P. J., Orme, J. P., Finlay, M. R. V., Ward, R. A., Mellor, M. J., Hughes, G., Rahi, A., Jacobs, V. N., Brewer, M. R., Ichihara, E., Sun, J., Jin, H., Ballard, P., Al-Kadhimi, K., Rowlinson, R., Klinowska, T., Richmond, G. H., Cantarini, M., Kim, D. W., Ranson, M. R., and Pao, W. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discovery 4*, 9 (sep 2014), 1046–1061.

[30] Crowley, E., Di Nicolantonio, F., Loupakis, F., and Bardelli, A. Liquid biopsy: Monitoring cancer-genetics in the blood. *Nature Reviews Clinical Oncology 10*, 8 (jul 2013), 472–484.

[31] Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., and Schaibly, J. H. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *Journal of Chemical Physics 59*, 8 (oct 1973), 3873–3878.

[32] Dagogo-Jack, I., and Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology 15*, 2 (2018), 81–94.

[33] Day, R. S. Treatment sequencing, asymmetry, and uncertainty: protocol strategies for combination chemotherapy. *Cancer Research 46*, 8 (1986), 3876–3885.

[34] Del Monte, U. Does the cell number 109 still really fit one gram of tumor tissue? *Cell Cycle 8*, 3 (feb 2009), 505–506.

[35] Dewanji, A., Luebeck, E. G., and Moolgavkar, S. H. A generalized Luria-Delbrück model.

[36] Eisenberg, M. C., and Jain, H. V. A confidence building exercise in data and identifiability: Modeling cancer chemotherapy as a case study. *Journal of Theoretical Biology 431* (oct 2017), 63–78.

[37] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer 136*, 5 (mar 2015), E359–E386.

[38] Foo, J., and Michor, F. Evolution of acquired resistance to anti-cancer therapy. *Journal of Theoretical Biology 355* (aug 2014), 10–20.

[39] Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. emcee : The MCMC Hammer. *Publications of the Astronomical Society of the Pacific 125*, 925 (feb 2013), 306–312.

[40] Gainor, J. F., Varghese, A. M., Ou, S. H. I., Kabraji, S., Awad, M. M., Katayama, R., Pawlak, A., Mino-Kenudson, M., Yeap, B. Y., Riely, G. J., Iafrate, A. J., Arcila, M. E., Ladanyi, M., Engelman, J. A., Dias-Santagata, D., and Shaw, A. T. ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: An analysis of 1,683 patients with non-small cell lung cancer. *Clinical Cancer Research 19*, 15 (aug 2013), 4273–4281.

[41] Gatenby, R. A., Silva, A. S., Gillies, R. J., and Frieden, B. R. Adaptive therapy. *Cancer research 69*, 11 (jun 2009), 4894–903.

[42] Gay, L., Baker, A. M., and Graham, T. A. Tumour Cell Heterogeneity. *F1000Research 5* (2016).

[43] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. Bayesian Data Analysis. In *Bayesian Data Analysis*. Chapman and Hall/CRC, nov 2013.

[44] Gerlee, P. The model muddle: In search of tumor growth laws. *Cancer Research 73*, 8 (apr 2013), 2407–2411.

[45] Gerlinger, M., and Swanton, C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *British Journal of Cancer 103*, 8 (2010), 1139–1143.

[46] Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry 81*, 25 (1977), 2340–2361.

[47] Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics 115*, 4 (jul 2001), 1716–1733.

[48] Goldie, J. H., and Coldman, A. J. *Drug Resistance in Cancer: Mechanisms and Models*. Drug Resistance in Cancer: Mechanisms and Models. Cambridge University Press, 1998.

[49] Goldie, J. H., Coldman, A. J., and Gudauskas, G. A. Rationale for the use of alternating non-cross-resistant chemotherapy. *Cancer Treatment Reports 66*, 3 (1982), 439–449.

[50] Goodman, J., and Weare, J. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science 5*, 1 (2010), 65–80.

[51] Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M., and Maire, P. The Hill equation: A review of its capabilities in pharmacological modelling. *Fundamental and Clinical Pharmacology 22*, 6 (2008), 633–648.

[52] Grossi, F., Aita, M., Follador, A., Defferrari, C., Brianti, A., Sinaccio, G., and Belvedere, O. Sequential, Alternating, and Maintenance/Consolidation Chemotherapy in Advanced Non-Small Cell Lung Cancer: A Review of the Literature. *The Oncologist 12*, 4 (apr 2007), 451–464.

[53] Haeno, H., Iwasa, Y., and Michor, F. The evolution of two mutations during clonal expansion. *Genetics 177*, 4 (dec 2007), 2209–2221.

[54] Hanahan, D., and Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell 144*, 5 (mar 2011), 646–674.

[55] Hata, A. N., Niederst, M. J., Archibald, H. L., Gomez-Caraballo, M., Siddiqui, F. M., Mulvey, H. E., Maruvka, Y. E., Ji, F., Bhang, H. E. C., Radhakrishna, V. K., Siravegna, G., Hu, H., Raoof, S., Lockerman, E., Kalsy, A., Lee, D., Keating, C. L., Ruddy, D. A., Damon, L. J., Crystal, A. S., Costa, C., Piotrowska, Z., Bardelli, A., Iafrate, A. J., Sadreyev, R. I., Stegmeier, F., Getz, G., Sequist, L. V., Faber, A. C., and Engelman, J. A. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nature Medicine 22*, 3 (2016), 262–269.

[56] Hayford, C. E., Tyson, D. R., Jack Robbins, C., Frick, P. L., Quaranta, V., and Harris, L. A. An in vitro model of tumor heterogeneity resolves genetic, epigenetic, and stochastic sources of cell state variability. *PLoS biology 19*, 6 (jun 2021).

[57] Herbst, R. S., Morgensztern, D., and Boshoff, C. The biology and management of non-small cell lung cancer. *Nature 553*, 7689 (jan 2018), 446–454.

[58] Hines, K. E., Middendorf, T. R., and Aldrich, R. W. Determination of parameter identifiability in nonlinear biophysical models: A bayesian approach. *Journal of General Physiology 143*, 3 (mar 2014), 401–416.

[59] Hirano, T., Yasuda, H., Tani, T., Hamamoto, J., Oashi, A., Ishioka, K., Arai, D., Nukaga, S., Miyawaki, M., Kawada, I., Naoki, K., Costa, D. B., Kobayashi, S. S., Betsuyaku, T., and Soejima, K. In vitro modeling to determine mutation specificity of EGFR tyrosine kinase inhibitors against clinically relevant EGFR mutants in non-small-cell lung cancer. *Oncotarget 6*, 36 (2015), 38789–38803.

[60] Hochhaus, A., Baccarani, M., Silver, R. T., Schiffer, C., Apperley, J. F., Cervantes, F., Clark, R. E., Cortes, J. E., Deininger, M. W., Guilhot, F., Hjorth-Hansen, H., Hughes, T. P., Janssen, J. J. W. M., Kantarjian, H. M., Kim, D. W., Larson, R. A., Lipton, J. H., Mahon, F. X., Mayer, J., Nicolini, F., Niederwieser, D., Pane, F., Radich, J. P., Rea, D., Richter, J., Rosti, G., Rousselot, P., Saglio, G., Saussele, S., Soverini, S., Steegmann, J. L., Turkina, A., Zaritskey, A., and Hehlmann, R. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia 34*, 4 (2020), 966–984.

[61] Holla, V. R., Elamin, Y. Y., Bailey, A. M., Johnson, A. M., Litzenburger, B. C., Khotskaya, Y. B., Sanchez, N. S., Zeng, J., Shufean, M. A., Shaw, K. R., Mendelsohn, J., Mills, G. B., Meric-Bernstam, F., and Simon, G. R. ALK: a tyrosine kinase target for cancer therapy. *Molecular Case Studies 3*, 1 (jan 2017), a001115.

[62] Holohan, C., Van Schaeybroeck, S., Longley, D. B., and Johnston, P. G. Cancer drug resistance: An evolving paradigm. *Nature Reviews Cancer 13*, 10 (oct 2013), 714–726.

[63] Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., and Sarkar, S. Drug resistance in cancer: An overview. *Cancers 6*, 3 (sep 2014), 1769–1792.

[64] Huang, L., and Fu, L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharmaceutica Sinica B 5*, 5 (sep 2015), 390–401.

[65] Ichihara, E., Ohashi, K., Takigawa, N., Osawa, M., Ogino, A., Tanimoto, M., and Kiura, K. Effects of vandetanib on lung adenocarcinoma cells harboring epidermal growth factor receptor T790M mutation in vivo. *Cancer Research 69*, 12 (jun 2009), 5091–5098.

[66] Iqbal, Z., Aleem, A., Iqbal, M., Naqvi, M. I., Gill, A., Taj, A. S., Qayyum, A., Ur-Rehman, N., Khalid, A. M., Shah, I. H., Khalid, M., Haq, R., Khan, M., Baig, S. M., Jamil, A., Abbas, M. N., Absar, M., Mahmood, A., Rasool, M., and Akhtar, T. Sensitive Detection of Pre-Existing BCR-ABL Kinase Domain Mutations in CD34+ Cells of Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia Patients Is Associated with Imatinib Resistance: Implications in the Post-Imatinib Era. *PLoS ONE 8*, 2 (2013).

[67] Iwasa, Y., Nowak, M. A., and Michor, F. Evolution of resistance during clonal expansion. *Genetics 172*, 4 (2006), 2557–2566.

[68] Jänne, P. A., Yang, J. C.-H., Kim, D.-W., Planchard, D., Ohe, Y., Ramalingam, S. S., Ahn, M.-J., Kim, S.-W., Su, W.-C., Horn, L., Haggstrom, D., Felip, E., Kim, J.-H., Frewer, P., Cantarini, M., Brown, K. H., Dickinson, P. A., Ghiorghiu, S., and Ranson, M. AZD9291 in EGFR Inhibitor–Resistant Non–Small-Cell Lung Cancer. *New England Journal of Medicine 372*, 18 (apr 2015), 1689–1699.

[69] Jiang, T., Su, C., Ren, S., Cappuzzo, F., Rocco, G., Palmer, J. D., van Zandwijk, N., Blackhall, F., Le, X., Pennell, N. A., and Zhou, C. A consensus on the role of osimertinib in non-small cell lung cancer from the AME Lung Cancer Collaborative Group. *Journal of Thoracic Disease 10*, 7 (jul 2018), 3909–3921.

[70] Jin, W., Shah, E. T., Penington, C. J., McCue, S. W., Maini, P. K., and Simpson, M. J. Logistic Proliferation of Cells in Scratch Assays is Delayed. *Bulletin of Mathematical Biology 79*, 5 (may 2017), 1028–1050.

[71] Kadi, N. E., Wang, L., Davis, A., Korkaya, H., Cooke, A., Vadnala, V., Brown, N. A., Betz, B. L., Cascalho, M., Kalemkerian, G. P., and Hassan, K. A. The EGFR T790M mutation is acquired through AICDA-mediated deamination of 5-methylcytosine following TKI treatment in lung cancer. *Cancer Research 78*, 24 (2018), 6728–6735.

[72] Kaznatcheev, A., Peacock, J., Basanta, D., Marusyk, A., and Scott, J. G. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nature Ecology & Evolution 3*, 3 (2019), 450–456.

[73] Kerr, K. M., and Lamb, D. Actual growth rate and tumour cell proliferation in human pulmonary neoplasms. *British Journal of Cancer 50*, 3 (1984), 343–349.

[74] King, K. L., and Cidlowski, J. A. Cell cycle and apoptosis: Common pathways to life and death. *Journal of Cellular Biochemistry 58*, 2 (jun 1995), 175–180.

[75] Kreutz, C., Raue, A., Kaschek, D., and Timmer, J. Profile likelihood in systems biology. *FEBS Journal 280*, 11 (jun 2013), 2564–2571.

[76] Kreutz, C., and Timmer, J. Systems biology: Experimental design. *FEBS Journal 276*, 4 (feb 2009), 923–942.

[77] Kuan, F. C., Kuo, L. T., Chen, M. C., Yang, C. T., Shi, C. S., Teng, D., and Lee, K. D. Overall survival benefits of first-line EGFR tyrosine kinase inhibitors in EGFR-mutated non-small-cell lung cancers: A systematic review and meta-analysis. *British Journal of Cancer 113*, 10 (nov 2015), 1519–1528.

[78] L., V. B. Quantitative laws in metabolism and growth. *The Quarterly review of biology 32*, 3 (1957), 217–231.

[79] Laird, A. K. Dynamics of tumor growth. *British Journal of Cancer 18*, 3 (1964), 490–502.

[80] Lavi, O., Gottesman, M. M., and Levy, D. The dynamics of drug resistance: A mathematical perspective. *Drug Resistance Updates 15*, 1-2 (2012), 90–97.

[81] Lee, Y., Lee, G. K., Lee, Y. S., Zhang, W., Hwang, J. A., Nam, B. H., Kim, S. H., Kim, J. H., Yun, T., Han, J. Y., Kim, H. T., and Lee, J. S. Clinical outcome according to the level of preexisting epidermal growth factor receptor T790M mutation in patients with lung cancer harboring sensitive epidermal growth factor receptor mutations. *Cancer 120*, 14 (jul 2014), 2090–2098.

[82] Leonetti, A., Sharma, S., Minari, R., Perego, P., Giovannetti, E., and Tiseo, M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *British Journal of Cancer 121*, 9 (oct 2019), 725–737.

[83] Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics 2*, 2 (jul 1944), 164–168.

[84] Ligon, T. S., Frohlich, F., Chis, O. T., Banga, J. R., Balsa-Canto, E., and Hasenauer, J. GenSSI 2.0: Multi-experiment structural identifiability analysis of SBML models. *Bioinformatics 34*, 8 (apr 2018), 1421–1423.

[85] Luria, S. E., and Delbrück, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics 28*, 6 (nov 1943), 491–511.

[86] Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., and Haber, D. A. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine 350*, 21 (may 2004), 2129–2139.

[87] Maini, P. K., McElwain, D. L., and Leavesley, D. I. Traveling Wave Model to Interpret a Wound-Healing Cell Migration Assay for Human Peritoneal Mesothelial Cells. *Tissue Engineering 10*, 3-4 (jul 2004), 475–482.

[88] Majeed, U., Manochakian, R., Zhao, Y., and Lou, Y. Targeted therapy in advanced non-small cell lung cancer: current advances and future trends. *Journal of Hematology and Oncology 14*, 1 (jul 2021), 1–20.

[89] Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., and Baradaran, B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Advanced pharmaceutical bulletin 7*, 3 (sep 2017), 339–348.

[90] Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology 254*, 1 (sep 2008), 178–196.

[91] Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics 11*, 2 (jun 1963), 431–441.

[92] Marusyk, A., Tabassum, D. P., Altrock, P. M., Almendro, V., Michor, F., and Polyak, K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature 514*, 7520 (oct 2014), 54–58.

[93] Mazza, V., and Cappuzzo, F. Treating EGFR mutation resistance in non-small cell lung cancer - role of osimertinib. *The application of clinical genetics 10* (2017), 49–56.

[94] Merlo, L. M., Pepper, J. W., Reid, B. J., and Maley, C. C. Cancer as an evolutionary and ecological process. *Nature Reviews Cancer 6*, 12 (dec 2006), 924–935.

[95] Meshkat, N., Er-zhen Kuo, C., and DiStefano, J. On finding and using identifiable parameter combinations in nonlinear dynamic systems biology models and combos: A novel web implementation. *PLoS ONE 9*, 10 (oct 2014), e110261.

[96] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*, 6 (dec 1953), 1087–1092.

[97] Miao, H., Xia, X., Perelson, A. S., and Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review 53*, 1 (2011), 3–39.

[98] Mok, T. S., Wu, Y.-L., Ahn, M.-J., Garassino, M. C., Kim, H. R., Ramalingam, S. S., Shepherd, F. A., He, Y., Akamatsu, H., Theelen, W. S., Lee, C. K., Sebastian, M., Templeton, A., Mann, H., Marotti, M., Ghiorghiu, S., and Papadimitrakopoulou, V. A. Osimertinib or Platinum–Pemetrexed in EGFR T790M–Positive Lung Cancer. *New England Journal of Medicine 376*, 7 (feb 2017), 629–640.

[99] Monro, H. C., and Gaffney, E. A. Modelling chemotherapy resistance in palliation and failed cure. *Journal of Theoretical Biology 257*, 2 (2009), 292–302.

[100] Murphy, S. A., and Van Der Vaart, A. W. On Profile Likelihood. *Journal of the American Statistical Association 95*, 450 (jun 2000), 449–465.

[101] Murtuza, A., Bulbul, A., Shen, J. P., Keshavarzian, P., Woodward, B. D., Lopez-Diaz, F. J., Lippman, S. M., and Husain, H. Novel third-generation EGFR tyrosine kinase inhibitors and strategies to overcome therapeutic resistance in lung cancer. *Cancer Research 79*, 4 (feb 2019), 689–698.

[102] Nakamura, R., Inage, Y., Tobita, R., Mori, K., Numata, T., Yanai, H., Endo, T., Ohtani, H., Satoh, H., Yuzawa, K., Koizumi, M., and Ueki, H. Epidermal growth factor receptor mutations: Effect on volume doubling time of non-small-cell lung cancer patients. *Journal of Thoracic Oncology 9*, 9 (2014), 1340–1344.

[103] Newville, M., Stensitzki, T., Allen, D. B., and Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python (Version 0.8.0). *Zenodo September* (sep 2014).

[104] Norton, L. A Gompertzian Model of Human Breast Cancer Growth. *Cancer Research 48*, 24 Part 1 (1988), 7067–7071.

[105] Norton, L., and Simon, R. Growth curve of an experimental solid tumor following radiotherapy. *Journal of the National Cancer Institute 58*, 6 (jun 1977), 1735–1741.

[106] Nowell, P. C. The clonal evolution of tumor cell populations. *Science 194*, 4260 (1976), 23–28.

[107] O'Leary, C., Gasper, H., Sahin, K. B., Tang, M., Kulasinghe, A., Adams, M. N., Richard, D. J., and O'Byrne, K. J. Epidermal growth factor receptor (EGFR)-mutated non-small-cell lung cancer (NSCLC). *Pharmaceuticals 13*, 10 (oct 2020), 1–16.

[108] Paczkowski, M., Kretzschmar, W. W., Markelc, B., Liu, S. K., Kunz-Schughart, L. A., Harris, A. L., Partridge, M., Byrne, H. M., and Kannan, P. Reciprocal interactions between tumour cell populations enhance growth and reduce radiation sensitivity in prostate cancer. *Communications Biology 4*, 1 (jan 2021), 1–13.

[109] Pao, W., Miller, V. A., Politi, K. A., Riely, G. J., Somwar, R., Zakowski, M. F., Kris, M. G., and Varmus, H. Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain. *PLoS Medicine 2*, 3 (feb 2005), e73.

[110] Pasi A. Janne. A Phase 1/2 Study of Osimertinib in Combination With Gefitinib in EGFR Inhibitor naïve Advanced EGFR Mutant Lung Cancer, 2017.

[111] Paz-Ares, L., Tan, E. H., O'Byrne, K., Zhang, L., Hirsh, V., Boyer, M., Yang, J. C., Mok, T., Lee, K. H., Lu, S., Shi, Y., Lee, D. H., Laskin, J., Kim, D. W., Laurie, S. A., Kölbeck, K., Fan, J., Dodd, N., Märten, A., and Park, K. Afatinib versus gefitinib in patients with EGFR mutation-positive advanced non-small-cell lung cancer: Overall survival data from the phase IIb LUX-Lung 7 trial. *Annals of Oncology 28*, 2 (feb 2017), 270–277.

[112] Pottier, C., Fresnais, M., Gilon, M., Jérusalem, G., Longuespée, R., and Sounni, N. E. Tyrosine kinase inhibitors in cancer: Breakthrough and challenges of targeted therapy. *Cancers 12*, 3 (mar 2020).

[113] Pucci, B., Kasten, M., and Giordano, A. Cell cycle and apoptosis. *Neoplasia 2*, 4 (2000), 291–299.

[114] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols 8*, 11 (oct 2013), 2281–2308.

[115] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics 25*, 15 (aug 2009), 1923–1929.

[116] Raue, A., Kreutz, C., Theis, F. J., and Timmer, J. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371*, 1984 (feb 2013).

[117] Richards, F. J. A flexible growth function for empirical use. *Journal of Experimental Botany 10*, 2 (jun 1959), 290–301.

[118] Román, M., Baraibar, I., López, I., Nadal, E., Rolfo, C., Vicent, S., and Gil-Bazo, I. KRAS oncogene in non-small cell lung cancer: clinical perspectives on the treatment of an old target. *Molecular cancer 17*, 1 (2018), 33.

[119] Rotow, J. K., Costa, D. B., Paweletz, C. P., Awad, M. M., Marcoux, P., Rangachari, D., Barbie, D. A., Sands, J., Cheng, M. L., Johnson, B. E., Oxnard, G. R., Jackman, D. M., Kwiatkowski, D. J., Kehl, K. L., Izdebski, M. D., Lau, C. J., Vasquez, K. A., and Janne, P. A. Concurrent osimertinib plus gefitinib for first-line treatment of EGFR-mutated non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology 38*, 15 (may 2020), 9507–9507.

[120] Saltelli, A., Tarantola, S., and Chan, P.-S. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics 41*, 1 (1999), 39–56.

[121] Sarapata, E. A., and de Pillis, L. G. A Comparison and Catalog of Intrinsic Tumor Growth Models. *Bulletin of Mathematical Biology 76*, 8 (aug 2014), 2010–2024.

[122] Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics 6*, 2 (1978), 461–464.

[123] Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., Mcdermott, U., Azizian, N., Zou, L., Fischbach, M. A., Wong, K.-k., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M., and Settleman, J. A chromatin-mediated reversible drug tolerant state in cancer cell subpopulations. *Cell 141*, 1 (2010), 69–80.

[124] Shi, P., Zhang, S., Zhu, L., Qian, G., Ren, H., Ramalingam, S. S., Chen, M., and Sun, S. Y. The Third-Generation EGFR Inhibitor, Osimertinib, Promotes c-FLIP Degradation, Enhancing Apoptosis Including TRAIL-Induced Apoptosis in NSCLC Cells with Activating EGFR Mutations. *Translational Oncology 12*, 5 (may 2019), 705–713.

[125] Simpson, M. J., Baker, R. E., Vittadello, S. T., and MacLaren, O. J. Practical parameter identifiability for spatio-temporal models of cell invasion. *Journal of the Royal Society Interface 17*, 164 (mar 2020).

[126] Simpson, M. J., Binder, B. J., Haridas, P., Wood, B. K., Treloar, K. K., McElwain, D. L., and Baker, R. E. Experimental and Modelling Investigation of Monolayer Development with Clustering. *Bulletin of Mathematical Biology 75*, 5 (apr 2013), 871–889.

[127] Simpson, M. J., Browning, A. P., Warne, D. J., Maclaren, O. J., and Baker, R. E. Parameter identifiability and model selection for sigmoid population growth models. *Journal of Theoretical Biology 535* (feb 2021), 110998.

[128] Smolle, E., Taucher, V., Lindenmann, J., Pichler, M., and Smolle-Juettner, F. M. Liquid biopsy in non-small cell lung cancer—current status and future outlook — A narrative review. *Translational Lung Cancer Research 10*, 5 (may 2021), 2237–2251.

[129] Soria, J.-C., Ohe, Y., Vansteenkiste, J., Reungwetwattana, T., Chewaskulyong, B., Lee, K. H., Dechaphunkul, A., Imamura, F., Nogami, N., Kurata, T., Okamoto, I., Zhou, C., Cho, B. C., Cheng, Y., Cho, E. K., Voon, P. J., Planchard, D., Su, W.-C., Gray, J. E., Lee, S.-M., Hodge, R., Marotti, M., Rukazenkov, Y., and Ramalingam, S. S. Osimertinib in Untreated EGFR -Mutated Advanced Non–Small-Cell Lung Cancer. *New England Journal of Medicine 378*, 2 (jan 2018), 113–125.

[130] Sosa Iglesias, V., Giuranno, L., Dubois, L. J., Theys, J., and Vooijs, M. Drug Resistance in Non-Small Cell Lung Cancer: A Potential for NOTCH Targeting? *Frontiers in Oncology 8* (jul 2018), 267.

[131] Steel, G. G. Cell loss as a factor in the growth rate of human tumours. *European Journal of Cancer (1965) 3*, 4-5 (1967), 381–387.

[132] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians 71*, 3 (may 2021), 209–249.

[133] Tabassum, D. P., and Polyak, K. Tumorigenesis: It takes a village. *Nature Reviews Cancer 15*, 8 (jul 2015), 473–483.

[134] Talkington, A., and Durrett, R. Estimating Tumor Growth Rates In Vivo. *Bulletin of Mathematical Biology 77*, 10 (oct 2015), 1934–1954.

[135] Tamirat, M. Z., Koivu, M., Elenius, K., and Johnson, M. S. Structural characterization of EGFR exon 19 deletion mutation using molecular dynamics simulation. *PLoS ONE 14*, 9 (sep 2019).

[136] Tsoularis, A., and Wallace, J. Analysis of logistic growth models. *Mathematical Biosciences 179*, 1 (jul 2002), 21–55.

[137] Vainstein, V., Eide, C. A., O'Hare, T., Shukron, O., and Druker, B. J. Integrating in vitro sensitivity and dose-response slope is predictive of clinical response to ABL kinase inhibitors in chronic myeloid leukemia. *Blood 122*, 19 (nov 2013), 3331–3334.

[138] Venzon, D. J., and Moolgavkar, S. H. A Method for Computing Profile-Likelihood-Based Confidence Intervals. *Applied Statistics 37*, 1 (1988), 87.

[139] Villaverde, A. F., Barreiro, A., and Papachristodoulou, A. Structural Identifiability of Dynamic Systems Biology Models. *PLoS Computational Biology 12*, 10 (2016), 1–22.

[140] Villaverde, A. F., Evans, N. D., Chappell, M. J., and Banga, J. R. Input-dependent structural identifiability of nonlinear systems. *IEEE Control Systems Letters 3*, 2 (2019), 272–277.

[141] Wang, Q., Yang, S., Wang, K., and Sun, S. Y. MET inhibitors for targeted therapy of EGFR TKI-resistant lung cancer. *Journal of Hematology and Oncology 12*, 1 (jun 2019).

[142] Wang, S., An, T., Wang, J., Zhao, J., Wang, Z., Zhuo, M., Bai, H., Yang, L., Zhang, Y., Wang, X., Duan, J., Wang, Y., Guo, Q., and Wu, M. Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer. *Clinical Cancer Research 16*, 4 (feb 2010), 1324–1330.

[143] Wang, S., Song, Y., and Liu, D. EAI045: The fourth-generation EGFR inhibitor overcoming T790M and C797S resistance. *Cancer Letters 385* (jan 2017), 51–54.

[144] Warne, D. J., Baker, R. E., and Simpson, M. J. Optimal Quantification of Contact Inhibition in Cell Populations. *Biophysical Journal 113*, 9 (2017), 1920–1924.

[145] Watanabe, M., Kawaguchi, T., Isa, S. I., Ando, M., Tamiya, A., Kubo, A., Saka, H., Takeo, S., Adachi, H., Tagawa, T., Kakegawa, S., Yamashita, M., Kataoka, K., Ichinose, Y., Takeuchi, Y., Sakamoto, K., Matsumura, A., and Koh, Y. Ultra-sensitive detection of the pretreatment EGFR T790M mutation in non-small cell lung cancer patients with an EGFR-activating mutation using droplet digital PCR. *Clinical Cancer Research 21*, 15 (aug 2015), 3552–3560.

[146] Weir, B. A., Woo, M. S., Getz, G., Perner, S., Ding, L., Beroukhim, R., Lin, W. M., Province, M. A., Kraja, A., Johnson, L. A., Shah, K., Sato, M., Thomas, R. K., Barletta, J. A., Borecki, I. B., Broderick, S., Chang, A. C., Chiang, D. Y., Chirieac, L. R., Cho, J., Fujii, Y., Gazdar, A. F., Giordano, T., Greulich, H., Hanna, M., Johnson, B. E., Kris, M. G., Lash, A., Lin, L., Lindeman, N., Mardis, E. R., McPherson, J. D., Minna, J. D., Morgan, M. B., Nadel, M., Orringer, M. B., Osborne, J. R., Ozenberger, B., Ramos, A. H., Robinson, J., Roth, J. A., Rusch, V., Sasaki, H., Shepherd, F., Sougnez, C., Spitz, M. R., Tsao, M.-S., Twomey, D., Verhaak, R. G. W., Weinstock, G. M., Wheeler, D. A., Winckler, W., Yoshizawa, A., Yu, S., Zakowski, M. F., Zhang, Q., Beer, D. G., Wistuba, I. I., Watson, M. A., Garraway, L. A., Ladanyi, M., Travis, W. D., Pao, W., Rubin, M. A., Gabriel, S. B., Gibbs, R. A., Varmus, H. E., Wilson, R. K., Lander, E. S., and Meyerson, M. Characterizing the cancer genome in lung adenocarcinoma. *Nature 450*, 7171 (dec 2007), 893–8.

[147] Wu, Y. L., Zhang, L., Kim, D. W., Liu, X., Lee, D. H., Yang, J. C. H., Ahn, M. J., Vansteenkiste, J. F., Su, W. C., Felip, E., Chia, V., Glaser, S., Pultar, P., Zhao, S., Peng, B., Akimov, M., and Tan, D. S. Phase Ib/II Study of Capmatinib (INC280) Plus Gefitinib after Failure of Epidermal Growth Factor Receptor (EGFR) Inhibitor Therapy in Patients with EGFR-Mutated, MET Factor-Dysregulated Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology 36*, 31 (nov 2018), 3101–3109.

[148] Yang, H., Liang, S. Q., Schmid, R. A., and Peng, R. W. New horizons in KRAS-mutant lung cancer: Dawn after darkness. *Frontiers in Oncology 9*, SEP (sep 2019), 953.

[149] Yang, Z., and Tam, K. Y. Combination Strategies Using EGFR-TKi in NSCLC Therapy: Learning from the Gap between Pre-Clinical Results and Clinical Outcomes. *International journal of biological sciences 14*, 2 (2018), 204–216.

[150] Yates, J. W., Dudley, P., Cheng, J., D'Cruz, C., and Davies, B. R. Validation of a predictive modeling approach to demonstrate the relative efficacy of three different schedules of the AKT inhibitor AZD5363. *Cancer Chemotherapy and Pharmacology 76*, 2 (aug 2015), 343–356.

[151] Yu, H. A., Sima, C., Feldman, D., Liu, L. L., Vaitheesvaran, B., Cross, J., Rudin, C. M., Kris, M. G., Pao, W., Michor, F., and Riely, G. J. Phase 1 study of twice weekly pulse dose and daily low-dose erlotinib as initial treatment for patients with EGFR-mutant lung cancers. *Annals of Oncology 28*, 2 (feb 2017), 278–284.

[152] Yuan, M., Huang, L. L., Chen, J. H., Wu, J., and Xu, Q. The emerging treatment landscape of targeted therapy in non-small-cell lung cancer. *Signal Transduction and Targeted Therapy 4*, 1 (dec 2019), 1–14.

[153] Yun, C. H., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K. K., Meyerson, M., and Eck, M. J. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences of the United States of America 105*, 6 (feb 2008), 2070–2075.

[154] Zappa, C., and Mousa, S. A. Non-small cell lung cancer: Current treatment and future advances. *Translational Lung Cancer Research 5*, 3 (jun 2016), 288–300.

[155] Zhang, J., Cunningham, J. J., Brown, J. S., and Gatenby, R. A. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature Communications 8*, 1 (dec 2017), 1–9.

[156] Zhang, R., Chen, B., Zhou, Y., Zhou, P., Jin, J., Zhao, L., and Li, W. Volume doubling time of lung adenocarcinomas considering epidermal growth factor receptor mutation status of exon 19 and 21: Three-dimensional volumetric evaluation. *Journal of Thoracic Disease 9*, 11 (nov 2017), 4387–4397.

[157] Zheng, Q. Progress of a half century in the study of the Luria-Delbruck distribution. *Mathematical Biosciences 162*, 1-2 (1999), 1–32.