Bayesian data assimilation and reinforcement learning for model-informed precision dosing in oncology

Corinna Sabrina Maier

Dissertation

zur Erlangung des akademischen Grades "doctor rerum naturalium" (Dr. rer. nat.) in der Wissenschaftsdisziplin "Angewandte Mathematik"

eingereicht an der Mathematisch-Naturwissenschaftlichen Fakultät Institut für Mathematik der Universität Potsdam



- 1. Reviewer: Prof. Dr. Wilhelm Huisinga
- 2. Reviewer: Ass. Prof. Dr. David Albers
- 3. Reviewer: Prof. Dr. Susanna Röblitz

Supervisors: Prof. Dr. Wilhelm Huisinga, Prof. Dr. Charlotte Kloft

Potsdam, 09. Juli 2021

Unless otherwise indicated, this work is licensed under a Creative Commons License Attribution-NonCommercial-NoDerivatives 4.0 International.

This does not apply to quoted content and works based on other permissions.

To view a copy of this license visit:

https://creativecommons.org/licenses/by-nc-nd/4.0/

Corinna Sabrina Maier

Bayesian data assimilation and reinforcement learning for model-informed precision dosing in oncology Dissertation zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. rer. nat), Potsdam, 09. Juli 2021 Reviewers: Prof. Dr. Wilhelm Huisinga (University of Potsdam), Associate Professor Dr. David Albers (University of Colorado), Prof. Dr. Susanna Röblitz (University of Bergen) Supervisors: Prof. Dr. Wilhelm Huisinga (University of Potsdam) and Prof. Dr. Charlotte Kloft (Freie Universität Berlin)

Universität Potsdam

Mathematische Modellierung und Systembiologie Institut für Mathematik Mathematisch-Naturwissenschaftliche Fakultät Karl-Liebknecht-Str. 24/25 14476 Potsdam

Published online on the Publication Server of the University of Potsdam: https://doi.org/10.25932/publishup-51587 https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-515870

Abstract

While patients are known to respond differently to drug therapies, current clinical practice often still follows a standardized dosage regimen for all patients. For drugs with a narrow range of both effective and safe concentrations, this approach may lead to a high incidence of adverse events or subtherapeutic dosing in the presence of high patient variability. Modelinformed precision dosing (MIPD) is a quantitative approach towards dose individualization based on mathematical modeling of dose-response relationships integrating therapeutic drug/biomarker monitoring (TDM) data. MIPD may considerably improve the efficacy and safety of many drug therapies. Current MIPD approaches, however, rely either on pre-calculated dosing tables or on simple point predictions of the therapy outcome. These approaches lack a quantification of uncertainties and the ability to account for effects that are delayed. In addition, the underlying models are not improved while applied to patient data. Therefore, current approaches are not well suited for informed clinical decision-making based on a differentiated understanding of the individually predicted therapy outcome.

The objective of this thesis is to develop mathematical approaches for MIPD, which (i) provide efficient fully Bayesian forecasting of the individual therapy outcome including associated uncertainties, (ii) integrate Markov decision processes via reinforcement learning (RL) for a comprehensive decision framework for dose individualization, (iii) allow for continuous learning across patients and hospitals. Cytotoxic anticancer chemotherapy with its major dose-limiting toxicity, neutropenia, serves as a therapeutically relevant application example.

For more comprehensive therapy forecasting, we apply Bayesian data assimilation (DA) approaches, integrating patient-specific TDM data into mathematical models of chemotherapy-induced neutropenia that build on prior population analyses. The value of uncertainty quantification is demonstrated as it allows reliable computation of the patient-specific probabilities of relevant clinical quantities, e.g., the neutropenia grade. In view of novel home monitoring devices that increase the amount of TDM data available, the data processing of sequential DA methods proves to be more efficient and facilitates handling of the variability between dosing events.

By transferring concepts from DA and RL we develop novel approaches for MIPD. While DA-guided dosing integrates individualized uncertainties into dose selection, RL-guided dosing provides a framework to consider delayed effects of dose selections. The combined DA-RL approach takes into account both aspects simultaneously and thus represents a holistic approach towards MIPD. Additionally, we show that RL can be used to gain insights into important patient characteristics for dose selection. The novel dosing strategies substantially reduce the occurrence of both subtherapeutic and life-threatening neutropenia grades in a simulation study based on a recent clinical study (CEPAC-TDM trial) compared to currently used MIPD approaches.

If MIPD is to be implemented in routine clinical practice, a certain model bias with respect to the underlying model is inevitable, as the models are typically based on data from comparably small clinical trials that reflect only to a limited extent the diversity in real-world patient populations. We propose a sequential hierarchical Bayesian inference framework that enables continuous cross-patient learning to learn the underlying model parameters of the target patient population. It is important to note that the approach only requires summary information of the individual patient data to update the model. This separation of the individual inference from population inference enables implementation across different centers of care.

The proposed approaches substantially improve current MIPD approaches, taking into account new trends in health care and aspects of practical applicability. They enable progress towards more informed clinical decision-making, ultimately increasing patient benefits beyond the current practice.

Zusammenfassung

Obwohl Patienten sehr unterschiedlich auf medikamentöse Therapien ansprechen, werden in der klinischen Praxis häufig noch standardisierte Dosierungsschemata angewendet. Bei Arzneimitteln mit engen therapeutischen Fenstern zwischen minimal wirksamen und toxischen Konzentrationen kann dieser Ansatz bei hoher interindividueller Variabilität zu häufigem Auftreten von Toxizitäten oder subtherapeutischen Konzentrationen führen. Die modellinformierte Präzisionsdosierung (MIPD) ist ein quantitativer Ansatz zur Dosisindividualisierung, der auf der mathematischen Modellierung von Dosis-Wirkungs-Beziehungen beruht und Daten aus dem therapeutischen Drug/Biomarker-Monitoring (TDM) einbezieht. Die derzeitigen MIPD-Ansätze verwenden entweder Dosierungstabellen oder einfache Punkt-Vorhersagen des Therapieverlaufs. Diesen Ansätzen fehlt eine Quantifizierung der Unsicherheiten, verzögerte Effekte werden nicht berücksichtigt und die zugrunde liegenden Modelle werden im Laufe der Anwendung nicht verbessert. Daher sind die derzeitigen Ansätze nicht ideal für eine fundierte klinische Entscheidungsfindung auf Grundlage eines differenzierten Verständnisses des individuell vorhergesagten Therapieverlaufs.

Das Ziel dieser Arbeit ist es, mathematische Ansätze für das MIPD zu entwickeln, die (i) eine effiziente, vollständig Bayes'sche Vorhersage des individuellen Therapieverlaufs einschließlich der damit verbundenen Unsicherheiten ermöglichen, (ii) Markov-Entscheidungsprozesse mittels Reinforcement Learning (RL) in einen umfassenden Entscheidungsrahmen zur Dosisindividualisierung integrieren, und (iii) ein kontinuierliches Lernen zwischen Patienten erlauben. Die antineoplastische Chemotherapie mit ihrer wichtigen dosislimitierenden Toxizität, der Neutropenie, dient als therapeutisch relevantes Anwendungsbeispiel.

Für eine umfassendere Therapievorhersage wenden wir Bayes'sche Datenassimilationsansätze (DA) an, um TDM-Daten in mathematische Modelle der Chemotherapie-induzierten Neutropenie zu integrieren. Wir zeigen, dass die Quantifizierung von Unsicherheiten einen großen Mehrwert bietet, da sie eine zuverlässige Berechnung der Wahrscheinlichkeiten relevanter klinischer Größen, z.B. des Neutropeniegrades, ermöglicht. Im Hinblick auf neue Home-Monitoring-Geräte, die die Anzahl der verfügbaren TDM-Daten erhöhen, erweisen sich sequenzielle DA-Methoden als effizienter und erleichtern den Umgang mit der Unsicherheit zwischen Dosierungsereignissen.

Basierend auf Konzepten aus DA und RL, entwickeln wir neue Ansätze für MIPD. Während die DA-geleitete Dosierung individualisierte Unsicherheiten in die Dosisauswahl integriert, berücksichtigt die RL-geleitete Dosierung verzögerte Effekte der Dosisauswahl. Der kombinierte DA-RL-Ansatz vereint beide Aspekte und stellt somit einen ganzheitlichen Ansatz für MIPD dar. Zusätzlich zeigen wir, dass RL Informationen über die für die Dosisauswahl relevanten Patientencharakteristika liefert. Der Vergleich zu derzeit verwendeten MIPD Ansätzen in einer auf einer klinischen Studie (CEPAC-TDM-Studie) basierenden Simulationsstudie zeigt, dass die entwickelten Dosierungsstrategien das Auftreten subtherapeutischer Konzentrationen sowie lebensbedrohlicher Neutropenien drastisch reduzieren.

Wird MIPD in der klinischen Routine eingesetzt, ist eine gewisse Modellverzerrung unvermeidlich. Die Modelle basieren in der Regel auf Daten aus vergleichsweise kleinen klinischen Studien, die die Heterogenität realer Patientenpopulationen nur begrenzt widerspiegeln. Wir schlagen einen sequenziellen hierarchischen Bayes'schen Inferenzrahmen vor, der ein kontinuierliches patientenübergreifendes Lernen ermöglicht, um die zugrunde liegenden Modellparameter der Ziel-Patientenpopulation zu erlernen. Zur Aktualisierung des Modells erfordert dieser Ansatz lediglich zusammenfassende Informationen der individuellen Patientendaten, was eine Umsetzung über verschiedene Versorgungszentren hinweg erlaubt.

Die vorgeschlagenen Ansätze verbessern die derzeitigen MIPD-Ansätze erheblich, wobei neue Trends in der Gesundheitsversorgung und Aspekte der praktischen Anwendbarkeit berücksichtigt werden. Damit stellen sie einen Fortschritt in Richtung einer fundierteren klinischen Entscheidungsfindung dar.

Acknowledgements

Throughout my doctoral studies and the writing of this thesis, I have received a great deal of both scientific and personal support.

First of all, I would like to thank my supervisor, Wilhelm Huisinga, for the regular meetings, his enthusiasm for research, his feedback and support, but also for giving me the freedom to develop my project in different directions. My sincere thanks also go to my co-supervisor, Charlotte Kloft, for her support regarding the application aspects of my work and for providing a clinical pharmacology perspective.

I am very grateful to be part of the Graduate Research Training Program PharMetrX, which provides a stimulating interdisciplinary environment, including academic and industry modules, research group meetings, the peer network, and financial support. I would like to thank my industry mentor, Sven Mensing, who shared his experience with me, discussed my PhD work with me, and gave me insights into the scientific challenges within drug discovery and development. I would like to thank him, Peter Nörtersheuser, and also all colleagues at AbbVie who made my stay at the Pharmacometrics group during my PhD time a unique, memorable, and enjoyable experience.

In addition, I am very thankful to Niklas Hartung without whom my PhD time would have been only half as fun and successful. My doctoral projects and this thesis benefited greatly from his extensive knowledge in the field, his critical thinking, and his eye for detail. I would also like to thank Jana de Wiljes who, with her incredible passion and enthusiasm for research, has kept me motivated and inspired me to explore the world of reinforcement learning.

I am also thankful that my project was associated to the Collaborative Research Center 1294 Data Assimilation. The seminars, colloquia, and spring schools were valuable scientific help and a great way to connect with fellow PhD students in different fields. In this respect, I am especially grateful to Sebastian Reich, Liv Heinecke, and to Alexandra Carpentier.

Also, I would like to thank all present and past working group members at the University of Potsdam: Jane Knöchel, Saskia Fuhrmann, Sabine Stübler, Christoph Hethey, Daniel Schindler, Undine Falkenhagen, Daniel Seeler, Andreas Braunß, and Katrin Kania as well as my PharMetrX colleagues Lena Klopp-Schulze, Lisa Ehmann, Sulav Duval, Ana-Marija Grišić, Franziska Kluwe, Francis Williams Ojara, Felix Jost, Sebastian Frank, Viktoria Stachanow, David Busse, Anna Müller-Schöll, Ferdinand Weinelt, Luis Ilia, Anja Lehmann, Alix Demaris, Yomna Nassar, Dimitra Eleftheriadou, Ayatallah Saleh and Robin Michelet. Also a big thank you goes to Olaf Dathe and Volker Gustavs for the support with computing on the university clusters. Special thanks go to Jana, Niklas, Yomna, Anna and Daniel for proof-reading parts of the thesis and to Andrea Henrich for answering all my questions regarding her PhD work.

A big thank you also goes to Carolin Schneider and Alexander Kutschera for a great week of thesis writing with mountain view and delicious hut food; and to Christine Huber and Martin Bachmeier for providing a valuable balance to thesis writing with climbing and hiking trips. Finally, I am very grateful to my parents Albert and Heidemarie Maier, my brother Markus Maier and Thomas Kleinert for always supporting me, encouraging me, and believing in me.

Publications

Parts of the present thesis are in preparation, are in the process of being published, or have been published in the following original articles:

Chapter 2.1, 2.2 and 3:

[CM1] C. Maier, N. Hartung, J. de Wiljes, C. Kloft, and W. Huisinga. "Bayesian Data Assimilation to Support Informed Decision Making in Individualized Chemotherapy". In: *CPT Pharmacometrics Syst. Pharmacol.* 9.3 (2020), pp. 153–164.

The implemented code is provided as supplementary material to the publication.

Chapter 2.3 and 4:

[CM2] C. Maier, N. Hartung, C. Kloft, W. Huisinga, and J. de Wiljes. "Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology". In: *CPT Pharmacometrics Syst. Pharmacol.* (2020). (accepted for publication).

The implemented code is publicly available under doi:10.5281/zenodo.3967011.

Chapter 5:

[CM3] C. Maier, J. de Wiljes, N. Hartung, C. Kloft, and W. Huisinga. "A continuous learning approach to model-informed precision dosing: accounting for model bias to meet clinical reality". (in preparation).

Contribution to the above three articles: The broad research ideas were conceptualized and results were continuously discussed with all co-authors. The author of this dissertation worked out the specific research objectives herself, performed the research in detail, implemented the code and wrote the first draft of the manuscripts. The co-authors have critically revised the article.

Introduction and Outlook:

[CM4] F. Kluwe, R. Michelet, A. Müller-Schöll, C. Maier, L. Klopp-Schulze, M. Dyk, G. Mikus, W. Huisinga, and C. Kloft. "Perspectives on Model-Informed Precision Dosing in the Digital Health Era: Challenges, Opportunities, and Recommendations". In: *Clin. Pharmacol. Ther.* (2020).

Contribution to the last article: The author of this dissertation contributed ideas, text suggestions and comments to the content of the article, and critically revised the manuscript.

Contents

1	Introduction						
2	Bac	kground	5				
	2.1	Oncology: chemotherapeutic agents and their dose-limiting toxicity	5				
		2.1.1 Chemotherapy-induced neutropenia	6				
		2.1.2 Dose individualization: CEPAC-TDM study as an example	8				
		2.1.3 PK/PD modeling of chemotherapy-induced neutropenia	9				
	2.2	Bayesian data assimilation (DA)	15				
	2.2	2.2.1 Batch DA approaches	$15 \\ 15$				
		2.2.1 Batch DA approaches 2.2.2 Sequential DA approaches	18				
	0.9		21				
	2.3	Reinforcement learning (RL)	24				
		2.3.1 Q-learning	26				
		2.3.2 Monte Carlo tree search (MCTS)	27				
		2.3.3 RL in health care	29				
3	Boy	resian therapy forecasting and clinical decision support	31				
J	Maximum a-posteriori (MAP) estimation: the current state-of-the-art	32					
	$3.1 \\ 3.2$,					
	3.2	Bayesian DA to quantify the uncertainty					
	ა.ა		34 34				
		3.3.1 Sampling Importance Resampling (SIR)					
		3.3.2 Markov Chain Monte Carlo (MCMC)	34				
		3.3.3 Particle filters (PFs)	35				
	3.4	Workflow in Bayesian forecasting	36				
		Informative decision support in oncology	39				
		3.5.1 Unfavorable properties of MAP-based predictions	40				
		3.5.2 Uncertainty quantification for informed decision-making	41				
		3.5.3 Comparable accuracy across full Bayesian approaches	43				
		3.5.4 Sequential DA processes patient data most efficiently	45				
	3.6	Discussion	46				
4	Dos	e individualization	49				
-	4.1	Offline approaches	43 52				
	4.1	4.1.1 Model-informed dosing tables	$52 \\ 52$				
			$\frac{52}{52}$				
	4.2	Online approaches	54				
		4.2.1 MAP-guided dosing	54				
		4.2.2 DA-guided dosing	55				

	4.34.44.5	Application to manage neutropenia in 3-weekly paclitaxel treatment54.4.1Novel approaches decrease occurrence of grade 4 & 0 neutropenia54.4.2Identification of relevant covariates via RL6	56 58 59 61 63
5	Con 5.1 5.2 5.3 5.4 5.5	Simulation study framework to investigate model bias	67 68 70 71 74 75 77 79
6	Out	look 8	33
Α	A.1	PK-guided dosing	85 85 85 85 86 87
В	B.1	Algorithmic details of MAP estimation 8 Additional analyses for the simulation studies 8 B.2.1 Single cycle study docetaxel 8	39 89 90 90 93
С	C.1 C.2 C.3 C.4 C.5	Comparison with reported CEPAC-TDM study outcomes9Sampling time points to infer neutropenia grade9Details on MAP-guided dosing10Details on DA-guided dosing10Details on RL-guided dosing10C.5.1 Computing the variance of the return10C.5.2 Investigating the effect of the tuning parameters10C.5.3 Q-planning as an alternative to MCTS10Details on DA-RL-guided dosing10C.6.1 Approaches for patient state estimation10C.6.2 RL-guided dosing based on DA state10C.6.3 PUCT algorithm10	97 98 01 02 03 03 04 07 08 09 09 09
	0.7	Comparison across an considered evaluation functions 1.	ŧυ

D	App	pendix related to Chapter 5	117
	D.1	Paclitaxel-induced neutropenia models	117
	D.2	Results including an estimation of γ	119
	D.3	Parameter identifiability	119
	D.4	Optimal design	120
	D.5	Impact on MIPD: parameter bias scenario	121
	D.6	Impact on MIPD: structural bias scenario	122

List of Figures

1.1	Overview of scientific challenges in MIPD	2
$2.1 \\ 2.2 \\ 2.3 \\ 2.4 \\ 2.5$	Typical neutropenia time course and statistics	$7 \\ 11 \\ 14 \\ 23 \\ 27$
3.1 3.2 3.3 3.4 3.5 3.6 3.7	Illustration of the particle filter algorithm	$35 \\ 37 \\ 38 \\ 41 \\ 42 \\ 44 \\ 45$
$\begin{array}{c} 4.1 \\ 4.2 \\ 4.3 \\ 4.4 \\ 4.5 \\ 4.6 \end{array}$	Overview of MIPD approaches	51 53 57 60 61 62
5.1 5.2 5.3 5.4 5.5	Hierarchical Bayesian model framework	70 76 77 78 79
A.1 B.1 B.2 B.3 B.4 B.5 B.6 B.7	Decision tree of the PK-guided dosing algorithm	 86 90 91 92 93 93 95 96

C.1	Comparison of simulation results with observed results in CEPAC-TDM trial	98
C.2	Comparison between model-predicted nadir and observation time points	98
C.3	Comparison of dosing policies for observation time point day 12	100
C.4	PK-guided dosing with observation day 12 vs. day 15	101
C.5	Comparison of utility versus target concentration for MAP-guided dosing	102
C.6	Exemplary dose selection in DA-guided dosing	103
C.7	Estimate of the action-value function across training stages of MCTS	104
C.8	Policy evaluation at different training stages of MCTS	105
C.11	Tuning parameters of MCTS: discount factor γ	105
C.12	Tuning parameters of MCTS: exploration/exploitation	106
C.13	Comparison of different reward functions in MCTS	107
C.14	Policy evaluation for different training stages in Q-planning	108
C.17	Comparison RL-guided dosing and DA-RL-guided dosing	110
C.18	Comparison of methods across all evaluation functions	111
C.9	Exemplary excerpt of the RL-guided dose selection	112
C.10	Statistics of MCTS for different states	113
C.15	Comparison of different approaches for state estimation	114
C.16	DA-RL-guided dosing without decision-time planning	115
D.1	Sequential updates illustrated for the rich sampling scheme	119
D.2	Parameter identifiability in the intermediate sampling scheme	120
D.3	Optimal design for the gold-standard model	121
D.4	Impact on MIPD: parameter bias scenario	122
D.5	Impact on MIPD: structural bias scenario with sparse sampling scheme	123
D.6	Impact on MIPD: structural bias scenario with rich sampling scheme	124

List of Tables

2.1	Parameter estimates of paclitaxel-induced neutropenia models	12
5.1	Hyper priors for the gold-standard model used in the simulation study	74
A.2	PK parameter estimates for docetaxel	87
	Notation related to RL-guided and DA-RL-guided dosing	
D.1	Paclitaxel-induced neutropenia models from literature	118

Acronyms

AAG	α 1-acid glycoprotein
ALB	albumin
ANC	absolute neutrophil counts
AI	artificial intelligence
BME	bone marrow exhaustion
BSA	body surface area
CEPAC-TDM	CESAR (Central European Society for Anticancer Drug Research) study of Paclitaxel Therapeutic Drug Monitoring
CI	
CTCAE	common terminology criteria for adverse events
\mathbf{CrI}	$ credible interval \dots \dots \dots \dots \dots 33 $
\mathbf{CV}	${\rm coefficient} \ of \ variation \ \ldots \ $
DA	data assimilation $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 3$
FIM	Fisher information matrix
G-CSF	granulocyte colony-stimulating factor
HPD	highest posterior density
iid	independent and identically distributed $\hdots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 17$
IIV	inter-individual variability $\ldots \ldots $ 8
ICU	intensive care unit
IOV	inter-occasion variability $\ldots \ldots \ldots$
i.v.	intravenous
KF	Kalman filter 19
MAP	maximum a-posteriori
MCMC	Markov chain Monte Carlo $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 17$
MCTS	Monte Carlo tree search $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 27$
MDP	Markov decision process
M-H	Metropolis-Hastings $\ldots \ldots 17$
MIDD	model-informed drug development $\ldots \ldots \ldots \ldots \ldots \ldots \ldots $
MIDT	model-informed dosing table $\ldots \ldots 50$

LIST OF ACRONYMS

MIPD	model-informed precision dosing $\ldots \ldots 1$
ML	machine learning
MLE	maximum likelihood estimate
MTT	mean transit time $\ldots \ldots \ldots$
NLME	nonlinear mixed effects
NAP	normal approximation
NSCLC	non-small cell lung cancer
ODE	ordinary differential equation
PD	pharmacodynamics
PF	particle filter
PI	prediction interval
PK	pharmacokinetics
POMDP	partially observable Markov decision process
PUCT	predictor+UCT
\mathbf{RL}	reinforcement learning
RMSE	root mean squared error
RSE	relative standard error
\mathbf{RUV}	residual unexplained variability
SIR	sampling importance resampling 17
SDE	stochastic differential equation
TD	temporal difference
TDM	the rapeutic drug/biomarker monitoring $\ldots \ldots \ldots \ldots \ldots \ldots 2$
\mathbf{TV}	typical value
UCB	upper confidence bound
UCT	upper confidence bound applied to trees
WBC	white blood cell

1 Introduction

It is well-known that individuals differ considerably in their capacity to absorb, metabolize, and eliminate drugs, which may lead to highly heterogeneous therapy outcomes [1, 2, 3]. Possible sources for this variability include genetic variations, environmental factors, age, and disease characteristics [4, 5]. Yet, the prevailing approach towards drug therapy regimens remains focused on an optimal uniform dosing that balances efficacy and safety across the entire patient population [6]. Accordingly, drug labels only provide very simplified dose recommendations that rarely take more than one patient characteristic into account [7]. In particular for drugs that exhibit narrow therapeutic ranges in relation to the variability between patients, this approach may lead to serious adverse events in some patients and to subtherapeutic exposure in others. Tailoring the dosing to each individual patient by balancing the patient-specific efficacy and safety bears, therefore, huge potential to improve overall drug treatment outcomes beyond the current practice [8].

One challenging example in which inadequate dosing may lead to serious patient outcomes is cytotoxic anticancer chemotherapy due to its associated severe toxicities [6]. It is common practice to adapt dosing of anticancer drugs to the patient's body surface area (BSA) including toxicity-related dose reductions [2]. Nevertheless, it has been observed that BSA-based dosing still leads to suboptimal and highly variable drug exposure, resulting in high variability of clinical outcomes with substantial occurrence of severe toxicities as well as unrecognized underdosing [1, 3, 9, 10, 11, 12]. Over the years, various patient factors and clinical markers (biomarkers) have been identified that are linked to the occurrence of severe side effects or to the overall therapy outcome (e.g., overall or progression-free survival) [10, 13, 14]. These established relationships provide a scientific rationale for dose adaptations by defining a well-informed target drug/biomarker concentration range (therapeutic window) for which a therapeutic effect is expected with minimal occurrence of toxicities [7]. To increase patient benefits, more elaborate approaches for dosing anticancer drugs are needed that reduce toxicities without compromising efficacy by integrating relevant patient information [6].

Model-informed precision dosing (MIPD) is a quantitative approach towards dose individualization, based on mathematical modeling and simulation integrating multiple sources of information [15]. The potential of MIPD to increase patient benefits has been largely recognized, however, the broad application in clinical practice is not yet established [16, 17]. There are multiple potential reasons and obstacles for this, e.g., regulatory and health care system barriers, as well as the lack of collaborative efforts, necessary clinical infrastructure

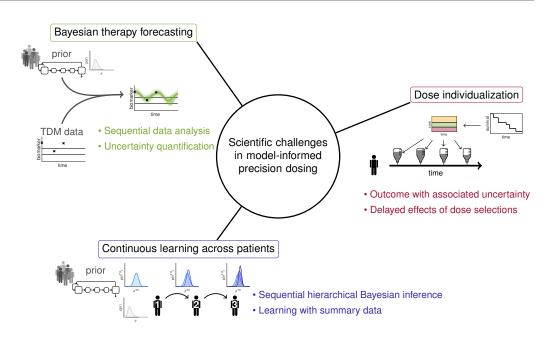


Figure 1.1: Overview of the scientific challenges in model-informed precision dosing addressed in this thesis.

and training of physicians in corresponding software tools [CM4]. But also scientific challenges remain (Figure 1.1) with respect to

- (i) reliable and efficient forecasting of the patient's individual therapy outcome,
- (ii) well-founded clinical decision support for individualized multiple dosage regimens, and
- (iii) continuous learning across patients to account for model bias.

Therapy forecasting. The first crucial aspect of MIPD is therapy forecasting, i.e., predicting the response of a patient to a given dose. Therapeutic drug/biomarker monitoring (TDM), i.e., collecting patient-specific data on drug or biomarker concentrations during ongoing treatment, provides a means to assess how the patient is responding to the treatment [18, 19]. TDM data, however, are often sparse and therefore as only source of information not well suited for informing a predictive individualized model. In this context, a Bayesian approach is very beneficial as it allows to integrate prior knowledge about the drug-patient-disease system [20]. The prior knowledge is generally obtained from population analyses of clinical studies which use a nonlinear mixed effects (NLME) framework to describe not only the typical therapy time course but also the associated variability between patients. This knowledge, in the form of mathematical models describing the pharmacokinetics (PK) and pharmacodynamics (PD) of the drug, can be used to predict the therapeutic outcome of a patient based on his/her characteristics (so called covariates) [21]. These (a-priori) predictions are, however, associated with uncertainty, resulting from the unexplained variability between patients. Applying Bayes formula to combine the prior knowledge with patient-specific TDM data results in a posterior distribution with reduced uncertainty. Ideally, all available information about the patient's individual parameters, quantified by the posterior, is used to predict the therapy outcome for a patient under uncertainty (*a-posteriori* predictions), see Figure 1.1 (top left). The currently

most widely used approach, however, is maximum a-posteriori (MAP) estimation, which only determines the mode of the posterior and thus neglects associated model uncertainties.

Bayesian data assimilation (DA) comprises algorithms for posterior inference by combining model-generated predictions with observational data for improved forecasting [22]. By approximating the full posterior, fully Bayesian DA algorithms enable a comprehensive uncertainty quantification on the level of the model parameters, which can be propagated consistently to the quantities of interest and thus provides a much more meaningful decision support. In the context of therapy forecasting, sequential DA approaches are of particular interest as they allow for a comprehensive uncertainty quantification as well as efficient sequential data processing while TDM data are collected. These approaches have become well-established in applications where real-time model predictions based on online/monitoring data are required. such as navigation [23], numerical weather prediction [24, 25] and object tracking [26, 27]. In health care these methods are only used to a limited extent, e.g., for glucose monitoring [28] or for intensive care unit (ICU) patients [29]. With the emergence of new mobile health care devices (e.g., wearables, home-monitoring, point-of-care testing [30, 17]) that enable more frequent assessment of the patient's health status, efficient sequential inference algorithms that also provide reliable and informative predictions based on quantified uncertainties are becoming increasingly important.

Dose individualization. The second aspect of MIPD is to provide well-informed decision support for optimal intervention for the individual patient. Optimal multiple dose selections can be regarded as a sequential decision-making problem under uncertainty [31]. Decisions must be made throughout the entire course of therapy, and a dose selection may have delayed or long-term consequences that influence subsequent dose decisions [3]. In addition, the therapeutic outcome in a patient for a certain dose is, as mentioned above, associated with uncertainty. Optimal dose selection should take into account the current patient status (including covariates and TDM measurements), the treatment history and should be made with respect to the evaluation of short-term markers as well as long-term therapeutic outcomes, e.g., overall or progression-free survival [3], see Figure 1.1 (right).

Reinforcement learning (RL) is a general class of algorithms in the field of machine learning (ML) that builds on the theory of Markov decision processes (MDPs) to formalize and solve decision-making problems where decisions are made in stages and are associated with uncertainty [32, 33]. The underlying idea is to learn from accumulated experience how to act best in uncertain environments, guided by a feedback or reward signal (as opposed to unsupervised learning), but without providing examples of 'correct behavior' (as in supervised learning). RL approaches have been developed mainly in the artificial intelligence (AI) community focusing on games with the goal that computers learn to play games better than humans [34, 35]. Over the years, RL has also gained popularity in other areas, including health care and also oncology; however, mainly focusing on clinical trial design [36, 37], and only few studies relate to optimal dosing in a PK/PD context [31, 38].

Continuous learning across patients. MIPD approaches are only as reliable as the underlying model used to predict the therapeutic outcome. When PK/PD models are used for MIPD a 'perfect model scenario' is often assumed, i.e., it is assumed that the model accurately reflects the drug-patient-disease system and that the variability observed in the patient population is adequately described. Structural models and prior model parameter distributions used in MIPD approaches build on prior clinical trials that include only a limited number of patients selected according to strict exclusion/inclusion criteria. Compared to the patient population in clinical practice, the prior clinical study population can be expected to

1 Introduction

introduce a model bias [15, 39]. The nature and extent of this bias are typically unknown, and it is not clear how a model bias could affect MIPD in routine clinical use. A critical aspect of applicability of MIPD is therefore to adapt the model to the target patient population as new patient data are observed. This requires a continuous learning approach that allows to update and improve the initial model across patients, see Figure 1.1 (bottom). With respect to data protection, an important consideration is that patient data may not be accessible across different hospitals or institutions and therefore model learning should be based on summary information of the data that can be shared. Continuous model learning based on an ever-growing amount of data has enormous potential to improve the predictive capabilities of the models and thus enable well-founded MIPD for the large variability encountered in the whole patient population.

Outline In this thesis the two fields Bayesian DA and RL are introduced in the context of clinical pharmacology and it is shown how these methods can be used and combined to advance the reliability, accuracy, and applicability of MIPD. The resulting approaches are discussed and illustrated using an example in oncology: chemotherapy-induced neutropenia, the most frequent dose-limiting toxicity in cytotoxic anticancer chemotherapy. In the next chapter (Chapter 2), the background to the application field, oncology, and the different methodological fields, DA and RL, is provided, and general concepts are introduced.

In Chapter 3, Bayesian DA approaches for therapy forecasting are comprehensively compared with respect to their reliability, efficiency, and ability to support informed decisionmaking. The unfavorable properties of MAP estimation, currently the most widely used approach, are outlined in a relevant clinical context. Further, we show that a comprehensive uncertainty quantification as provided by fully Bayesian DA approaches provides a more informative and differentiated understanding of the forecasted therapy outcome and thus better-informed decision-making. Finally, the advantages of sequential DA approaches are presented in view of novel point-of-care devices.

Next, novel approaches for MIPD based on DA and/or RL are developed in Chapter 4. DAguided dosing includes the individualized uncertainty quantification into the objective function for dose selection. RL-guided dosing is based on the solution of the stochastic sequential decision-making problem and DA-RL-guided dosing provides a combined framework leveraging the posterior information provided by DA in RL. In a simulation study, we compare the three proposed approaches, DA-guided, RL-guided and DA-RL-guided dosing with current dosing strategies in terms of dosing performance and their ability to provide insights into the factors driving dose selection.

In order to increase the applicability of MIPD in clinics, we discuss in Chapter 5 how a bias in the underlying models can be corrected via continuous learning across patients to bridge the gap between academia/industry and clinical practice. In particular, we propose a sequential hierarchical Bayesian framework with two stages, that separates the inference on the individual patient level from the update of the prior knowledge (on the population level) for the next patient. This separated framework facilitates practical implementation as the patient data themselves do not need to be shared across hospitals, which is advantageous compared to approaches based on pooling patient data.

Finally, in Chapter 6 the proposed approaches for MIPD are positioned in the broader context of health care and general aspects of MIPD that need to be addressed in the future are outlined.

2 Background

This thesis leverages, adapts, and combines the methodology from different research areas to advance approaches towards MIPD in oncology and beyond. Here, a brief overview of the different areas is given first, introducing general and important concepts as well as domain-specific terminology used throughout this work.

2.1 Oncology: chemotherapeutic agents and their doselimiting toxicity

Cancer is the second leading cause of death worldwide (after cardiovascular diseases) and is estimated to have accounted for 9.6 million deaths in 2018 [40]. Cancer is a family of many diseases, characterized by uncontrolled, excessive growth, and spread of abnormal cells [41]. The malignant transformation of normal cells into cancer cells is a multi-step process involving cellular and genetic changes that give cancer cells a growth advantage, enable them to invade other tissues, and overcome cell death [42]. Depending on the site of origin, different types of cancer are distinguished; the most common being lung, breast, colorectal, prostate, stomach, and cervical cancer [40]. Lung cancer is not only the most common but also the deadliest type [43], accounting for the highest number of deaths (18% of cancer deaths [44]).

General treatment strategies in oncology include surgery, radiotherapy, chemotherapy, novel targeted therapies as well as immunotherapy [45, 46]. Despite advances in novel treatment options, traditional chemotherapy remains an integral component of cancer treatment, e.g., in chemoimmunotherapy [12, 47, 48]. Cytotoxic chemotherapy has a non-specific mechanism of action affecting the cell cycle of rapidly dividing cells of all types [49]. This induces the desired cytotoxic effect on malignant cancer cells but also affects healthy dividing cells, causing severe side effects. Therefore, improved therapeutic management is essential to minimize the side effects of cancer treatment while increasing the benefits and hence address the major public health burden of cancer.

The cytotoxic chemotherapeutic agents of interest in this thesis are paclitaxel (Taxol[®]), marketed since 1993) and docetaxel (Taxotere[®]), marketed since 1996). Both drugs belong to the class of taxanes and function as mitotic inhibitors via enhancing and stabilizing the polymerization of microtubules [50, 51]. Since microtubules are essential for the formation of the spindle apparatus during mitosis, the drugs prevent cell division. Paclitaxel is used either as a single agent or in combination with other drugs, e.g., carboplatin or cisplatin [52], against ovarian [53], breast [54], and non-small cell lung cancer (NSCLC) [55]. Docetaxel is indicated (alone or in combination) for breast cancer [51], NSCLC, prostate, gastric [56] as well as head and neck cancer. The main associated toxicities of docetaxel and paclitaxel chemotherapy are peripheral neuropathy [57] and hematological toxicity, most importantly neutropenia [58].

2.1.1 Chemotherapy-induced neutropenia

Neutropenia is a severe reduction of neutrophil granulocytes, which can be induced by cytotoxic agents [58]. Neutrophils are the most abundant type of white blood cells and play a key role in the immune system as they serve as primary responders to infections [59]. Therefore, neutropenia can lead to life-threatening immunodeficiencies associated with the incidence of infections and fever [58].

Healthy hematopoiesis. Hematopoiesis refers to the formation of blood cells, which all originate from hematopoietic stem cells in the bone marrow. Stem cells have the ability to replicate in order to sustain the stem cell pool (self-renewal) as well as to differentiate and mature to all cell types (pluripotency) to maintain stable blood counts (homeostasis) [60]. Proceeding from the stem cells, the common myeloid or lymphoid progenitors form the starting point for the two main cell lineages: the myeloid arm comprising red blood cells, platelets, and white blood cells (granulocytes), and the lymphoid arm giving rise to T and B-lymphocytes. Hematopoiesis is a highly complex hierarchical process involving the differentiation and maturation over various progenitor (mitotic, i.e., proliferating) and precursor (post-mitotic) stages until ultimately mature blood cells are released into the systemic circulation. The mature cell types have often a rather short life span, e.g., the half-life of neutrophils is approximately 7 h [61]. Therefore, a balance between proliferation, differentiation, and release is essential, which is ensured by strong regulation via different cytokines, e.g., granulocyte colony-stimulating factor (G-CSF) for granulopoiesis (production of granulocytes) [62].

The role of neutropenia in clinical decision-making. The perturbation of normal hematopoiesis caused by cytotoxic agents has a major impact on the dosing and scheduling of chemotherapy. For illustration, a typical time course of neutropenia after administration of docetaxel is depicted in Figure 2.1. The baseline neutrophil value, prior to the dose administration, should be above $1.5 \cdot 10^9$ cells/L for docetaxel (Taxotere[®] label) and paclitaxel (Taxol[®] label). After drug administration (at time t = 0), the neutrophil concentration is decreased with a delay. This delay in response is due to the fact that mainly proliferating hematopoietic progenitor cells in the bone marrow are vulnerable to the cytotoxic effect and that the maturation process to mature neutrophils takes several days (mitotic 5–7.5 days [64, 65] & post-mitotic 4–6.6 days [66, 62]). The nadir, the lowest neutrophil concentration, is typically reached around day 9 for docetaxel [67] and day 12 for paclitaxel [11]. The different grades g of neutropenia range according to the common terminology criteria for adverse events (CTCAE) [63] from no neutropenia (g = 0) over mild (g = 1), moderate (g = 2), severe (g = 3), to life-threatening (g = 4), and death (g = 5), see Figure 2.1. Recovery of neutropenia is initiated via increased G-CSF levels due to low neutrophil counts [68]. Only when the neutrophil concentration is back at a sufficient level $(>1.5 \cdot 10^9 \text{ cells/L})$, the next dose can be administered, leading to treatment cycles of 1-4 weeks [69, 70].

Due to the high risk of life-threatening infections associated with neutropenia grade 4, neutropenia restricts the maximum tolerable dose and insufficient recovery of neutropenia could delay subsequent treatments [58]. In contrast, neutropenia during chemotherapy has

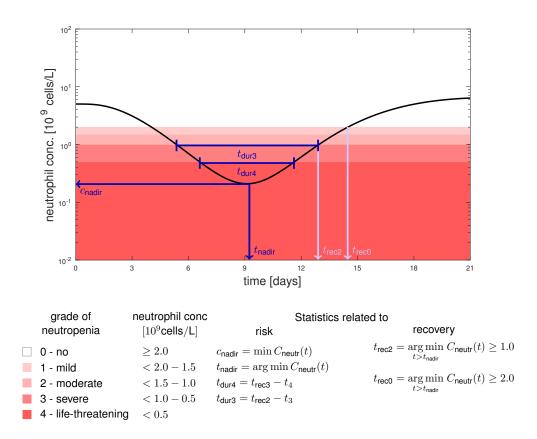


Figure 2.1: Typical time course of neutropenia after administration of an anticancer drug. The neutrophil concentration $C_{\text{neutr}}(t)$ over time is shown for docetaxel (100 mg/m² BSA, 1 h intravenous infusion). Neutropenia grades are defined according to the common terminology criteria for adverse events (CTCAE) [63]. Note that the shades of red are related to increasing toxicity, however, grade 0 (white) over the whole cycle is associated with ineffective treatment. Key characteristics for decision-making in cytotoxic chemotherapy related to risk (dark blue) and recovery (light blue) of neutropenia are indicated on the typical time course: the lowest neutrophil concentration (c_{nadir}), the time at which the nadir is reached (t_{nadir}), the duration of neutropenia grade 3 and grade 4 (t_{dur3} and t_{dur4} , respectively), as well as the times until recovery to neutropenia grade 2 and 0 (t_{rec2} and t_{rec0} , respectively). Note that the rebound (overshooting of the neutrophil concentration above the baseline) is not as apparent due to the log-scale.

been also associated with increased median survival [13, 71, 72]. A possible explanation for this observation is that the absence of neutropenia could indicate subtherapeutic exposure, resulting in a sub-optimal effect [73]. A moderate degree of toxicity is therefore desirable, and consequently, neutropenia also functions as an early and easily measurable marker (surrogate) for treatment efficacy [74]. Based on this established relationship between neutropenia and efficacy as well as toxicity, neutrophil counts were suggested to be used as a biomarker to guide dosing [75, 76, 73].

For decision-support in cytotoxic chemotherapy certain characteristics of the neutropenia time course related to risk and recovery should be considered [73]. The time t_{nadir} , at which the nadir is reached, is important for time management of intervention, e.g., intravenous antibiotics and hospitalization [58]. In addition, risk is also related to the duration t_{dur3} and t_{dur4} of an individual being in grade 3 and 4 neutropenia, respectively [58]. The patient could be considered out of risk at time t_{rec2} , when neutropenia grade 2 is reached post nadir. The recovery time to grade 0, $t_{\rm rec0}$, could be conservatively used to schedule the initiation of the next treatment cycle.

Measuring neutrophil concentrations. Due to the associated risks, neutropenia is routinely assessed during chemotherapy treatment (as recommended by drug labels). Absolute neutrophil counts (ANC) are measured within complete blood cell counts using hematology analyzers [77]. In general clinical settings, neutrophil counts are measured before the start of the treatment (baseline absolute neutrophil counts, ANC_0), during the cycle to assess toxicity and before administration of the next dose to ensure recovery [52, 30]. Recent advancements in point-of-care testing now also allow more frequent assessment of neutrophil counts, e.g., HemoCue[®] WBC DIFF system for white blood cell (WBC) counts (HemoCue AB, Angelholm, Sweden) [78]. These point-of-care testing devices rely on capillary finger-prick blood samples which can be taken by the patients themselves at home [30]. It has been demonstrated that the capillary neutrophil counts measured via HemoCue WBC Diff correlate well with neutrophil counts measured in venous blood using state-of-the-art hematology analyzers [79, 80] [30, chapter 5]. The precision, in terms of the coefficient of variation (CV), of HemoCue WBC Diff is stated in the operating manual as $\sim 7\%$ CV for low WBC levels and $\sim 4\%$ CV for normal and high WBC levels. In comparison, the precision of state-of-the-art hematology analyzers is given as 10% CV for low neutrophil levels and as 2.5% CV for normal levels [77]. Overall, the data situation can be expected to change in the near future, allowing for accurate and frequent hematologic assessment, which enables more adaptive dosing approaches based on patient-specific neutrophil concentration measurements.

2.1.2 Dose individualization: CEPAC-TDM study as an example

To achieve therapeutic success, informed decision-making with respect to the dose and the schedule is critical [81, 45]. The current administration of anticancer drugs is generally adapted to BSA, e.g., 100 mg/m² BSA 1 h intravenous (i.v.) infusion for docetaxel [82]. In addition, toxicity-related dose reductions are indicated [3]. Despite dose adjustments to BSA, high inter-individual variability (IIV) in efficacy and toxicity outcomes is observed, which demands more refined approaches to increase patient benefits [2, 12, 6].

To individualize paclitaxel treatment, a dosing algorithm (called *PK-quided dosing*) based on the patient's sex, age, BSA, as well as drug exposure and toxicity was developed [11], and evaluated in the clinical trial CESAR (Central European Society for Anticancer Drug Research) study of Paclitaxel Therapeutic Drug Monitoring (CEPAC-TDM) [ClinicalTrials.gov Identifier: NCT01326767 [83] versus standard dosing 200 mg/m² BSA (including toxicity-related dose reductions of 20% if grade 4 neutropenia was observed in the previous treatment cycle) [52]. Paclitaxel was given as 3 h i.v. infusion in combination with carboplatin or cisplatin. In the PK-guided dosing, the dose of the first cycle was determined based on the patient's age and sex. For subsequent cycles, the dose was adjusted according to exposure (time during which the drug concentration is above $0.05 \,\mu\text{M}$) and neutropenia grade of the previous cycle (inferred from observed ANC at day 15), see Figure A.1. The algorithm relies on a mathematical model describing the PK of paclitaxel to infer the individual exposure measure [11]. The study population of the CEPAC-TDM study comprised 365 patients with advanced NSCLC. NSCLC accounts for 84% of lung cancer cases and has poor prognosis [41]. The study did not succeed in reducing life-threatening grade 4 neutropenia, however, paclitaxel-induced neuropathy was reduced, which could substantially improve the patient's quality of life. Also, a similar treatment response has been achieved by administering significantly lower doses [52].

Neutrophil-guided dosing has been suggested earlier using a PK/PD model that describes

chemotherapy-induced neutropenia [75, 76]. The patient's neutrophil counts can be used to infer the individual model parameters within a Bayesian forecasting framework to predict the individual neutropenia time course. The dose could then be optimized with respect to a target nadir concentration or a utility function [76, 84]. The dosing strategies mentioned serve in Chapter 4 as a benchmark for the novel proposed MIPD dosing strategies.

PK/PD modeling of chemotherapy-induced neutropenia 2.1.3

Mathematical models describing the PK/PD of a drug play a key role in academia/industry, e.g., for dose selection, clinical trial design and to demonstrate safety as well as efficacy of a drug candidate [85, 86]. For model-informed drug development (MIDD) PK/PD models are typically developed based on clinical trial data using an NLME framework to describe not only the typical therapy outcome but also the variability in outcome between patients (population analysis) [87]. The PK model links a dose d to a drug's concentration time course $C_{\rm drug}(t)$ in plasma. PK models generally have a compartmentalized structure that groups together regions of the body that behave kinetically similar [88]. The PD model relates the drug concentration to a response via a mechanism of drug action [89]. The models can incorporate mechanistic/semi-mechanistic aspects, i.e., explicitly describing the underlying physiological process, or data-driven/empirical, i.e., without physiological interpretation. In PK/PD models the rate of change of concentrations or amounts over time for one individual *i* is mathematically described via a system of ordinary differential equations (ODEs), the structural model,

$$\frac{\mathrm{d}x_i}{\mathrm{d}t}(t) = f(x_i(t); \theta_i, d_i), \qquad x_i(t_0) = x_0(\theta_i)$$

$$h_i(t) = h(x_i(t), \theta_i), \qquad (2.2)$$

$$h_i(t) = h(x_i(t), \theta_i), \qquad (2.2)$$

with state vector $x_i = x_i(t) \in \mathbb{R}^{n_x}$ (incl. drug/biomarker concentrations), individual parameter values $\theta_i \in \mathbb{R}^{n_{\theta}}$ (e.g., volumes, clearances) and rates of change $f(x_i; \theta_i, d_i)$ of all state variables for a given input $d_i \in \mathbb{R}^{n_d}$ (e.g., dose). Since typically only a part of the state variables is observed, the observational model Eq. (2.2) maps x_i to the observed quantities $h(x_i(t), \theta_i)$, e.g., plasma drug or neutrophil concentration, including potential state-space transformations (e.g., log-transformed output). The initial conditions $x_0(\theta_i)$ are defined by the pre-treatment levels (e.g., baseline values).

Population analyses integrate data from multiple patients, $i = 1, \ldots, N$ of a clinical study. The individual model parameters are not directly estimated based on individual data but modeled as random variables with specified probability density functions of which the parameters (population parameters) are estimated.

The covariate and statistical model link patient-specific factors (covariates) ' cov_i ' and observations $(t_{ij}, y_{ij})_{j=1,...,n_i}$ to the model predictions $h_{ij}(\theta_i) = h(x_i(t_j), \theta_i)$,

$$\Theta_i \sim p_{\Theta}(\cdot; \theta^{\mathrm{TV}}(\mathrm{cov}_i), \Omega), \qquad (2.3)$$

$$[Y_{ij}|\Theta_i = \theta_i] \sim p(\cdot |\theta_i; h_{ij}(\theta_i), \Sigma), \qquad j = 1, \dots, n_i \quad \text{(independent)}$$
(2.4)

where $\theta^{\text{TV}}(\text{cov}_i)$ denotes the typical value (TV) that might depend on covariates, Ω denotes the magnitude of the unexplained IIV, i.e., the variability between patients in clinical outcome that cannot be explained by covariates, and Σ the residual unexplained variability (RUV), accounting for measurement errors and (possibly) model misspecification. Often an additive error model is chosen, $[Y_{ij}|\Theta_i = \theta_i] = h_{ij}(\theta_i) + \epsilon_{ij}$ with $\epsilon_{ij} \sim_{\text{iid}} \mathcal{N}(0, \Sigma)$ potentially on a log-scale. In case of positive physiological parameters, i.e., $\theta_i \in \mathbb{R}^{n_{\theta}}_+$, the most widely used parametric assumption for the individual parameters Eq. (2.3) is the lognormal distribution, i.e., $\Theta_i = \theta^{\text{TV}}(\text{cov}_i) \cdot e^{\eta_i}$, with $\eta_i \sim_{\text{iid}} \mathcal{N}(0, \Omega)$. The goal of a population analysis within a NLME framework is to estimate the population parameters θ^{TV} , Ω and Σ , and to identify significant covariates that influence the PK or PD of the drug and thus can explain parts of the observed variability in clinical outcome, i.e., estimate the functional relationship $\text{cov} \mapsto \theta^{\text{TV}}(\text{cov})$ [67]. These established relationships can be used to forecast the therapy outcome for a patient (under uncertainty) based on his/her characteristics.

Docetaxel pharmacokinetic model. We employed a published model which describes docetaxel PK by three compartments with first-order elimination [82]. Clearance of docetaxel was modeled as function of the covariates α 1-acid glycoprotein (AAG), age, BSA, and ALB (plasma albumin). The system of ODEs is provided in Section A.2.1 and parameter estimates by Bruno et al. [82] in Table A.1. The typical drug concentration time course is depicted in Figure 2.3 A.

Paclitaxel pharmacokinetic model Paclitaxel PK was previously described by a three compartment model with nonlinear distribution to the first peripheral compartment and nonlinear elimination [11], see Figure 2.2 (left part) for a schematic representation of the compartmental model structure. A typical concentration time course is shown in Figure 2.3 B. For our analyses, we used the (re-)estimated parameter values by Henrich et al. [90, Table 1], see also Table A.2. The PK model includes a covariate model on the maximum elimination capacity VM_{EL} :

$$\begin{split} \mathrm{VM}_{\mathrm{EL,TV},i} = & \mathrm{VM}_{\mathrm{EL,pop}} \cdot \left(\frac{\mathrm{BSA}_{i}}{1.8 \, \mathrm{m}^{2}}\right)^{\theta_{\mathrm{VM}_{\mathrm{EL}}\text{-}\mathrm{BSA}}} \cdot \left(\theta_{\mathrm{VM}_{\mathrm{EL}}\text{-}\mathrm{SEX}}\right)^{\mathrm{SEX}_{i}} \cdot \\ & \left(\frac{\mathrm{AGE}_{i}}{56 \, \mathrm{years}}\right)^{\theta_{\mathrm{VM}_{\mathrm{EL}}\text{-}\mathrm{AGE}}} \cdot \left(\frac{\mathrm{BILI}_{i}}{7 \, \mathrm{\mu M}}\right)^{\theta_{\mathrm{VM}_{\mathrm{EL}}\text{-}\mathrm{BILI}}}, \end{split}$$

with SEX (0/1 for female/male) of patient *i*, AGE (in years) and bilirubin concentration BILI, a marker for liver function, (in μ M). A lognormal distribution was assumed for the statistical IIV model, Eq. (2.3). In addition to IIV and RUV, inter-occasion variability (IOV) was included in the model, on the parameters, VM_{EL} as well as the central volume of distribution V_1 . An occasion was defined as the start of a treatment cycle c,

$$\Theta_{i,c} = \theta^{\mathrm{TV}}(\mathrm{cov}_i) \cdot e^{\eta_i + \kappa_{i,c}}, \qquad \eta_i \stackrel{iid}{\sim} \mathcal{N}(0,\Omega), \ \kappa_{i,c} \stackrel{iid}{\sim} \mathcal{N}(0,\Pi).$$
(2.5)

Therefore, model parameters may vary to some extent (given by Π) between dosing events within one patient.

Models for chemotherapy-induced neutropenia The most well-known and widely used model (the gold-standard) describing chemotherapy-induced neutropenia for various drugs is a semi-mechanistic model developed by Friberg et al. [91], Figure 2.2 (right part, black). The model consists of five compartments reflecting the maturation from rapidly proliferating progenitor cells ('Prol') in the bone marrow (gray dashed box) to neutrophils in the systemic circulation ('Circ'). Three transit compartments ('Transit 1–3') approximate the delay (also known as 'linear chain trick' [92, 93]) caused by the different maturation stages, cf. Section 2.1.1 with $k_{\rm tr} = 4/{\rm MTT}$ denoting the transition rate constant given by the number of transit compartments plus one divided by the mean transit time (MTT). The

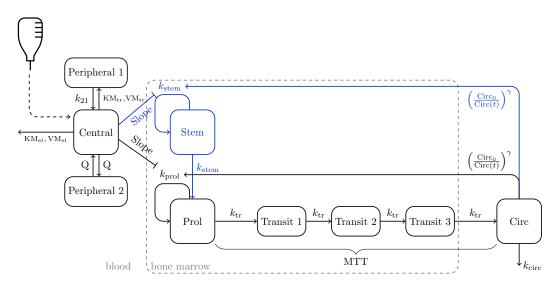


Figure 2.2: Schematic representation of paclitaxel-induced neutropenia models. The model describing paclitaxel pharmacokinetics (PK) is depicted on the left with a central compartment and two peripheral compartments. In black the gold-standard pharmacodynamic (PD) model for neutropenia is depicted, including a compartment for proliferating progenitor cells 'Prol' in the bone marrow and three transit compartments 'Transit 1–3' representing the maturation to circulating neutrophils 'Circ' in the blood [91]. The extension including a stem cell compartment 'Stem' to account for the long-term exhaustion of the bone marrow is depicted in blue [90]. The parameters and the corresponding model equations are described in the text. The figure is based on [90, Fig. 1].

drug concentration in the central compartment of the PK model, $C_1 = \text{Cent}/V_1$, is modeled to have a linear (inhibitory) effect on the proliferation rate k_{prol} of 'Prol', $E_{\text{drug}} = \text{Slope} \cdot C_1$. The transit compartments represent post-mitotic stages and are therefore not (directly) affected by the drug. The feedback of low neutrophils (Circ(t) small) compared to baseline neutrophils (Circ₀) increases the proliferation rate initiating the recovery. The degradation rate of neutrophils k_{circ} is typically set to equal k_{tr} .

The gold-standard model for neutropenia [91] serves as a starting point for the development of various models for different drugs, see Table A.3 for docetaxel, and for different patient populations for the same drug, e.g., for paclitaxel [9, 11, 67, 94]. For paclitaxel we will focus, in this thesis, on the models summarized in Table 2.1. The gold-standard (original) model was used to develop the previously described *PK-guided dosing* algorithm [11] (Section 2.1.2 & A.1). Retrospectively, the model parameters were re-estimated for the CEPAC-TDM data (gold-standard (re-estimated) [95]).

In the CEPAC-TDM study [52], cumulative neutropenia was observed over multiple treatment cycles, i.e., the nadir and the maximum neutrophil concentration decreased over the course of treatment. A potential hypothesis for this cumulative behavior is that the drug also affects the long-term recovery of the bone marrow, causing bone marrow exhaustion (BME) [90]. The gold-standard model for neutropenia by Friberg et al. [91] does not describe this long-term effect and was shown to overpredict neutrophil concentrations at later cycles [95, section 3.3]. Therefore, Henrich et al. [90] extended the model to include a stem cell compartment 'Stem', representing pluripotent stem cells with slower proliferation, which are also affected by the drug, see Figure 2.2 (black & blue). The system of ODEs describing the structural model reads

2.1 Oncology: chemotherapeutic agents and their dose-limiting toxicity

Table 2.1: Parameter estimates for paclitaxel-induced neutropenia models retrieved from [95]. The parameter estimates of the gold-standard (original) model are based on [11]. The gold-standard (re-estimated) model denotes the unchanged structural model but the parameter estimates were re-estimated using the CEPAC-TDM study data [95]. The bone marrow exhaustion (BME) model [90] describes the long-term effect over multiple cycles. The typical value (TV) parameters describe the fixed effects and the inter-individual variability (IIV) parameters the magnitude of the variability between patients. The residual unexplained variability (RUV) an exponential model was chosen for all models. The IIV and RUV parameters are provided as coefficient of variation (CV). If originally published, the relative standard errors (RSEs) are provided in brackets (in %). Note that baseline method B2 [96] was used for baseline absolute neutrophil counts ANC₀, i.e., the IIV was estimated together with RUV as one single parameter. The table is modified from [95, Table 3.7]

Parameter	$\operatorname{Gold-standard}(\operatorname{original})$	Gold-standard (re-estimated)	Bone marrow exhaustion
	Joerger et al. (2012)	Henrich (2017) Thesis	Henrich et al. (2017)
TV parameters			
MTT [h]	141	128 (2.03)	145 (2.65)
$\mathrm{Slope} \; [\mathrm{L}/\mu\mathrm{mol}]$	2.6	4.48(4.55)	$13.1 \ (4.56)$
γ	0.2	$0.231 \ (6.79)$	0.257 (5.53)
ftr	-	-	0.787(2.76)
IIV parameters $(CV\%)$			
MTT	27.0	-	-
Slope	44.9	43.8(8.23)	44.8(6.54)
$ANC_0 (=Circ_0)$	31.6	60.3(3.27)	51.5(3.61)
RUV parameters (CV%)			
exp. model	31.6	60.3 (3.27)	51.5(3.61)

$$\begin{split} \frac{d\operatorname{Stem}(t)}{dt} &= k_{\operatorname{stem}}\operatorname{Stem}(t) \cdot (1 - E_{\operatorname{drug}}(t)) \cdot \left(\frac{\operatorname{Circ}_{0}}{\operatorname{Circ}(t)}\right)^{\gamma} \\ &- k_{\operatorname{stem}}\operatorname{Stem}(t) , & \operatorname{Stem}(0) = \operatorname{Circ}_{0} \\ \frac{d\operatorname{Prol}(t)}{dt} &= k_{\operatorname{prol}}\operatorname{Prol}(t) \cdot (1 - E_{\operatorname{drug}}(t)) \cdot \left(\frac{\operatorname{Circ}_{0}}{\operatorname{Circ}(t)}\right)^{\gamma} \\ &+ k_{\operatorname{stem}}\operatorname{Stem}(t) - k_{\operatorname{tr}}\operatorname{Prol}(t) , & \operatorname{Prol}(0) = \operatorname{Circ}_{0} \\ \frac{d\operatorname{Transit1}(t)}{dt} &= k_{\operatorname{tr}}\operatorname{Prol}(t) - k_{\operatorname{tr}}\operatorname{Transit1}(t) , & \operatorname{Transit1}(0) = \operatorname{Circ}_{0} \\ \frac{d\operatorname{Transit2}(t)}{dt} &= k_{\operatorname{tr}}\operatorname{Transit1}(t) - k_{\operatorname{tr}}\operatorname{Transit2}(t) , & \operatorname{Transit2}(0) = \operatorname{Circ}_{0} \\ \frac{d\operatorname{Transit3}(t)}{dt} &= k_{\operatorname{tr}}\operatorname{Transit2}(t) - k_{\operatorname{tr}}\operatorname{Transit3}(t) , & \operatorname{Transit3}(0) = \operatorname{Circ}_{0} \\ \frac{d\operatorname{Circ}(t)}{dt} &= k_{\operatorname{tr}}\operatorname{Transit3}(t) - k_{\operatorname{circ}}\operatorname{Circ}(t) , & \operatorname{Circ}(0) = \operatorname{Circ}_{0} \end{split}$$

with $k_{\text{circ}} = k_{\text{tr}}$ and $E_{\text{drug}} = \text{Slope} \cdot C_1$ representing the linear (inhibitory) drug effect and $C_1 = \text{Cent}/V_1$, the drug concentration in the central compartment of the paclitaxel PK model. Note that the model implicitly assumes that the volumes of all compartments are identical. Henrich et al. [90] calibrated this model in a population analysis to the CEPAC-TDM study data [52], see Table 2.1. The proliferation rates for the two compartments Prol and Stem are given by

$$k_{\text{prol}} = \text{ftr} \cdot k_{\text{tr}}$$

 $k_{\text{stem}} = (1 - \text{ftr}) \cdot k_{\text{tr}},$

respectively, with ftr denoting the fraction of replication in 'Prol'. The IIV, Eq. (2.3), is modeled to be lognormal. The baseline neutrophil count $\text{Circ}_0 = \text{ANC}_0$ was modeled based on the baseline data point y_{i0} (using the baseline method B2 in [96])

$$\operatorname{Circ}_{0,i} = y_{i0} \cdot e^{\sigma \cdot \eta_{\operatorname{Circ}_{0},i}}, \qquad \eta_{\operatorname{Circ}_{0},i} \sim \mathcal{N}(0,1), \qquad (2.6)$$

with σ denoting the univariate RUV. This extension leads to a gradual decrease in the nadir and the maximum neutrophil concentration (recovery) over multiple treatment cycles, see typical time course in Figure 2.3 C and comparison to the models based on the structure of the gold-standard model in panel D. Thus, the model was able to describe cumulative neutropenia as observed in the CEPAC-TDM study [90]. It has been shown that paclitaxel is the main reason for neutrophil toxicity when given in combination with carboplatin [9]. Therefore, we restricted our analysis to paclitaxel-induced neutropenia alone although paclitaxel was given in combination with carboplatin or cisplatin in the CEPAC-TDM study. The gold-standard model for docetaxel is used in Chapter 3, and the BME model builds the basis for the simulation studies in Chapters 3 & 4. In Chapter 5 the three different models in Table 2.1 are used to investigate the effect of a model bias on MIPD.

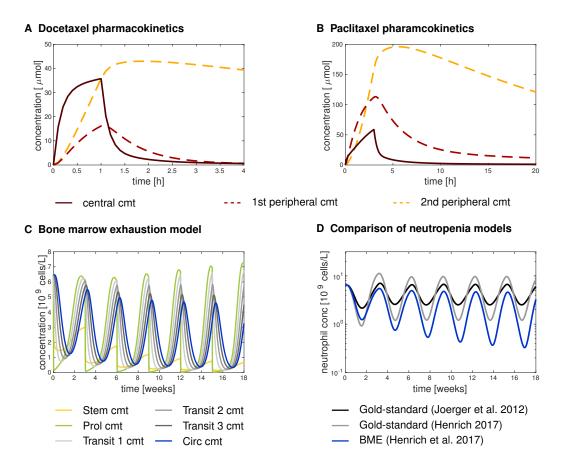


Figure 2.3: Typical model predictions for docetaxel pharmacokinetics ($100 \text{ mg/m}^2 \text{BSA}$, 1 h i.v. infusion) (A), paclitaxel pharmacokinetics ($200 \text{ mg/m}^2 \text{BSA}$, 3 h i.v. infusion) (B), paclitaxel-induced bone marrow exhaustion (BME) (C). (D) the typical model predictions are compared for the gold-standard model using the parametrization of Joerger et al. [11], and using the (re-)parametrization of Henrich (2017) [95] with the model extension to describe bone marrow exhaustion (BME) by Henrich et al. [90]. Typical patient characteristics were for docetaxel: female, age = 56 years, AAG = 1.34 g/L, BSA = 1.76 m^2 , ALB = 41 g/L [67] and for paclitaxel: male, age = 56 years, Circ₀ = $6.48 \cdot 10^9 \text{ cells/L}$, BSA = 1.8 m^2 and BILI = 7 µmol/L [52, 95].

2.2 Bayesian data assimilation (DA)

DA refers to a family of procedures that allows incorporating observations into a mathematical model of a system in order to improve the prediction skills of the model [22]. These methods are highly valuable tools to fill the gap between insufficient or partial data for purely data-driven approaches and imperfect or incomplete models for purely model-driven approaches [28]. An overview of the different approaches towards DA is provided in Figure 2.4. DA algorithms comprise methods that process data in a batch (all data points at once) or sequentially as they become available (one data point at a time). Only selected algorithms are presented here; a more comprehensive overview and discussion of DA algorithms can be found for example in [22, 27, 97].

2.2.1 Batch DA approaches

Within a Bayesian framework, the data are not regarded to be stand-alone information about the system of interest, rather system parameters are seen as random variables, and prior information about them is incorporated using Bayes' formula

$$p(\theta|y_{1:n}) = \frac{p(y_{1:n}|\theta)p(\theta)}{p(y_{1:n})},$$
(2.7)

with $p(\theta)$ denoting the prior distribution of the model parameters, $p(y_{1:n}|\theta)$ denoting the likelihood, the probability of the data $y_{1:n} = (y_1, \ldots, y_n)^T$ given the model parameters θ , and $p(\theta|y_{1:n})$ the posterior, i.e., the conditional distribution of the parameters given the data. The denominator in Eq. (2.7)

$$p(y_{1:n}) = \int_{\mathbb{R}^{n_{\theta}}} p(y_{1:n}|\theta) p(\theta) d\theta$$
(2.8)

serves as a normalization factor, denoting the marginal probability of the data (also often called the evidence). In most realistic problems, the normalization constant Eq. (2.8) is difficult to compute due to the involved (possibly high-dimensional) integral that makes posterior inference analytically intractable [27, section 2.5] [98]. Even if it is possible to calculate the normalization constant, sampling from the posterior may not be, which might be required to derive posterior-based quantities, e.g., marginals or posterior expectations. Typically, the unnormalized posterior

$$\tilde{p}(\theta|y_{1:n}) = p(y_{1:n}|\theta)p(\theta) \tag{2.9}$$

can only be evaluated point-wise, which might also be computationally expensive for complex models as evaluation of the likelihood might require solving complex/stiff ODEs or partial differential equations.

Maximum a-posteriori (MAP) estimation Variational DA methods transform the posterior inference problem into an optimization problem, minimizing a cost functional over all data points [97]. A widely used approach is MAP estimation, i.e., seeking the parameters that maximize the posterior distribution,

$$\hat{\theta}_{n}^{\text{MAP}} = \arg\max_{\theta} p(\theta|y_{1:n}) = \arg\max_{\theta} p(y_{1:n}|\theta)p(\theta) , \qquad (2.10)$$

which only requires the point-wise evaluation of the unnormalized posterior Eq. (2.9). The posterior distribution is summarized by a point-estimate, the mode (the MAP estimate), without measure of the uncertainty associated with it, see Figure 2.4 left.

Normal approximation (NAP) For an extended resolution of the posterior, a parametric approximation to the posterior can be used. Most often a normal approximation (NAP) centered at the mode of the posterior is considered, approximating the variance of the distribution using the curvature at the mode:

$$p(\cdot|y_{1:n}) \approx \mathcal{N}\left(\hat{\theta}_n^{\mathrm{MAP}}, \mathcal{I}^{-1}\left(\hat{\theta}_n^{\mathrm{MAP}}\right)\right),$$
 (2.11)

where \mathcal{I} denotes the total observed Fisher information matrix (FIM) [99, section 2.5]

$$\mathcal{I}(\theta) := \mathcal{I}^{\text{post}}(\theta) = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log p(\theta|y_{1:n})$$

of the posterior. The approximation Eq. (2.11) can be derived using a Taylor expansion of the log-posterior at its mode (following [100, section 4.1]):

$$\log p(\theta|y_{1:n}) \approx \underbrace{\log p(\hat{\theta}_n^{\text{MAP}}|y_{1:n})}_{=\text{const.}} + (\theta - \hat{\theta}_n^{\text{MAP}})^T \underbrace{\left[\frac{\mathrm{d}\log p(\theta|y_{1:n})}{\mathrm{d}\theta}\right]_{\theta = \hat{\theta}_n^{\text{MAP}}}}_{=0} + \frac{1}{2}(\theta - \hat{\theta}_n^{\text{MAP}})^T \Big[\frac{\mathrm{d}^2\log p(\theta|y_{1:n})}{\mathrm{d}\theta^2}\Big]_{\theta = \hat{\theta}_n^{\text{MAP}}}(\theta - \hat{\theta}_n^{\text{MAP}}).$$

Exponentiating both sides and normalizing leads to

$$p(\theta|y_{1:n}) \approx \mathcal{N}\left(\hat{\theta}_n^{\mathrm{MAP}}, \left[-\frac{\mathrm{d}^2 \log p(\theta|y_{1:n})}{\mathrm{d}\theta^2}\right]_{\theta=\hat{\theta}_n^{\mathrm{MAP}}}^{-1}\right).$$

The inverse of the variance can be decomposed using Bayes' formula Eq. (2.7)

$$-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log p(\theta|y_{1:n}) = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log p(y_{1:n}|\theta) - \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log p(\theta), \qquad (2.12)$$

where we retrieve the total observed FIM

$$\mathcal{I}^{\text{likelihood}}(\theta) = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log p(y_{1:n}|\theta)$$

If the definition for the observed FIM of the likelihood is transferred to prior and posterior, the previous decomposition Eq. (2.12) can be written as

$$\mathcal{I}^{\text{post}}(\theta) = \mathcal{I}^{\text{likelihood}}(\theta) + \mathcal{I}^{\text{prior}}(\theta) \,,$$

which allows to rewrite the normal approximation of the posterior as given in Eq. (2.11). This closed-form approximation is only well suited for uni-modal and symmetric posterior distributions, see Figure 2.4. The approximation, however, is supported by the asymptotic normality of the posterior provided by large-sample theory (Bernstein-von-Mises Theorem [101], see [100, appendix B] for a proof) and can often be improved by transformations [100].

Since these assumptions do not hold in many problem settings, non-parametric approximations can be used to resolve more complex shapes of the posterior, e.g., multi-modality or skewness. Instead of optimization (as in variational DA) or a parametric approximation, full Bayesian inference methods employ a sample approximation to the full posterior

$$p(\theta|y_{1:n}) \approx \sum_{m=1}^{M} w_n^{(m)} \mathbf{1}_{\{\theta_n^{(m)} = \theta\}},$$
(2.13)

based on a sample (called ensemble in DA literature)

$$\mathcal{E}_n := \left\{ \left(\theta_n^{(m)}, w_n^{(m)}\right), \quad m = 1, \dots, M \right\},$$
(2.14)

with sample parameters $\theta_n^{(m)}$ of the posterior [27, section 2.5], weights $w_n^{(m)}$, which sum to one, i.e., $\sum_{m=1}^{M} w_n^{(m)} = 1$. If $w_n^{(m)} = 1/M$ for all $m = 1, \ldots, M$, the sample is called unweighted (or uniformly weighted); leading to the well-known Monte Carlo approximation. Since direct sampling from the posterior is in general not possible, alternative approaches need to be employed to generate a sample of the posterior.

Sampling Importance Resampling (SIR) The sampling importance resampling (SIR) algorithm is based on importance sampling [27, 102, 103]. It generates an unweighted sample \mathcal{E}_n from the posterior $p(\theta|y_{1:n})$ using a sample from a so called importance distribution π_I , from which samples can easily be generated [100, section 10.4]. Here, the non-iterative SIR is presented as opposed to the sequential importance resampling algorithm (also often abbreviated with SIR; see also the following paragraph about particle filters) and proceeds in only three steps:

- S Step: independent and identically distributed (iid) sampling from the importance distribution π_I results in a sample $\tilde{\mathcal{E}}_n = {\{\tilde{\theta}_n^{(m)}\}}_{m=1}^M$.
- I Step: An importance weight is assigned to each sample point $\tilde{\theta}_n^{(m)} \in \tilde{\mathcal{E}}_n$

$$\tilde{w}_{n}^{(m)} = \frac{p(y_{1:n}|\bar{\theta}_{n}^{(m)})p(\bar{\theta}_{n}^{(m)})}{\pi_{I}(\tilde{\theta}_{n}^{(m)})}, \qquad (2.15)$$

given by the ratio of the unnormalized posterior Eq. (2.9) and the importance distribution evaluated at the sample point.

R Step: After normalization of the weights $w_n^{(m)} = \tilde{w}_n^{(m)} / \sum_m \tilde{w}_n^{(m)}$, a resampling step is performed: *M* unweighted samples $\theta_n^{(m)}$ are drawn from $\tilde{\mathcal{E}}_n$ according to weights $w_n^{(m)}, m = 1, \ldots, M$. There exist many different strategies on how to resample efficiently and effectively. The most widely used are multinomial, residual and systematic resampling, see e.g., [104].

Generally, this algorithm needs a large number of samples, especially if there is a large disagreement between the importance distribution and the target distribution. This is computationally expensive but can be run in parallel up to the normalization of the weights.

Markov chain Monte Carlo (MCMC) In contrast, Markov chain Monte Carlo (MCMC) methods construct a Markov chain $\{\theta^{(m)}|m = 0, 1, ...\}$ with the posterior as stationary distribution [100]. Generating draws from the Markov chain to approximate the posterior can be done with different algorithms; popular choices are the Metropolis-Hastings (M-H) algorithm [105, 106] or the Gibbs sampler [107, 108].

The M-H sampler requires a proposal distribution $\pi_P(\cdot|\theta^{(m)})$ which specifies how to move from a given state $\theta^{(m)}$ in the Markov chain to the next state $\theta^{(m+1)}$, and must satisfy certain regularity conditions so that the Markov chain converges to the correct stationary distribution, i.e., irreducibility and ergodicity (aperiodicity and positive recurrence). First, the Markov chain is initialized, i.e., $\theta^{(0)}$ is defined. Then, in each iteration $m = 1, \ldots, M$ a proposal θ^* is sampled from $\pi_P(\cdot|\theta^{(m-1)})$. The proposal is accepted with probability

$$\alpha = \min\left[1, \frac{\tilde{p}(\theta^*|y_{1:n})/\pi_P(\theta^*|\theta^{(m-1)})}{\tilde{p}(\theta^{(m-1)}|y_{1:n})/\pi_P(\theta^{(m-1)}|\theta^*)}\right],$$
(2.16)

which only involves pointwise evaluation of the unnormalized posterior as the normalization constant cancels in the ratio. If the proposal is accepted, the Markov chain moves to the proposal $\theta^{(m)} = \theta^*$, else the Markov chain remains at its current location $\theta^{(m)} = \theta^{(m-1)}$. The choice of the proposal distribution is key for constructing an efficient simulation algorithm. It is required to be able to sample from the proposal distribution, the parameter space should be reasonably explored, and the proposals should not be rejected too often (the optimal acceptance rate in one dimension is 44% and in higher dimensions 23%) [100, section 11-12]. The original Metropolis sampler requires the proposal to be symmetric which simplifies the acceptance probability to $\alpha = \min[1, \tilde{p}(\theta^*|y_{1:n})/\tilde{p}(\theta^{(m-1)}|y_{1:n})]$. A special case is the random walk Metropolis sampler that relies on a normal proposal distribution. In the independence sampler, the proposal distribution does not depend on the current location of the Markov chain, which leads to

$$\alpha = \min\left[1, \frac{\tilde{p}(\theta^*|y_{1:n})/\pi_P(\theta^*)}{\tilde{p}(\theta^{(m-1)}|y_{1:n})/\pi_P(\theta^{(m-1)})}\right].$$

For the independence sampler to be efficient, the proposal needs to be close to the target distribution.

A different approach to generate draws from a Markov chain is the Gibbs sampler, which can be very efficiently employed in multidimensional inference problems and is based on alternate conditional sampling [100, section 11]. Suppose θ can be split into K parts, $\theta = (\theta_1, \ldots, \theta_K)$. Define $\theta_{-k} = (\theta_1, \ldots, \theta_{k-1}, \theta_{k+1}, \ldots, \theta_K)$, the parameter vector without component θ_k . The Markov chain is generated by sampling in each iteration m from each of the full conditionals $p(\theta_k | \theta_{-k}^{(m-1)}, y_{1:n})$. Therefore, the algorithm requires that sampling from the full conditional distributions of the components is possible. For hierarchical models, the conditionals are often constructed as conjugate distributions, which allow straightforward sampling within the Gibbs sampler. The advantage of the method is that no rejections take place; the acceptance probability equals one which means that fewer iterations are required.

A critical aspect of MCMC methods is, that the Markov chain needs to run long enough to ensure convergence. There exist various diagnostics to monitor convergence, see [109]. Generally, a certain number of initial samples is discarded, a so called 'burn-in' or 'warm-up'. However, there is no method for determining the hyperparameters to achieve a good acceptance rate, which means that for many application problems a considerable tuning effort is required. In addition, the generated samples $\theta_n^{(m)}$ are not iid samples from $p(\theta|y_{1:n})$. Within MCMC research one important aspect is therefore to investigates methods to reduce autocorrelation. Standard MCMC methods cannot be used efficiently in a sequential inference context, i.e., in which data points are collected over time, as for every updated posterior distribution $p(\theta|y_{1:n+1})$ a new Markov chain has to be generated [110] and parallelization is restricted to running multiple chains in parallel.

2.2.2 Sequential DA approaches

Many application fields demand efficient real-time predictions based on monitoring data, e.g., as mentioned meteorology, navigation, or tracking. This need motivated the development of sequential DA approaches, in which the posterior is iteratively updated via Bayes' formula by combining computer-generated Bayesian forecasts with real-time incoming data [22]. The recursive update of the posterior relies on a sequential formulation of Bayes' formula

$$p(\theta|y_{1:n+1}) \propto p(y_{n+1}|\theta) \cdot p(\theta|y_{1:n}), \qquad (2.17)$$

in which the current posterior functions as prior for the upcoming data point, see also Figure 2.4 (right). For n = 0, the distribution $p(\theta|y_{1:0})$ is identical to the prior $p(\theta)$ [27, section 3]. Note that for the likelihood holds $p(y_{n+1}|\theta) = p(y_{n+1}|y_{1:n},\theta)$ due to independence in the statistical error model Eq. (2.4).

The most well-known sequential DA method is the Kalman filter (KF) [111, 112], which relies on the assumptions of linear model dynamics and Gaussian uncertainty. Thus, it constitutes a parametric approach, which reduces the Bayesian inference problem to tracking only the first and second moment (mean and variance) of the posterior density forward in time as more data are observed.

Particle filters (PF) As aforementioned many application problems do not satisfy the necessary assumptions for the closed-form solution of the KF. Therefore, particle filters (PFs) [113, 114] were developed that allow for non-Gaussian error models and nonlinear structural models. PFs are based on an ensemble prediction with ensemble members referred to as *particles*. Filtering algorithms were mainly developed for state estimation with fixed parameters. However, the parameters can be added to the state, which results in an augmented state space $z = (x, \theta)$,

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = f(x(t);\theta,d)$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}t}(t) = 0.$$
(2.18)

For static parameters, the rate of change of the parameters is zero. In conjunction with importance sampling, a weighted sample of the posterior is carried forward in time and re-weighted as new data points become available [22]. The basic version of PF is closely related to the previously presented batch algorithm SIR, but relies on the sequential Bayes' formula Eq. (2.17) and is therefore also referred to as sequential importance resampling. To avoid confusion with the batch SIR algorithm, this term will not be used in the following.

Given a weighted sample \mathcal{E}_n of the posterior $p(\cdot|y_{1:n})$ and a new data point y_{n+1} , PF generates a weighted sample \mathcal{E}_{n+1} of the posterior $p(\cdot|y_{1:n+1})$ by updating \mathcal{E}_n using Eq. (2.17). Analogously to the I-step in SIR, the weights $w_n^{(m)}$ are updated proportionally to the (local) likelihood

$$\tilde{w}_{n+1}^{(m)} \propto p(y_{n+1}|\theta) \cdot w_n^{(m)},$$
(2.19)

involving, however, only the new data point y_{n+1} , and normalized (as in the R-step of SIR). The bootstrap filter [113] performs resampling at every step (as in SIR), however, it is more efficient to only perform a resampling if too many samples carry an almost negligible weight and the total weight is limited to only a few samples (weight degeneracy). An often used criterion is based on the effective sample size

$$M_{\rm eff}(t_n) := \frac{1}{\sum_{m=1}^{M} \left(w_n^{(m)} \right)^2}$$
(2.20)

to decide whether to resample. Starting initially with uniform weights $w_0^{(m)} = 1/M$ with $M_{\text{eff}}(t_0) = M$, resampling should be carried out once $M_{\text{eff}} < M/2$ (effective ensemble size half of the initial ensemble size). If resampling is performed, it is followed by a so called rejuvenation step [22] to prevent sample impoverishment by fixation to limited parameter values:

$$\theta_n^{(m)} = \tilde{\theta}_n^{(m)} + \xi_n^{(m)}, \quad \text{with } \xi_n^{(m)} \sim_{\text{iid}} \mathcal{N}(0, \tau B)$$
(2.21)

with rejuvenation parameter $\tau > 0$ and covariance matrix $B \in \mathbb{R}^{n_{\theta} \times n_{\theta}}$, where $\tilde{\theta}_{n}^{(m)}$ denotes the resampled parameters. For sufficiently small τ this procedure only introduces small perturbations in the parameter space, hence it can be assumed that $x(\theta) \approx x(\tilde{\theta})$. These two steps, resampling and rejuvenation, ensure that the weighted sample \mathcal{E}_{n} adequately represents areas of posterior probability. Algorithm 1 summarizes the described steps of the standard particle filter. For smoothing over the past predictions the previously predicted paths could be resampled along with the states and parameters according to the current particle weights at each step.

There exist many extensions, modifications, and add-on techniques for particle filters. Depending on how the analysis ensemble is generated from the forecast ensemble, different filter algorithms are distinguished, see e.g., [115, 116]. Among these, the class of ensemble transform filters [116] is very promising as it replaces the stochastic resampling and rejuvenation step of the standard particle filter by a deterministic transformation solving an optimal transport problem, which allows to ensure certain properties, e.g., 2nd order accuracy [115]. However, in the augmented state space the connection between the parameters and states is lost as larger steps in the parameter space are undertaken, which means that the assumption $x(\theta) \approx x(\tilde{\theta})$ is no longer valid, requiring a post-processing step to fulfill the corresponding constraints (see e.g., [117]).

Algorithm 1 Standard particle filter	
Initialization of particles $\{(x_0^{(m)}, \theta_0^{(m)}, w_0^{(m)})\}_{m=1}^M$	
for $j = 1: n$ do	
• Propagation under model equations	\triangleright Eq. (2.18)
to generate the forecast ensemble $\{(x_j^{(m)f}, \theta_{j-1}^{(m)}, w_{j-1}^{(m)})\}_{m=1}^M$.	
• Update of importance weights $\tilde{w}_j^{(m)} = p(y_j h(x_j^{(m)f}), \Sigma) \cdot w_{j-1}^{(m)}$.	\triangleright Eq. (2.19)
• Normalization of importance weights: $w_j^{(m)} = \tilde{w}_j^{(m)} / \sum_{m=1}^M \tilde{w}_j^{(m)}$.	
• Calculate M_{eff} .	\triangleright Eq. (2.20)
if $M_{\text{eff}} < M/2$ then	
• Resampling of particles according to weights $\{w_j^{(m)}\}_{m=1}^M$	
and set $w_j^{(m)} = 1/M$ for $m = 1,, M$.	
• Rejuvenation of particles.	\triangleright Eq. (2.21)
end if	
end for	

The presented batch and sequential DA algorithms are used in Chapter 3 for Bayesian

forecasting, which also includes an interpretation within the pharmacometric setting. The Gibbs sampler is applied to the hierarchical Bayesian inference problem in Chapter 5.

2.2.3 DA in health care

Health care applications, or more specific Bayesian forecasting in a PK/PD context, confronts DA algorithms with different challenges compared to typical DA applications in terms of data availability and model specifications. In this section, the different challenges arising from the specific application in health care are described. Then specific application examples of DA in health care are outlined.

Typical DA applications deal with the estimation of the state of a time-varying system, e.g., in navigation and tracking [27]. Diffusion models based on stochastic differential equations (SDEs) are commonly used to describe uncertainties in the underlying models. The SDE formulation allows separating model misspecification from measurement error, which is typically summarized in PK/PD models in the RUV as defined in Eq. (2.4). SDE models have also been studied in the PK/PD context [118, 119, 120, 121] but have so far not been widely adopted. It has been found, that SDE-based models on the state level might not be well suited for describing PK/PD dynamics [87]. The dynamics are represented too irregularly and fluctuating, and biological laws and constraints are not well preserved, e.g., concentrations should not be negative, or drug concentrations should not increase, which might require post-processing steps [117].

Often, the model parameters are considered fixed or known, e.g., in numerical weather prediction [24]. In contrast, in PK/PD models the parameters are to be estimated from data. Rather than through stochastic fluctuations at the state level, differences between individuals but also within one individual (IIV and IOV) are expected on the level of the parameters, and deterministic state dynamics are considered. Therefore, it could be more reasonable to model a stochastic process on the level of the parameters [122], which better preserves biological constraints. The resampling and rejuvenation step in PF also allows a certain temporal parameter evolution depending on the size of the rejuvenation parameter. The time-varying nature of PK/PD parameters is already to some extent accounted for by IOV, see e.g., Eq. (2.5). The hierarchical structure of PK/PD models already adds one more level of variability, the IIV, which is generally not considered in typical DA applications.

In addition to the differences on the model level, also the data situation is very distinct. SDE models require high-frequency data, which are generally not available in health care [122], even with novel point-of-care devices (except in the ICU). The hierarchy of the NLME framework makes use of the structure of the data in health care; data might be sparse on the individual patient level, however, data might be available from many patients. In addition, health care data provide varying information content about the model parameters, i.e., not all data points are informative about all parameters, and intervention is required to learn parameters (except for steady-state parameters). Also, health care data pose further challenges comprising missing data, systematic errors, subjective scores based on physicians' assessments, unmeasured confounders, and mixtures of discrete and continuous variables [123]. The sparse data situation in combination with the nonlinearity of the underlying dynamics hinders the application of linearization approaches as frequently used in DA.

Despite the outlined differences, DA algorithms have been applied in health care; mainly in a rich clinical data context, e.g., forecasting glucose values [124, 125, 28] or leveraging ICU monitoring data [29]. The studies generally do not build on prior knowledge from previous population analyses, i.e., the prior parameter uncertainty does not reflect the observed variability in the patient population as estimated in NLME modeling. Novel digital health care devices as outlined in Section 2.1.1 demand more efficient real-time inference within the existing MIPD and Bayesian forecasting framework that also enables model-based predictions with quantified uncertainty for well-founded clinical decision-making.

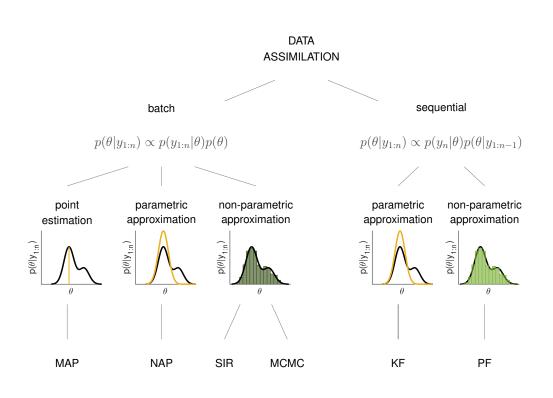


Figure 2.4: Overview of data assimilation (DA) approaches. We categorized the methods according to the way data are processed—all at once (in a batch) or one data point at a time (sequentially). The two different approaches rely on different formulations of Bayes' formula. Within these two approaches we distinguished between the different levels of approximation of the posterior distribution. Methods that are based on minimization of a cost functional (also called variational DA) determine a point estimate, the mode, of the posterior, i.e., maximum a-posteriori (MAP) estimation. Parametric approximations assume a parametric distribution, e.g., the normal distribution to approximate the shape of the posterior (normal approximation (NAP)). Non-parametric approximations, in contrast, do not assume any distributional form and generally rely on sample approximations to the full posterior (displayed as histograms), e.g., sampling importance resampling (SIR), or Markov chain Monte Carlo (MCMC) methods. Sequential DA approaches also comprise parametric and non-parametric methods. Parametric methods, e.g., the Kalman filter (KF), rely on the conjugate update of gaussian distributions. The proposed categorization is not universal and many hybrid methods, i.e., combinations between different algorithms exist.

2.3 Reinforcement learning (RL)

RL, a subfield of ML, comprises approaches for solving decision problems in which decisions have to be made in stages and the outcome is to some extent uncertain [126, 31, 32]. These kinds of problems appear in many application fields, e.g., games [34], robot motion control [127], or online advertisement [128]. A comprehensive overview of the history of RL is given in [32]. At the core of an RL setting is the sequential trial-and-error interaction between a goal-directed agent (or controller) with an uncertain environment. The agent's task is to learn how to act best with respect to optimizing a specific long-term goal [129, 32]. The search is guided by a feedback signal (the so called reward) that evaluates the consequences of the chosen actions. The field of RL is closely related to dynamic programming [130, 32], stochastic optimal control [33], AI [131], and bandit problems [132]. The RL problem of repeatedly making decisions in a stochastic environment can be formalized as an MDP.

Markov decision process (MDP). MDPs build the theoretical foundation for modeling sequential decision-making problems under uncertainty [133, 134]. An MDP comprises the tuple (S, A, R, p, γ) , with

- States S: a set of states. The current situation of the environment at a certain time point t is summarized in a state $s_t \in S$, e.g., the current location of a robot (in x and y coordinates) in a navigation task. The state should satisfy the Markov property, i.e., the present state represents the history relevant for future states. If the state is not fully observable, the MDP is called *partially observable* (POMDP), e.g., only noisy or partial measurements of the state are available.
- Actions A: A set of actions. The agent can interact with the environment by choosing an action $a_{t+1} \in A$ at decision time point t which will change (or control) the state of the environment, e.g., the robot can choose to go left, straight ahead, or right.
- Rewards R: The reward is the immediate (numerical) value of taking an action, i.e., it evaluates the change of state of the environment as a consequence of actions. It can be a function of the state, or the state and the action, $R: S \to \mathbb{R}$ or $R: S \times A \to \mathbb{R}$. The definition of the reward function requires domain knowledge and is crucial for achieving the desired goal. The reward might also be delayed, as e.g., in games like chess, where often all rewards are zero except for the terminal states (a win or a loss).
- *Transition probabilities p*: Due to the uncertain nature (or incomplete knowledge) of the environment, transitions between states as a consequence of an action are characterized by transition probabilities

$$p(s_{t+1}|s_t, a_{t+1}) = \mathbb{P}[S_{t+1} = s_{t+1}|S_t = s_t, A_{t+1} = a_{t+1}], \qquad (2.22)$$

i.e., the probability of being in state s_{t+1} given that action a_{t+1} was taken in state s_t . The transition probability depends only on the current state and not on the history due to the Markov property.

• Discount factor $\gamma \in [0, 1]$: The discount factor represents the weighting of future rewards. If $\gamma = 0$, the agent is only interested in collecting immediate rewards, future rewards are neglected. As $\gamma \to 1$, future rewards are increasingly taken into account. For $\gamma = 1$ the MDP is considered undiscounted; all future rewards are weighted equally. The behavior of the agent, i.e., choosing actions in given states of the environment, is modeled by the policy π ,

$$\pi(a|s) = \mathbb{P}[A_{t+1} = a|S_t = s].$$
(2.23)

If the agent follows a random policy, the agent will select each action with uniform probability, i.e., $\pi(a|s) = 1/|A|$, $\forall a \in A$. Solving an MDP corresponds to finding an optimal policy π^* , i.e., an optimal interaction of the agent with the environment. The goodness of a policy is evaluated based on the so called return G_t at time step t, defined as the weighted sum of rewards over the *remaining* time horizon

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{T-(t+1)} R_T = \sum_{k=t+1}^T \gamma^{k-(t+1)} R_k, \qquad (2.24)$$

in the case of a finite-time horizon RL problem, i.e., there are sequences of limited length T, called episodes. In contrast, infinite-time horizon RL problems are ongoing tasks without a clearly defined end (i.e., $T = \infty$). The objective of the agent is to maximize the expected long-term return

$$q_{\pi}(s,a) := \mathbb{E}_{\pi}[G_t | S_t = s, A_{t+1} = a], \qquad (2.25)$$

given the current state $S_t = s$ and action $A_{t+1} = a$ over the space of policies π . The function q_{π} is called the *action-value function* as it describes the value of taking action a in state s [32]. Similarly, the *value* of a state

$$v_{\pi}(s) := \mathbb{E}_{\pi}[G_t | S_t = s],$$
 (2.26)

is the expected return starting from state s, following policy π , and averaging over all possible actions.

The value functions Eq. (2.25) and Eq. (2.26) take a central role in RL as they represent an important intermediate step towards improving policies in a state-space search rather than searching the space of plans/actions. Learning an optimal policy (i.e., solving an MDP) involves maximizing the expected long-term return q_{π} , which in turn depends on the current estimate of π . Therefore, RL approaches are typically based on an iterative process of value estimation (\hat{q}_{π}) and policy improvement [32]. Dynamic programming can be used to solve MDPs for small problems (i.e., small state and action spaces) in which the transition probabilities are known. However, for most real-world problems dynamic programming is not feasible [135]. RL can be seen as approximate dynamic programming to approach intractable real-world problems. To reach this goal, RL techniques generally use sampled data of the transition functions to represent the underlying dynamics [31, 135]. We will focus, in the following, on solving MDPs via model-based RL and consider finite sets of states and actions.

Model-based RL. Most RL algorithms are data-driven approaches that do not require a mathematical/computational model of the data generating/underlying process but solely rely on experience (model-free RL). Model-based RL methods can, however, be used in applications for which models exist that describe a system and/or if not enough data are available for a purely data-driven approach. In model-based RL (also called planning), the model is used to generate data (simulated experience), which is subsequently used within model-free RL methods instead of 'real' experience.

Model-based RL methods that rely on sampling (sample-based planning) estimate the expected value in Eq. (2.25) via a sample approximation, where each sample $k = 1, \ldots, K$ corresponds to a model-generated trajectory (episode)

$$s_0^{(k)} \xrightarrow{a_1^{(k)}} (s_1^{(k)}, r_1^{(k)}) \xrightarrow{a_2^{(k)}} (s_2^{(k)}, r_2^{(k)}) \xrightarrow{a_3^{(k)}} \dots \xrightarrow{a_T^{(k)}} (s_T^{(k)}, r_T^{(k)})$$

using a policy π_k . For each sample k the current sample approximation of Eq. (2.25)

$$q_k(s,a) = \frac{1}{N_k(s,a)} \sum_{k'=1}^k \sum_{t=1}^T \mathbf{1}_{(s_t^{(k')} = s, a_{t+1}^{(k')} = a)} G_t^{(k')}$$
(2.27)

is computed. Here, $N_k(s, a)$ denotes the number of times that action a was chosen in state s amongst the first k episodes, and $G_t^{(k)} = r_{t+1}^{(k)} + \gamma r_{t+2}^{(k)} + \ldots$ is the discounted sum of sampled rewards. Ideally, $N_k(s, a)$ should be large for each state-action combination to guarantee a good approximation of the expected return (law of large numbers). This, however, is infeasible for most applications (curse of dimensionality). A major aspect of RL is therefore the trade-off between exploration and exploitation; the agent should take actions that—at the current stage of approximation—give a high return (exploitation), but should also try new actions that potentially may lead to an even higher return (exploration), i.e., to avoid local maxima. Different algorithms were developed to provide a good approximation of Eq. (2.25).

In this work, only tabular representations of the action-value function are considered in the following algorithms assuming discrete states and actions. However, for continuous state or action representations, or large state-action spaces, function approximations should be used, e.g., a linear approximation, $q_k(s, a) \approx \hat{q}(s, a, \psi) = \psi^T \phi(s, a)$ with weights ψ and features ϕ , or neural networks for a nonlinear function approximation, see e.g. [136].

2.3.1 Q-learning

Q-learning [137, 138] is a model-free RL method based on the decomposition of the action-value function, Eq. (2.25), into the immediate reward and the discounted action-value of succeeding state and action

$$q_{\pi}(s, a) = \mathbb{E}_{\pi}[G_t | S_t = s, A_t = a]$$

= $\mathbb{E}_{\pi}[R_{t+1} + \gamma G_{t+1} | S_t = s, A_t = a]$
= $\mathbb{E}_{\pi}[R_{t+1} + \gamma q_{\pi}(S_{t+1}, A_{t+1}) | S_t = s, A_t = a].$ (2.28)

In contrast to Monte Carlo learning (as described above), the target of so-called temporal difference (TD) approaches [139] is not the total return G_t but $R_{t+1} + \gamma q_{\pi}(S_{t+1}, A_{t+1})$ or in terms of value functions $R_{t+1} + \gamma v_{\pi}(S_{t+1})$. Equation (2.28) is generally called Bellman expectation equation (for the state-action function) [130]. This decomposition is especially useful for long sequences (T large) or infinite time horizons $(T = \infty)$. In Q-learning the sample approximation of Eq. (2.25) is computed via

$$q_{k+1}(s,a) = q_k(s,a) + \alpha_k \cdot \left(R_{t+1} + \gamma \cdot \max_{a'} \left[q_k(s',a') \right] - q_k(s,a) \right),$$

where s' is the next state when taking action a in state s at decision time t. A crucial parameter to ensure convergence to the optimal action-value function (i.e., $q_k(s, a) \rightarrow q^*(s, a)$ as $k \rightarrow \infty$ [138]) is the learning rate α , which needs to decay appropriately with the number of iterations, e.g., as a Robbins-Monro sequence [140].

In Q-learning, frequently the ϵ -greedy approach is used to address the aforementioned exploration-exploitation trade-off: With probability ϵ , a random action is chosen, and with probability $1 - \epsilon$, the greedy action, i.e., the current argmax of q_k (policy improvement). Also, ϵ can be chosen in a decreasing manner, to encourage exploration in the beginning and exploitation at later training episodes.

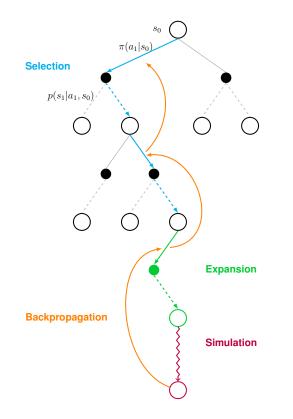


Figure 2.5: Illustration of Monte Carlo tree search (MCTS). 1.) Selection: Within the search tree, actions are selected according to the tree policy, until a not yet fully expanded node is reached. 2.) Expansion: The action is chosen among the unvisited actions according to a roll-out policy (e.g., random) and the new node is added to the search tree. 3.) Simulation: From the new child node, a simulation is started using the roll-out policy until the end of the episode/therapy. 4.) Back-propagation: The return is back-propagated through the tree, i.e., action-values functions and number of visits are updated for each selected state-action pair within the search tree.

2.3.2 Monte Carlo tree search (MCTS)

Monte Carlo tree search (MCTS) combines the Monte Carlo method with an incremental asymmetric tree search [141]. It was mainly developed and applied to game tree search, e.g., AlphaGo [142]. In contrast to Q-learning, MCTS requires a model or simulator to simulate or sample from the underlying process (simulation-based search). MCTS comprises four recursive steps which are repeated in each episode, k = 1, ..., K, for building a search tree based on random samples in the decision space which explores promising directions more exhaustively [143], see Figure 2.5:

- Selection: Starting at the root node s_0 actions are selected according to the tree policy (π_k) until a not yet fully expanded node, i.e., a node with an unvisited action, is reached.
- *Expansion:* If the selected node is expandable (nonterminal state), one child node is added by selecting an unvisited action.
- *Simulation:* Following a default policy (often random) a single simulation is run until a terminal state is reached. The return of this episode is calculated.

• Backpropagation: The simulated return is backpropagated to the selected nodes in the search tree. More specifically, back up means computing incrementally the expected return via a running mean, i.e., in episode k

$$q_k(s_t, a) = q_{k-1}(s_t, a) + \frac{1}{N_k(s_t, a)} \left(G_t^{(k)} - q_{k-1}(s_t, a) \right).$$

Note that just a basic version of MCTS is presented here and various modifications are possible, e.g., adding several child nodes in the expansion step or running multiple simulations in parallel. MCTS can also be combined with an efficient exploration and exploitation strategy—called the upper confidence bound applied to trees (UCT) [144, 132], which was originally developed in the context of bandit problems.

The upper confidence bound applied to trees. The simplest version of an RL problem is the setting of a multi-armed bandit, i.e., a one-step decision problem [135]. The agent can pull the arms of multiple 'one-armed bandit' (gambling) machines and observes the resulting (random) reward. The arms correspond to the actions that the agent in an RL problem can choose at one time instance (MDP with a single state). Each machine produces rewards with unknown expectation. As a measure of success, the regret is typically considered in bandit literature, i.e., the loss associated with not following the optimal policy from the beginning (not pulling the arm of the machine with the highest expectation).

The bandit algorithm underlying UCT assigns an upper confidence bound (UCB) of the average reward for each machine based on Hoeffding's inequality [145]. The principle of *optimism in the face of uncertainty* is implemented in the algorithm; instead of choosing the arm (action) with the highest average return (greedy), one chooses the one with the highest upper bound (UCB). Thus, integrating additionally how certain we are about the estimate of the return. We consider the setting of Hoeffding's second Theorem [145]:

Let X_1, \ldots, X_n be independent, bounded random variables (the rewards generated by repeatedly pulling the arm of one machine), i.e., $lb_i \leq X_i \leq ub_i$ with sample mean $\bar{X} = (X_1 + \ldots + X_n)/n$, then

$$P[E[X] > \bar{X} + U] \le \exp\left(\frac{-2n^2 U^2}{\sum_{i=1}^n (\mathbf{u}\mathbf{b}_i - \mathbf{l}\mathbf{b}_i)^2}\right),$$
(2.29)

where U represents the size of the one-sided confidence interval (upper confidence bound). Translating Eq. (2.29) into the RL setting for X = G, i.e., each tree node is considered as a distinct multi-armed bandit, see [34, 132]

$$P[q_{\pi}(s,a) > q_k(s,a) + U_k(s,a)] \le \exp\left(-\frac{2N_k(s,a)^2 U_k(s,a)^2}{\sum_i (\mathbf{ub}_i - \mathbf{lb}_i)^2}\right)$$

Choosing α to represent the probability that the true expected action-value exceeds the upper bound $(q_k(s, a) + U_k(s, a))$

$$\exp\left(-\frac{2N_k(s,a)^2U_k(s,a)^2}{\sum_i(\mathrm{ub}_i-\mathrm{lb}_i)^2}\right) = \alpha\,,$$

leads to

$$U_k(s,a) = \sqrt{(\mathrm{ub} - \mathrm{lb})^2} \cdot \sqrt{\frac{-\log \alpha}{2N_k(s,a)}},$$

with ub = ub_i and lb = lb_i for all $i = 1, ..., N_k(s, a)$, i.e., the upper and lower bounds are equal for all *i*. For the bandit problem, it was shown that $\alpha = N_k(s)^{-4}$ ensures logarithmic regret [144], where $N_k(s) := \sum_{a'} N_k(s, a')$ are the number of visits of state *s*. This leads to the UCB 1 algorithm [144] for the case [lb, ub] = [0, 1] with

$$U_k(s,a) = \sqrt{\frac{2\log N_k(s)}{N_k(s,a)}}.$$
(2.30)

In [34] a variant of the algorithm

$$U_k(s,a) = \frac{\sqrt{N_k(s)}}{1 + N_k(s,a)}$$
(2.31)

was successfully used for game tree search (AlphaGo), which corresponds to

$$\alpha = \exp\left(-2N_k(s,a)^2 \cdot \frac{N_k(s)}{(1+N_k(s,a))^2}\right) \approx \exp\left(-2N_k(s)\right)$$

for large $N_k(s, a)$. Note, the UCB 1 Algorithm is actually initialized by taking each action once; this is, however, not possible in the tree setting due to the higher dimensionality, therefore, the denominator of Eq. (2.31) was modified to avoid division by zero. The choice in Eq. (2.31) encourages exploration as the numerator is larger than in Eq. (2.30). In a purely model-based learning setting, exploration can be encouraged to a greater extent since the measure of success is not the cumulative regret, but the goal is to search the tree efficiently and exhaustively (during training). In contrast, to the often-used ϵ -greedy policy, which chooses with probability ϵ a random action, UCT avoids choosing repeatedly actions with low return ($q_k(s, a)$ and $U_k(s, a)$ small) and also chooses actions more frequently that are close to the maximum but have not been taken often ($q_k(s, a)$ and $U_k(s, a)$ large).

We will employ MCTS in conjunction with UCT in Chapter 4 to determine RL-based dosing policies for neutrophil-guided dosing in individualized chemotherapy.

2.3.3 RL in health care

Computing optimal drug administrations can be regarded as a sequential decision-making problem under uncertainty in which dose decisions have to be made in stages and the outcome in a patient for a given dose choice is associated with uncertainty [31]. The agent can be seen as a virtual physician (also called AI clinician [146]) whose task is to learn a policy (or strategy) of how to dose (act) best with respect to optimizing a specific expected long-term response (return) of a virtual patient (an uncertain and delayed feedback environment)[32, 129, 147, 148]. Traditionally, these problems have been addressed with optimal control or model predictive control, e.g. the artificial pancreas [149] The high potential of RL in health care has been highlighted in recent review articles [129, 148], especially for delayed feedback environments and various applications in different areas of health care exist, see [37] for a comprehensive survey.

One major application area is the design of clinical trials, e.g., when to initiate the second treatment line and which chemotherapeutic agent to choose [36] but also precision dosing has been investigated (so called dynamic treatment regimens [37]), e.g., for anemia treatment in hemodialysis patients [31] or management of sepsis in the ICU [146]. Unlike typical RL applications, such as robotics or games, model-free RL in health care needs to learn dosing policies from potentially sparse historical clinical data, i.e., the RL agent cannot directly

interact with the environment to learn a policy. This restricts the action space to actions taken by clinicians previously, which could lead to suboptimal policies (as exploration is not possible). PK/PD models, however, allow extrapolating, i.e., simulating the response of patients to actions that have not yet been taken. Hence, the RL agent can interact with the model, which enables exploration of the action space to find even better policies than already existent in clinical practice, e.g., in model-based RL a tumor growth model was used to simulate the change in mean tumor diameter [38] or an erythropoiesis model to simulate hemoglobin levels [31].

A decisive quantity for deriving dosing policies using RL is the reward function. Ideally, it is related to the overall therapy outcome, e.g., survival [146], however, often surrogate markers of the treatment response are better accessible, e.g., the tumor diameter [38] or hemoglobin concentration [31]. Improperly defined relationships between surrogate markers to the overall outcome may have an unintended impact on the resulting policy, e.g., important additional aspects such as side effects may be neglected. Weighting the various markers of efficacy and toxicity relative to each other is a difficult task that requires an expert discussion and preferences may vary from clinician to clinician [37]. Evaluation of the resulting policies is, therefore, critical and has to be performed with great care before deployment [150, 151]. In this regard, off-policy evaluation could be beneficial, which is an approach for estimating the value of a policy (RL-learned policy), without actually applying it, based on data collected by another policy (e.g., historical data using the current practice by physicians) [146]. Another key aspect for applicability is to examine the interpretability and transparency of RL policies in order to build confidence in the dose recommendations so that they are accepted by health care professionals and eventually integrated into everyday healthcare use.

3

Bayesian therapy forecasting and clinical decision support

A crucial element of MIPD and clinical decision support is model-driven forecasting of the individual therapy outcome integrating patient-specific TDM data and prior knowledge on the patient-disease-drug system from previous clinical studies [15]. In clinical practice, it is important that patient data can be processed efficiently, but at the same time, comprehensive and reliable information about the therapy outcome and potential risks is required for thorough clinical decision-making.

The challenge in therapy forecasting for MIPD is to infer information on the individual model parameter values θ_i of a patient based on his/her covariate values and TDM measurements. As introduced in Chapter 1, a Bayesian approach is very advantageous for this task. Bayesian forecasting in the context of MIPD builds on prior knowledge in form of a structural (Eq. (2.1)), observational (Eq. (2.2)), covariate (Eq. (2.3)) and statistical model (Eq. (2.4)). The functional relationships of the models as well as the parameter estimates are the result of a population analysis of clinical study data as described in Section 2.1.3. The unexplained IIV and IOV in the population model (Eq. (2.3)) defines the prior uncertainty about the individual parameter values. Then, assimilating measurements $y_{1:n} = (y_1, \ldots, y_n)^T$ into the model based on Bayes' formula Eq. (2.7) allows learning about individual parameter values from the data. The remaining uncertainty of parameter values is encoded in the posterior $p(\cdot|y_{1:n})$.

The currently most common approach for Bayesian forecasting is MAP estimation, as this approach transforms the posterior inference problem into an optimization problem that can be solved efficiently (Section 2.2.1). In MAP estimation, only the most probable individual parameter values, i.e., the MAP estimates, are used to predict the individual therapy outcome without quantifying associated uncertainties [20]. Thus, relevant risks associated with a dosing regimen selection, e.g., treatment inefficacy or unacceptable toxicity, are not determined hindering well-founded therapeutic decision-making. Quantifying associated uncertainties is at the heart of making more informed decisions. In this chapter, which is based on [CM1], we thoroughly compare in a TDM context different Bayesian DA methods, as introduced in Section 2.2, that either (i) estimate only the posterior mode (MAP estimation) without uncertainty quantification, or (ii) estimate the full posterior distribution (termed full Bayesian approaches) to quantify uncertainties. The DA approaches comprise not only methods that process patient data collected over time in a batch (i.e., all at once), like MCMC, SIR, and NAP (see Section 2.2.1), but also PFs that allow for efficient sequential data processing (see Section 2.2.2). In the context of chemotherapy-induced neutropenia, the clear benefits of uncertainty quantification compared to purely MAP-based predictions (as, e.g., in [73]) are presented using the gold-standard model for docetaxel-induced neutropenia [91, 67], see Section A.2 for model details. Further, we compare the full Bayesian approaches regarding the quality of uncertainty quantification and computational runtime for multiple cycle chemotherapy using the BME model described in Section 2.1.3 [90]. Finally, the efficient data processing of the sequential DA approach is discussed in a frequent monitoring context, where novel health care devices (e.g., home-monitoring devices) allow patients to measure and report individual biomarker concentrations online.

In this chapter, TDM data are considered for a single individual and therefore there is no running index for individuals (i.e., we drop the *i*). The hyperparameters or population parameters, i.e., all parameters after the semicolon in Eq. (2.3), are assumed to be known (fixed) in the Bayesian forecasting framework. As a consequence, we drop them as well as the subscripts in the notation in this chapter. As a result, the likelihood at the individual level reads $p(\cdot|\theta) = p(\cdot|\theta; h_j(\theta), \Sigma)$ and the prior $p(\cdot) = p_{\Theta}(\cdot; \theta^{\text{TV}}(\text{cov}), \Omega)$. In the case of IOV (Eq. (2.5)), the parameter vector contains parameter values, which are constant across occasions (cycles) θ^{IIV} and parameters that are specific for each cycle c, θ_c^{IOV} . For a lognormal distributional assumption for the parameters, this is given by

$$\theta_c = \underbrace{e^{\log(\theta^{\mathrm{TV}}) + \eta}}_{=\theta^{\mathrm{IIV}}} \cdot \underbrace{e^{\kappa_c}}_{=\theta^{\mathrm{IOV}}_c} . \tag{3.1}$$

The θ^{IIV} parameters of a patient need to be learned across all treatment cycles and the cycle specific parameters θ_c^{IOV} based on the data observed in cycle c.

3.1 Maximum a-posteriori (MAP) estimation: the current state-of-the-art

MAP estimation infers the mode of the posterior distribution, i.e., the most probable individual parameter values given patient-specific measurements $y_{1:n}$, Eq. (2.10). It is, however, more convenient and numerically more stable to minimize the negative log-posterior instead

$$\hat{\theta}_n^{\text{MAP}} = \underset{\theta}{\operatorname{arg\,min}} - \log p(\theta|y_{1:n}) = \underset{\theta}{\operatorname{arg\,min}} - \log p(y_{1:n}|\theta) - \log p(\theta) \,. \tag{3.2}$$

Choosing an additive normal residual error model $(Y_j = h_j(\theta) + \epsilon_j \text{ with } \epsilon_j \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2))$ and a lognormal IIV model for the parameters $(\theta_k = \theta_k^{TV} \cdot e^{\eta_k} \text{ with } \eta_k \stackrel{iid}{\sim} \mathcal{N}(0, w_k^2))$ as in the models considered in this thesis (Section 2.1.3) yields

$$\hat{\theta}_{n}^{\text{MAP}} = \arg\min_{\theta} \frac{n}{2} \log(2\pi) + \frac{n}{2} \log\sigma^{2} + \frac{1}{2} \sum_{j=1}^{n} \frac{(y_{j} - h_{j}(\theta))^{2}}{\sigma^{2}} + \frac{n_{\theta}}{2} \log(2\pi) + \frac{1}{2} \sum_{k=1}^{n_{\theta}} \log\omega_{k}^{2} + \sum_{k=1}^{n_{\theta}} \log\theta_{k} + \frac{1}{2} \sum_{k=1}^{n_{\theta}} \frac{(\log(\theta_{k}) - \log(\theta_{k}^{\text{TV}}))^{2}}{\omega_{k}^{2}},$$
(3.3)

with data $y_{1:n} = (y_1, \ldots, y_n)^T$ observed up to time point t_n . The MAP estimate $\hat{\theta}_n^{\text{MAP}}$ is a one-point summary of the posterior distribution, i.e., the distribution is approximated with a delta peak, without quantification of the associated uncertainty. The extension for estimating

also IOV parameters is given in Section B.2.2. The optimization problem Eq. (3.2) can be solved efficiently using gradient descent approaches. Since these are, generally, only local optimizers they should be combined with a sensitivity analysis or multi-start search, see e.g., [152, 153]. MAP estimation was performed in Matlab R2017b using the interior-point algorithm in fmincon. The Matlab toolbox AMICI [152] was used for simulation of the system of ODEs and for computations of sensitivities used in gradient calculations, see Section B.1 for details. The individual prediction of the therapy outcome is typically generated by solving the structural Eq. (2.1) and observational model Eq. (2.2) for the MAP estimate $\hat{\theta}_n^{MAP}$ (in the sequel called MAP-based predictions).

3.2 Normal approximation (NAP)

To overcome the one-point summary limitation of the MAP estimate, the posterior $p(\cdot|y_{1:n})$ may be approximated locally by a normal distribution located at the MAP estimate, see Eq. (2.11). As discussed in Section 2.2.1, this approximation can be very precise in the case of highly informative data sets, however, this is often not the case in Bayesian forecasting as not all data points are informative about all parameters.

The uncertainty of the model parameters can be propagated to the observables using either a simulation-based approach or the delta method. In the simulation-based approach, a sample \mathcal{E}_n is generated from the normal distribution in Eq. (2.11). Based on this posterior sample, we may approximate quantities of interest in the observable space by solving Eq. (2.1)+(2.2) for all elements in \mathcal{E}_n [154]. This serves as the basis for credible intervals (CrIs); applying subsequently the residual error model (Eq. (2.4)) is the basis for prediction intervals (PIs).

Alternatively, the delta method [155, section 5.5] could be used to determine the limiting distribution of a differentiable function of the parameters $T(\theta)$ [155, section 5.5]. For forecasting in the observable space: $T(\theta) = h(x(t), \theta) =: h_t(\theta)$. Using the delta method, the uncertainties are propagated from the parameters to the observable via the output sensitivities $S_t^h(\hat{\theta}_n^{\text{MAP}}) = \nabla_{\theta} h_t(\theta) |_{\hat{\theta}^{\text{MAP}}}$ [155, section 5.5]:

$$p(h_t(\hat{\theta}_n^{\text{MAP}})|y_{1:n}) \approx \mathcal{N}\Big(h_t(\hat{\theta}_n^{\text{MAP}}), S_t^h(\hat{\theta}_n^{\text{MAP}})^T \Sigma^{\text{MAP}} S_t^h(\hat{\theta}_n^{\text{MAP}})\Big)$$

with $\Sigma^{\text{MAP}} = [\mathcal{I}(\hat{\theta}_n^{\text{MAP}})]^{-1}$. The CrI in observable space is then given by

$$\operatorname{CrI}^{\alpha} = h_t(\hat{\theta}_n^{\mathrm{MAP}}) \pm z_{1-\alpha/2} \sqrt{\sigma_{\mathrm{CrI}}^2(t)}, \qquad (3.4)$$

where $z_{1-\alpha/2}$ is the quantile of the standard normal distribution and $\sigma_{CrI}^2(t)$ is computed using the output sensitivities

$$\sigma_{\rm CrI}^2(t) \approx S_t^h(\hat{\theta}_n^{\rm MAP})^T \Sigma^{\rm MAP} S_t^h(\hat{\theta}_n^{\rm MAP}) \,,$$

see also Section B.1 for more details regarding sensitivity computations. For a more conservative choice, the quantiles of the Student's t distribution could be used in Eq. (3.4)[154].

For determining the PI, the residual variability is additionally taken into account

$$\sigma_{\rm PI}^2(t) \approx S_t^h(\hat{\theta}_n^{\rm MAP})^T \Sigma^{\rm MAP} S_t^h(\hat{\theta}_n^{\rm MAP}) + \sigma^2 \,,$$

with σ^2 denoting the univariate RUV Σ . Consequently, the prediction interval is given by

$$\mathrm{PI}^{\alpha} = h_t(\hat{\theta}_n^{\mathrm{MAP}}) \pm z_{1-\alpha/2} \sqrt{\sigma_{\mathrm{PI}}^2(t)} \,.$$

Since the delta method involves differentiation of T, it is not straightforward to apply the method to any desired quantity of interest, e.g., $T(\theta) = c_{\text{nadir}}(\theta) = \min h_t(\theta)$ (Figure 2.1).

3.3 Bayesian DA to quantify the uncertainty

Quantifying associated uncertainties is key to making well-founded decisions, particularly in clinical applications for which dosing decisions are associated with potentially serious risks for the patient. In the following, the DA algorithms introduced in Section 2.2 for full posterior inference are described in the specific context of Bayesian forecasting of the therapeutic outcome combining prior knowledge from NLME analysis with patient-specific TDM data.

3.3.1 Sampling Importance Resampling (SIR)

The SIR algorithm is a fully Bayesian approach that processes data in a batch based on importance sampling (Section 2.2.1). We considered for Bayesian forecasting the prior as importance distribution, $\pi_I(\theta) := p(\theta) = p(\theta|\theta^{\text{TV}}(\text{cov}), \Omega)$ assuming that the patient under consideration is sufficiently well represented by the clinical patient population given by the prior. In this case, the importance weight in Eq. (2.15) is given by the likelihood

$$\tilde{w}_{n}^{(m)} = \frac{\tilde{p}(\tilde{\theta}_{n}^{(m)}|y_{1:n})}{p(\tilde{\theta}_{n}^{(m)})} = p(y_{1:n}|\tilde{\theta}_{n}^{(m)}).$$

Note that once a new TDM data point y_{n+1} becomes available, the SIR algorithm does not simply update the present sample points $\theta_n^{(m)}$ in \mathcal{E}_n , but re-performs all three steps based on the updated posterior $p(\cdot | y_{1:n+1})$ to determine \mathcal{E}_{n+1} .

3.3.2 Markov Chain Monte Carlo (MCMC)

A popular alternative to SIR in Bayesian inference are MCMC methods with a wide range of different algorithms (Section 2.2.1). In Bayesian forecasting, MCMC was previously considered with the prior as fixed proposal distribution (independence sampler) for sparse patient monitoring data [156]. To counteract large rejection rates and ensure efficient sampling as the posterior is becoming narrower with more data, we used an adaptive M-H sampler with lognormally distributed proposal distribution located at the current position in the Markov chain

$$\pi_P(\cdot|\theta^{(m-1)}) := \mathcal{LN}(\cdot|\theta^{(m-1)}; \operatorname{Cov}[\hat{p}(\log(\theta)|y_{1:n-1}]).$$

The variance is computed from the previously estimated posterior based on TDM data available up to time t_{n-1} . Since we employed the M-H algorithm with a lognormal, hence asymmetric, distribution as a proposal, a correction term is required in the acceptance probability to account for the asymmetry. Using the probability density function of the lognormal distribution, the acceptance ratio Eq. (2.16) in the case of a lognormal proposal distribution is given by

$$\alpha = \min\left[1, \frac{p(\theta^*|y_{1:n})}{p(\theta^{(m-1)}|y_{1:n})} \cdot \frac{\mathcal{LN}(\theta^{(m-1)}|\theta^*, \Omega)}{\mathcal{LN}(\theta^*|\theta^{(m-1)}, \Omega)} = \frac{p(\theta^*|y_{1:n})}{p(\theta^{(m-1)}|y_{1:n})} \cdot \frac{\prod_k (\theta^*)_k}{\prod_k (\theta^{(m-1)})_k}\right]$$

where $k = 1, ..., n_{\theta}$. For the first data point y_1 the Markov chain is initialized at the typical value $\theta_0 = \theta^{\text{TV}}(\text{cov})$ and for posterior inference at time point t_n at the posterior expectation based on data $y_{1:n-1}$: $\theta_0 = \mathbb{E}[\hat{p}(\log(\theta)|y_{1:n-1})].$

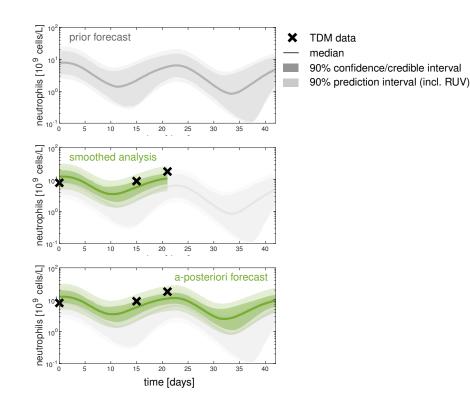


Figure 3.1: Step-by-step description of the particle filter. The different steps of the particle filter are depicted in the context of forecasting the time course of neutropenia. In grey the prior forecast is depicted with median (solid line), 90% confidence interval (CI) (shaded area) and 90% prediction interval (PI) (lighter shaded area). The updated a-posteriori (smoothed) prediction for the first cycle (21 days) and the a-posteriori forecast for the next cycle based on three observations (black crosses) is depicted in green (solid line: median, shaded area: 90% CrI, lighter shaded area: 90% PI).

3.3.3 Particle filters (PFs)

In contrast to SIR and MCMC, which process data in a batch, PF constitutes a sequential approach to DA (Section 2.2.2), see also [22, 113, 114] for a detailed introduction. As in SIR and MCMC, evaluation of the likelihood involves solving the structural model in Eq. (2.1). Since the structural model is deterministic, one may either solve Eq. (2.1) with initial condition $x_0(\theta)$ for the total timespan $[t_0, t_{n+1}]$ (as in SIR and MCMC) or with initial condition $x_n(\theta)$ for the incremental timespan $[t_n, t_{n+1}]$. The latter approach requires to store for each sample point $(\theta_n^{(m)}, w_n^{(m)})$ also the corresponding state $x_n^{(m)}$ at time t_n , since typically the structural model cannot be solved analytically. The incremental approach makes use of the Markov property that the future state is independent of the past when the present state is known.

The resulting triple $(\theta_n^{(m)}, x_n^{(m)}, w_n^{(m)})$ is called a particle. In our setting, the ensemble of particles can be interpreted as the state of a population of virtual individuals at time t_n , whose 'diversity' represents the uncertainty about the state/parameters of the patient at time t_n , given the individual TDM measurements $y_{1:n}$. The posterior $p(\cdot|y_{1:n})$ obtained by nsequential update steps in Eq. (2.17) is mathematically identical to the posterior obtained in Eq. (2.7) by assimilating all data $y_{1:n}$ in a batch [103, section 3.3.3]. However, the sequential update is much more efficient as it involves a reduced integration time span. In addition, in the case of IOV only the IOV parameters of the current occasion (cycle) need to be taken into account, thus, reducing the dimensionality of the inference problem. The rejuvenation was chosen proportional to the absolute value of the parameter in contrast to the general form Eq. (2.21), i.e., $\xi_n^{(m)} \sim_{\text{iid}} \mathcal{N}(0, \tau \cdot |\tilde{\theta}_n^{(m)}|)$. The IOV parameters are re-initialized at each occasion, which has also a rejuvenation effect.

The steps of the particle filter (Algorithm 1) are illustrated in Figure 3.1 including the prior ensemble forecast, the smoothed analysis of the TDM data, and the a-posteriori forecast in the context of forecasting the neutropenia time course integrating neutrophil measurements.

3.4 Workflow in Bayesian forecasting

In full Bayesian forecasting, uncertainty is quantified on the parameter level and subsequently propagated to the observable level, possibly further summarized for some key quantities of interest, see Figure 3.2. Prior to observing patient-specific data, the parameter uncertainty is characterized by the prior (Eq. (2.3)). Sampling from the prior allows to make a-priori predictions of the neutropenia time course and its uncertainty in form of a $(1 - \alpha)$ -confidence interval (CI). Also, a-priori predictions for quantities of interest can be derived, e.g., the neutropenia grade (Figure 3.2, left column). These predictions (including uncertainties) correspond to our expectations prior to observing any individual patient data of the patient to be treated. Once patient-specific data are assimilated into the Bayesian model, the remaining uncertainty about the parameter values is characterized by the posterior. A posterior sample \mathcal{E}_n allows to update also the uncertainty in the observable space (CrIs) and the quantities of interest (Figure 3.2, middle column).

Forward uncertainty propagation corresponds to transforming a probability distribution (prior or posterior) under a typically nonlinear mapping $T(\cdot)$, resulting in a transformed quantity $\psi = T(\theta)$. For illustration, we assume the one-dimensional case with strictly increasing T and $\theta = T^{-1}(\psi)$. Then, the posterior in terms of ψ is given by [157, section 1]

$$p_{\Psi}(\psi|y_{1:n}) = p_{\Theta}(\theta|y_{1:n}) \cdot \frac{dT^{-1}(\psi)}{d\psi}, \qquad (3.5)$$

which is approximated in sampling-based approaches (cf. Eq. (2.13)) by

$$\hat{p}_{\Psi}(\psi|y_{1:n}) = \sum_{m=1}^{M} w_n^{(m)} \mathbf{1}_{\{\psi^{(m)}=\psi\}},$$

with $\psi^{(m)} = T(\theta^{(m)})$. This allows the computation of any desired summary statistic, e.g., posterior expectation or quantiles. MAP estimation, in contrast, characterizes the posterior by a single value and allows only to make a single MAP-based prediction by mapping the MAP estimate $\hat{\theta}_n^{\text{MAP}}$ to the quantity of interest $T(\hat{\theta}_n^{\text{MAP}})$, lacking crucial information on its uncertainty (Figure 3.2, right column). Importantly, for nonlinear T this does *not* result in the most probable outcome, due to the Jacobian factor $\frac{d\theta}{d\psi} = \frac{dT^{-1}(\psi)}{d\psi}$ in Eq. (3.5) [158, 87]: The most probable outcome is defined as the outcome with maximum posterior probability

$$\hat{\psi}_n^{\text{MAP}} = \operatorname*{arg\,max}_{\psi} p_{\Psi}(\psi|y_{1:n}),$$

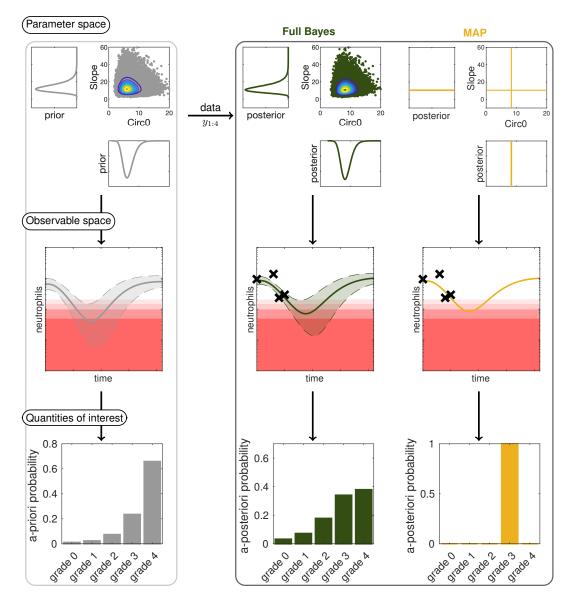


Figure 3.2: Overview of the workflow in Bayesian forecasting for clinical decision support comparing full Bayesian inference to maximum a-posteriori (MAP)-based prediction. In full Bayesian inference uncertainties in the parameter values are propagated to uncertainties in the observable space and quantities of interest. The posterior $p(\theta|y_{1:n})$ is displayed for the parameters 'Slope' (drug effect parameter) and 'Circ₀' (pretreatment neutrophil concentration). For the prior and full Bayes (reference) approach (sampling importance resampling (SIR) with $M = 10^6$) samples (dots) from the distributions are shown with contour levels. In the observable space the point estimates (solid lines) are displayed with the central 90% CrIs (dashed lines and shaded area) along with the therapeutic drug/biomarker monitoring (TDM) data (crosses). The a-priori/a-posteriori probabilities are calculated for the neutropenia grades (grade 0 to 4). Note that y_1 corresponds to the measurement of baseline neutrophil counts ('Circ₀') and is taken into account in the posterior.

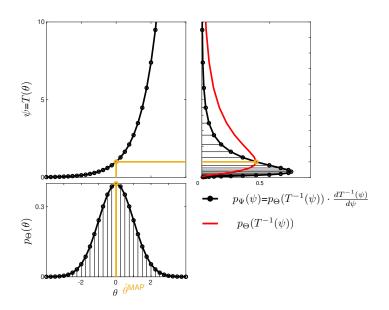


Figure 3.3: Illustration of a nonlinear transformation of a probability density. For illustration a normal distribution $p_{\Theta}(\theta) = \mathcal{N}(\cdot|0,1)$ is transformed with $\psi = T(\theta) = \exp(\theta)$, which results in the lognormal distribution $p_{\Psi}(\psi) = \mathcal{LN}(\cdot|0,1)$ (black line with dots). The red line shows the transformation of $p_{\Theta}(\theta)$ as a function, $p_{\Theta}(T^{-1}(\psi))$ not integrating to one. The transformed mode of the normal distribution, $\exp(\mu) = \exp(\hat{\theta}^{\text{MAP}})$, is not the mode of the lognormal distribution but the median. The illustration is based on [159, Figure 1].

which satisfies (assuming for illustration that T is strictly increasing)

$$0 = \frac{d}{d\psi} p_{\Psi}(\psi|y_{1:n}) \stackrel{\text{Eq. (3.5)}}{=} \frac{d}{d\psi} \left[p_{\Theta}(T^{-1}(\psi)|y_{1:n}) \cdot \frac{dT^{-1}(\psi)}{d\psi} \right]$$
$$= \frac{d}{d\theta} p_{\Theta}(T^{-1}(\psi)|y_{1:n}) \cdot \left(\frac{dT^{-1}(\psi)}{d\psi}\right)^2$$
(3.6a)

$$+ p_{\Theta}(T^{-1}(\psi)|y_{1:n}) \cdot \frac{d^2 T^{-1}(\psi)}{d\psi^2}.$$
(3.6b)

For the transformed MAP estimate $\psi = T(\hat{\theta}_n^{\text{MAP}})$, the first term, Eq. (3.6a), is zero, since its first factor vanishes by definition. The second term, Eq. (3.6b), however, is non-zero, since both its factors are non-zero for nonlinear T. Therefore, the transformed MAP estimate does not satisfy the condition for the mode of the transformed posterior probability and hence, $T(\hat{\theta}_n^{\text{MAP}}) \neq \hat{\psi}_n^{\text{MAP}}$.

The transformation of random variables is illustrated in Figure 3.3 for the case of transforming a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with the exponential function $T(\theta) = \exp(\theta)$, which results in the lognormal distribution $\mathcal{LN}(\mu, \sigma^2)$ with probability density function

$$p_{\Psi}(\psi) = \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln\psi-\mu)^2}{2\sigma^2}\right)}_{=p_{\Theta}(T^{-1}(\psi))} \cdot \underbrace{\frac{1}{\psi}}_{=\frac{dT^{-1}(\psi)}{d\psi}}$$

The mode of the normal distribution is $\mu = \hat{\theta}^{MAP}$ but the mode of the lognormal is defined as $\exp(\mu - \sigma^2) \neq \exp(\mu)$.

In the following, we investigate this theoretical aspect in the specific application of interest and compare the different algorithms proposed for Bayesian forecasting in the therapeutically relevant example of chemotherapy-induced neutropenia.

3.5 Comprehensive and informative therapy forecasting and decision-support in oncology

Two simulation studies were performed in MATLAB R2017b/2018b to analyze the approaches regarding their suitability to support MIPD for the frequently used anticancer agents docetaxel and paclitaxel.

Biomarker data for single/multiple cycle simulation studies. For the single cycle study with docetaxel (100 mg/m² BSA, 1 h i.v. infusion), we used the NLME model in [67] as prior knowledge, see Section 2.1.3 and Section A.2.1+A.2.3. It is based on the well-known PD model in [91] and describes the effect of a single dose of the anticancer drug docetaxel based on monitoring neutrophil counts. The covariates AGE and AAG were sampled from a normal distribution with mean given by the median and an estimated variance from the given observed range. The parameter estimates used for the PK/PD model are given in Table A.1 & A.3. As in [73] we assume γ to be fixed and estimated the parameters 'Circ₀', 'MTT', and 'Slope'. A virtual population (N = 100) was generated based on the patient characteristics provided in [67]. For inference, neutrophil concentrations were considered on a log-scale at time points $t = 0, 3, \ldots, 21$ days post-dose. This simulation study aims to demonstrate the limitations of MAP estimation for a model frequently used in MIPD (see, e.g. [73, 75]).

Since recursive data processing and decision-support gain in relevance for long-term monitoring, we performed a simulation study for multiple cycle therapy with paclitaxel using the BME model in [90] (Figure 2.2). It describes the effect of the anticancer drug paclitaxel $(200 \text{ mg/m}^2 \text{ BSA}, 3 \text{ h i.v. infusion})$ over 6 cycles of 3 weeks each, corresponding to treatment arm A of the CEPAC-TDM study [52], see Section 2.1.3 and Section B.2.2. The model includes IOV on PK parameters describing the variability between cycles within one patient. Therefore, the parameter values comprise the interindividual parameters (θ_{IIV}), and a parameter for each occasion (θ_{IOV}), see Eq. (3.1). As a consequence, the size of θ increases with every occasion/cycle:

$$\theta = (\theta_{\rm IIV}, \theta_{\rm IOV}^1, \dots, \theta_{\rm IOV}^C), \qquad (3.7)$$

where C denotes the number of cycles, see Section B.2.2 for details.

We were interested in a setting where data becomes available sequentially (one-by-one) and assumed that neutrophil concentrations are monitored every third day, e.g., via a homemonitoring device, see Section 2.1.1. To this end, neutrophil concentration data were simulated for a virtual patient using Eq. (2.3)-(2.4) and the corresponding model (Section 2.1.3). Then the individual parameter values were inferred based on the simulated neutrophil concentration data available up to a certain time point, using the same model. For the statistical analysis, this procedure was repeated for N = 100 virtual patients (with covariate characteristics mirroring the real study population underlying the NLME model).

Method comparison

For all sampling-based methods (NAP, SIR, MCMC, PF) we used a sample of size $M = 10^3$. For MCMC we chose a burn-in of 100 samples in our analysis. For the MAP estimation, parameter bounds are needed for the MATLAB optimizer fmincon. The lower bounds were taken from the code provided in [160] and the upper bounds were tested so that the optimizer did not reach the bound. Since the posterior is analytically intractable, an extensive sample of size $M = 10^6$ was used as a reference, called 'full Bayes (reference)' in the sequel, which was generated by SIR and cross-checked with MCMC (see Figure B.1, since these approaches are exact in the limit $M \to \infty$). As a statistical measure for the quality of uncertainty quantification, we considered the Hellinger distance

$$H(\hat{P}, P^{\text{ref}}) := \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{b} \left(\sqrt{\hat{p}_i} - \sqrt{p_i^{\text{ref}}}\right)^2}, \qquad (3.8)$$

which measures the difference between the discrete sampling-based a-posteriori probability distribution $\hat{P} = (\hat{p}_1, \ldots, \hat{p}_b)$ and the reference solution $P^{\text{ref}} = (p_1^{\text{ref}}, \ldots, p_b^{\text{ref}})$ generated with SIR and $M = 10^6$ for *b* fixed bins.

First, we show the limitations of MAP estimation for MIPD and how full Bayesian approaches can overcome these limitations (using SIR with $M = 10^6$ as reference). Next, we compare different full Bayesian approaches with reduced sample sizes ($M = 10^3$) regarding accuracy and computational efficiency.

3.5.1 Unfavorable properties of MAP-based predictions

The first example of decision support in individualized chemotherapy employs the most frequently used model of neutropenia [91]. The MAP estimate $\hat{\theta}_n^{\text{MAP}}$ is derived given experimental data $y_{1:n} = (y_1, \ldots, y_n)^T$, see Eq. (2.10) and Eq. (3.3) for the specific setting of interest. In the context of TDM, it is used to predict the future time course $x(t; \hat{\theta}_n^{\text{MAP}})$ of the patient and thereon based observables. In mathematical terms, $\hat{\theta}_n^{\text{MAP}}$ is mapped to some quantity of interest $T(\hat{\theta}_n^{\text{MAP}})$, e.g., the nadir concentration. As pharmacometric models are generally nonlinear, this does, however, not result in the most probable outcome (see Section 3.4). This is due to the fact that first determining the MAP estimate and then applying a nonlinear mapping is in general different from first applying the mapping to the full parameter posterior and then determining its MAP estimate: $T(\hat{\theta}_n^{MAP}) \neq \widetilde{T(\theta_n)}^{MAP}$. Figure 3.4 illustrates this aspect with $T(\theta) = c_{nadir}(\theta)$. Note that the mode of the posterior is correctly identified in parameter space (left panel), however, due to the nonlinear transformation the MAP-based nadir concentration $c_{\text{nadir}}(\hat{\theta}_n^{\text{MAP}})$ does not correspond to the a-posteriori most probable nadir (mode of the posterior of the nadir concentration). The MAP-based nadir concentration predicts grade 3 neutropenia, while the a-posteriori most probable nadir concentration is within the range of grade 4 neutropenia. In this example the severity of neutropenia is critically underestimated with possible clinical implications. Additional details and analyses are provided in Figure B.2.

Thus, MAP-based estimation lacks both, a measure of uncertainty and the feature to predict the most probable outcome/quantity of interest. In addition, relevant outcomes such as the risk of grade 4 neutropenia can not be evaluated from the point estimate alone. MAP-based estimation, therefore, provides a biased basis for clinical decision-making. In contrast, full Bayesian inference provides access to the full posterior distribution of the parameters

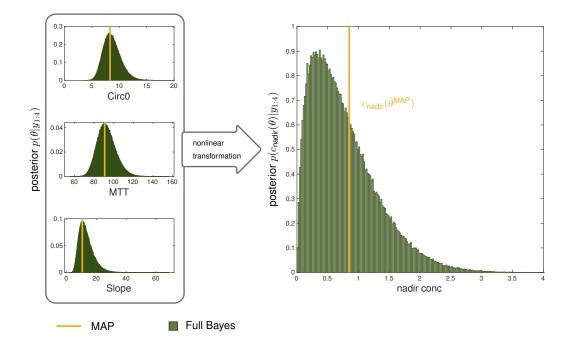


Figure 3.4: MAP-based predicted nadir concentration is not the most (a-posteriori) probable nadir concentration. We considered the single cycle study with docetaxel with four observed data points $y_{1:4}$ and forecasted the nadir concentration based on the posterior $p(\theta|y_{1:4})$ using MAP estimation and the full Bayes (reference) approach (SIR with $M = 10^6$), where θ comprises the parameters 'Circ₀' (baseline neutrophil counts), 'MTT' (mean transit time) and 'Slope' (drug effect). The MAP estimate of the parameters coincides with the mode of the posterior distribution of the parameters (left panel), however, the mode is not preserved under nonlinear transformation (see text). Therefore, $c_{\text{nadir}}(\hat{\theta}^{\text{MAP}})$ with $c_{\text{nadir}}(\cdot)$ denoting some observable $T(\cdot)$ does not equal the mode of the a-posteriori probability $p_{T(\Theta)}(\cdot|y_{1:n})$ of the nadir concentration. Please also refer to Figure B.2 for further illustration and analysis.

and correctly transforms uncertainties forward to the observables and quantities of interest, allowing to compute any desired summary statistic and relevant risks [103, Section 5.2].

3.5.2 Uncertainty quantifications for more comprehensive, differentiated understanding and thus better informed decisionmaking

The first scenario served to demonstrate the limitations of MAP-based predictions for the gold-standard model [91], however, the model does not account for the observed cumulative neutropenia over multiple cycles. Therefore, we considered for dose adaptations the BME model accounting for bone marrow exhaution over multiple cycles [90], see Section 2.1.3. We exemplarily considered the dose selection for the third treatment cycle based on prior information and patient-specific measurements during the first two cycles. The patient-specific data together with the full Bayes (reference) model fit and prediction are shown in Figure 3.5 A. The CrIs (dashed) and PIs (dotted) show the uncertainty about the therapy outcome (in observable space), without and with measurement errors, respectively.

For optimizing the dose of the third cycle, different dosing scenarios were considered: the standard dose and a -15%, -30%, and +10% adapted dose. Figure 3.5 B shows the

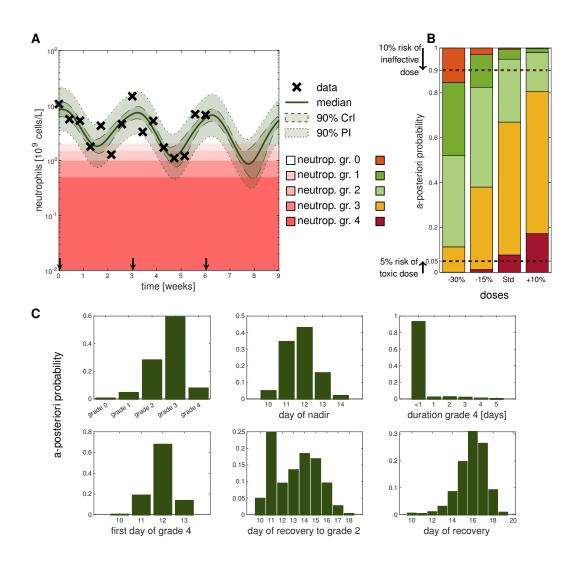


Figure 3.5: Uncertainty quantification by full Bayesian methods gives important information for therapy dosing selection. The scenario described in the 'multiple cycle study with paclitaxel' is used and the results are shown for the full Bayes (reference) solution with SIR using $M = 10^6$ samples. (A) Forecasting the third cycle for different doses based on the patient's covariates and measurements of the first two cycles. (B) Full Bayesian inference allows for probabilistic statements of the different grades. Color coding of neutropenia grades shows the trade-off between efficacy and toxicity. No toxicity (grade 0) is associated with ineffective treatment (orange) but severe neutropenia (grade 3 and 4) is also not desired (yellow and red). (C) A-posteriori probabilities of quantities of interest for the third cycle based on the posterior at the end of second cycle (week 6) for the standard dose. Statistics such as day of grade 4 were computed given that grade 4 is reached.

probability of the predicted grades of the third cycle for each dose. To find an effective and safe dose, the risk of being ineffective (neutropenia grade 0) should be minimized jointly with the risk of being unsafe (neutropenia grade 4). For illustration in Figure 3.5 B, the dashed horizontal lines indicate a 10% and 5% level of being ineffective and unsafe, respectively. The standard dose and the increased dose have a risk of toxicity larger than 5% (lower horizontal line). A decrease in dose also leads to an increased risk of an ineffective dose (upper horizontal line). The 15% reduced dose is with 96% probability safe and efficacious (grade 1–3), with 3% probability ineffective (grade 0) and with 1% probability unacceptably toxic (grade 4). If grade 3 is also to be avoided, the 30% reduction would be preferable, as it is with 74%probability safe and efficacious (grade 1-2), with 15% probability ineffective (grade 0), and with 11% probability toxic (grade 3–4). Thus, the choice of an optimal dose might depend on how priority is given to the risk of inefficacy and toxicity. As both risks are described by the tails of the posterior distribution, a point estimate is not able to adequately capture them. The MAP-based predicted grades were: grade 2 (standard dose and +10% dose), grade 1 (-15% dose), and grade 0 (-30% dose), which do not only make it difficult to distinguish between some doses but also do not reflect the true most probable grades.

Posterior-based predictions of important statistics related to the neutropenia time course can help to answer questions like "How probable is it that the patient will suffer from grade 4 neutropenia?" or "How probable is it that the patient will recover in time for the next scheduled dose so that the therapy can be continued as planned?". To answer such questions, Figure 3.5 C shows important predicted quantities of interest, illustrated for the standard dose in cycle 3. We inferred that the risk of grade 4 neutropenia is 8%, and if the patient were to reach grade 4, it would be most probable (68%) on day 12. The probability that the patient's duration in grade 4 is a day or longer is very small (<7%). As the probability of no recovery until day 21 is negligible, the administration can remain scheduled on day 21 for cycle 4. Therefore, uncertainty quantification improves the decision-making process regarding dose adjustments as well as therapy management by quantifying the a-posteriori probabilities of relevant risks and quantities of interests. Repeating the above analysis for different doses therefore allows for an improved distinction between dose adjustments.

3.5.3 Approximation accuracies comparable across different full Bayesian approaches

We next compared different established methods for uncertainty quantification with regard to their approximation accuracy. To this end, posterior inference was investigated for a patient on day 5 of the first cycle (Figure 3.6). Whereas the marginal posterior distribution for the parameter ' Circ_0 ' (pre-treatment neutrophil concentration) is close to a normal distribution, the marginal posterior for the drug effect parameter ('Slope') is closer to a lognormal distribution. Accordingly, the NAP is rather reasonable for 'Circ₀', but is questionable for the 'Slope' parameter. In addition, sampling from the normal distribution can lead to unrealistic (negative) parameter values (Figure 3.6 A, right panel). Also, NAP very clearly underestimated the patient's risk to reach grade 4 neutropenia (Figure 3.6 B and C), which could possibly lead to a fatal dose selection. Here shown for the simulation-based approach to propagate the uncertainty. Using the delta method leads to a similar biased prediction and does not allow to compute the uncertainties for the nadir concentration, see Figure B.3. Considering a Student's t distribution instead of the normal approximation, as in [154], did not lead to an adequate improvement (Figure B.4). Consequently, the NAP approach can result in over-optimistic, over-pessimistic and unrealistic predictions. In contrast, the full Bayesian methods (SIR, MCMC, and PF) adequately represent the tails of the posterior

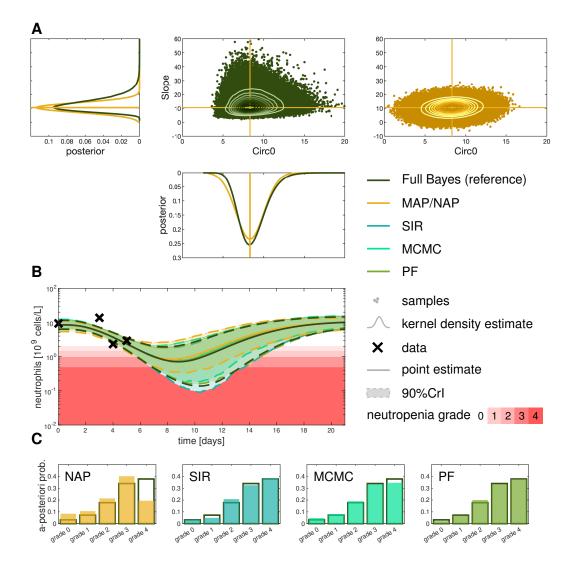


Figure 3.6: Uncertainty quantification of the different methods at the level of parameters, observables and quantities of interest. Exemplary comparison of the different methods for one patient after having observed four data points up to day 5. (A) The posterior is shown for parameters 'Slope' and 'Circ₀' showing the kernel density estimates of the marginal sampling distribution and as scatter plots for the bivariate sampling distributions with contour plots for the full Bayesian approach (full Bayes (reference), SIR with $S = 10^6$) and the normal approximation located at the MAP estimate. (B) On the level of the observable (neutrophil concentration) the point estimates (median or MAP) are displayed along with the 90% credible intervals (CrI). For illustration purposes the prediction intervals are not shown here. (C) The forecasted a-posteriori probability of the different neutropenia grades (0-4) is shown for the different approximations (filled bars) in comparison with the full Bayes (reference) (dark green outlined bars).

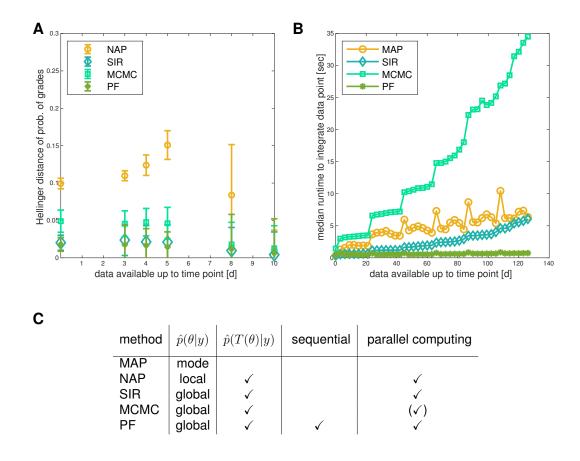


Figure 3.7: Comparison of methods regarding important aspects for model-informed precision dosing (MIPD). (A) Approximation error (measured as Hellinger distance) of the probability of neutropenia grades. ('Single cycle study docetaxel'). (B) Qualitative runtime comparison to sample from the parameter posterior. Median of N = 100 repeated analyses ('Multiple cycle study paclitaxel'). (C) Comparison of method properties. For MCMC several chains could be run in parallel, however, in this study only one chain was considered.

and respect the positivity constraint of parameter values. The resulting CrIs are comparable to the reference CrIs. For the proposed MCMC method, the adaptive M-H sampler, the acceptance rates are compared with the M-H sampler with fixed variance Ω in Figure B.5. For illustration, Figure 3.7 A shows the approximation error for the predicted probability of neutropenia grades, measured in the Hellinger distance Eq. (3.8). Overall, SIR and PF showed the best approximation, while NAP resulted in the largest errors. MAP is not included as it does not allow to infer the a-posteriori probability of the neutropenia grades. We also compared the methods for the previously described dose selection scenario (Section 3.5.2), see Figures B.6 and B.7.

3.5.4 Sequential DA processes patient data most efficiently

The need for real-time inference algorithms is increasing with the possibility to more frequently collect patient-specific data (online collection) during treatment. Sequential DA methods (e.g.,

PF) provide an efficient framework for real-time data processing. At any time, all information (incl. associated uncertainty) is present in a collection of particles that can be interpreted as representing the current state and associated uncertainty of a patient via a virtual population. With a new datum, this information is updated. Approaches that rely on batch data analysis, i.e., MAP, SIR, MCMC, need to redo the inference from scratch including one more data point. This has an impact on the computational effort as the number of data points increases. Figure 3.7 B shows a comparison of the computational cost to assimilate an additional data point. All approaches show some kind of increase in effort every 21 days—due to the IOV on some parameters. Clearly, PF shows lowest and almost constant costs, while for batch mode approaches computational costs increase over time due to an increasing number of parameters (one additional parameter for every cycle due to the IOV, see Eq. (3.7) and Section B.2.2) and an increasing integration time span to determine the likelihood. This could become computationally expensive in view of long term treatments and higher time resolution of data points provided by new digital health care devices. Figure 3.7 C summarizes the features of the different inference approaches. Note that all sampling-based approaches can be accelerated by parallel computing (MCMC only in terms of multiple chains). In summary, it was found that PF processes patient monitoring data most efficiently and facilitated the handling of IOV because only the IOV parameter of the current occasion needs to be considered (the size of the parameter vector is constant).

3.6 Discussion

In the context of chemotherapy-induced neutropenia, we illustrated the severe drawbacks of MAP-based approaches for Bayesian forecasting and thereon based decision-making. A prediction based on the MAP estimate does neither necessarily correspond to the most probable outcome, nor does it allow to quantify relevant risks as the uncertainties are not quantified. Both are highly undesirable characteristics and make MAP-based inference difficult to interpret in a TDM setting and unsuitable for MIPD. A normal approximation of the posterior at the MAP estimate is no alternative, as it retains the same point estimate and is inappropriate in case of skewed parameter distributions. We demonstrated that full Bayesian approaches, like SIR, MCMC, or PF provide accurate approximations to the posterior distribution, enabling comprehensive uncertainty quantification of the quantities of interest (e.g., nadir concentration). Amongst the three considered approaches, PF is a sequential approach that is beneficial in a more frequent monitoring context.

Uncertainty quantification in TDM is scarce. In [161] the SIR algorithm was previously used in the TDM setting to construct CrIs using a Student's t distribution located at the MAP estimate as importance function. A sequential approach in the context of MAP estimation is discussed in [162] and in [163] with a moving estimation horizon (window of data points that are considered). A sequential DA approach has been investigated previously for glucose forecasting [125, 28]. Here, we present DA in combination with a NLME modeling framework, i.e., leveraging the prior distributions resulting from previous population analyses and provide decision support statistics based on a-posteriori probabilities. A systematic comparison of approaches for posterior inference in the TDM context, as presented in this chapter is lacking.

In this study, particle filtering is applied in a TDM setting within an NLME modeling framework to represent the current patient status via an uncertainty ensemble. A challenge in the application of PF is the potential for weight degeneracy, i.e., a gradual separation into a few large and many very small weights. A rejuvenation approach (as applied in this study) resolves this problem but requires to specify an additional parameter (magnitude of the rejuvenation). A too large value might result in an artificially increased uncertainty, while a too small value might hinder exploration of the parameter space. In the present application context, however, IOV counteracts, in addition to the rejuvenation step, weight degeneracy.

Possible future work. Among the possible extensions of the described approaches is the use of alternative sequential DA algorithms. In the present thesis, the classical particle filter has been applied with an augmented state to allow simultaneous state and parameter estimation. Particle filters are, however, only feasible in a low-dimensional augmented state space as the necessary ensemble size increases exponentially with the state dimension (curse of dimensionality) [164]. For more complex models with a larger number of states and parameters, alternative sequential DA approaches should be considered, e.g., ensemble Kalman filters or more recently developed hybrid methods that balance robustness, efficiency, and accuracy [165, 166, 167].

Future research should also explore aspects related to parameter sensitivity. In contrast to state estimation, where each new data point provides information about the current state, e.g., in the case of object tracking, not every new data point is informative for all parameters, e.g., data points taken in the elimination phase will not be informative about the absorption phase of the drug. Therefore, rejuvenation could lead to an inflated variance in the direction of the parameters for which the data are not informative. This could result in an overestimation of the a-posteriori uncertainty. Scaling the rejuvenation with the information content of the data as given by the FIM could be a possible extension. This could allow for a larger rejuvenation parameter and thus a better exploration of the parameter space. In this respect, optimal experimental design approaches could also be beneficially exploited to better control the timing of TDM for improved parameter estimation, see e.g., [168].

Sequential data processing is not only computationally efficient and convenient for IOV handling, but has the additional advantage that already assimilated experimental data need not be stored to assimilate future data points. Sampling approaches allow an extension for hierarchical models to include the uncertainties in the population parameters for an even more holistic uncertainty quantification [156]. This would enable a continuous learning process between clinical trials from drug development and continue during the acquisition of real-world data after market authorization, in quantifying the diverse population of patients that have taken a given drug. For a future patient, this 'historic' diversity would transform into well-quantified uncertainty in a TDM setting. The absence of the need to store 'historic' experimental data can also be helpful for the exchange of information between clinics, health insurances, and pharmaceutical companies. The current knowledge, present in form of a sample of particles, can easily be exchanged without the need to exchange the experimental data. The 'historic' data are implicitly present in the particles. This idea will be taken up again and pursued further in Chapter 5.

In view of new treatments and new mobile health care devices (e.g., wearables) gathering data from various sources, clinicians have to deal with new challenges and an increasing complexity of treatment decision-making, which demands comprehensive approaches that integrate data efficiently and provide informative and reliable decision-support. In this chapter, we illustrated that a comprehensive uncertainty quantification can lead to a more meaningful, reliable, and differentiated therapy forecasting and subsequent decision-support. This is not limited to individualized chemotherapy but has the potential to improve patient care in various therapeutic areas in which TDM is indicated, such as oncology, infectious diseases, inflammatory diseases, psychiatry, and transplantation patients.

4 Dose individualization

The previous chapter was concerned with forecasting the individual therapy outcome for a given standard dose, including clinical decision support based on forecasting therapy outcomes for certain fixed dose adjustments. In this chapter, which is based on [CM2], we investigate quantitative approaches to determine model-informed individual dose recommendations. Dose individualization is at the core of MIPD, and the way in which an individual dose can be selected depends on the method used for Bayesian forecasting.

Therefore, a popular approach towards MIPD is to evaluate MAP-based outcomes with respect to a utility function or a target concentration to determine the optimal next dose (MAP-guided dosing) [84, 76]. The definition of a target concentration or utility function is, however, difficult since in many therapies rather subtherapeutic or toxic *ranges* are known. For therapeutic ranges, MAP-guided dosing is not readily suited [21], since only a (potentially biased) point estimate is used, neglecting associated uncertainties, as demonstrated in Chapter 3 (Figure 3.4). A post-hoc uncertainty quantification for MAP-based predictions often relies on a normal approximation located at the MAP estimate, which was shown to not necessarily transform accurately into quantities of interest for nonlinear models, e.g., to the a-posteriori probabilities of the neutropenia grades (Figure 3.6).

Building on the results of the previous chapter, which showed that Bayesian DA approaches provide more informative therapy forecasting by fully exploiting patient-specific information, we develop new approaches for dose selection within MIPD. DA enables individualized uncertainty quantification, which allows straightforwardly (i) to integrate both, safety and efficacy aspects into the objective function of determining the optimal dose, or (ii) to compute the probability of being within/outside the target range (Figure 3.5). However, optimizing across a whole therapy time frame for multiple dosage regimens to also account for delayed effects of dose selections can be hard and potentially too costly for real-time decision support.

RL has been applied to various fields in health care, however, mainly focusing on clinical trial design [36, 37], and only a few studies relate to optimal dosing in a PK/PD context [31, 38], as discussed in Section 2.3.3. In model-based RL, model simulations are used to learn how to act best in an uncertain environment where decisions have to be made in stages and might have delayed effects. A key aspect of learning is to make successively use of knowledge already acquired, while also exploring yet unknown sequences of actions. The result is typically a decision tree (or some functional relationship). In other words, the physician's decision is supported via a pre-calculated, extensive, and detailed look-up table without additional computation during the course of therapy. So far, RL approaches in health care are limited to

rather simple exploration strategies (so-called ϵ -greedy approaches) with one time-step ahead approximations of the look-up table (Q-learning) [37].

In this chapter, we demonstrate how DA and RL can be very beneficially exploited to develop new approaches to MIPD. The first approach referred to as DA-guided dosing, improves existing online MIPD by integrating model uncertainties into the dose selection process. For the second approach (RL-guided dosing) we propose MCTS in conjunction with UCT [142, 169] (Section 2.3) as a sophisticated learning strategy for an optimal dosing policy. The third approach combines DA and RL (DA-RL-guided dosing) to make full use of patient TDM data and to provide a flexible, interpretable, and extendable framework. We compared the three proposed approaches with current dosing strategies (standard, PK-guided, and MAP-guided dosing) in terms of dosing performance and their ability to provide insights into the factors driving dose selection.

Simulation study framework. We consider a single dose every three weeks schedule for paclitaxel-based chemotherapy, $c = 1, \ldots, C$, for a total of six cycles (C = 6), as in the multiple cycle study in Chapter 3. We denote the decision time point for the dose of cycle c by T_c , and assume $T_1 = 0$ (therapy start). For dose selection, the physician has different sources of information available, such as the patient's covariates 'cov' (sex, age, etc), the treatment history (drug, dosing regimen, etc), and TDM data related to PK/PD (drug concentrations, response, toxicity, etc). Despite these multiple sources of information, it remains a partial and imperfect information problem, as only noisy measurements of few quantities of interest at certain time points are available. MIPD aims to provide decision support by linking prior information on the drug-patient-disease system with patient-specific TDM data.

The standard dosing for 3-weekly paclitaxel, as applied in the CEPAC-TDM study arm A, is 200 mg/m² BSA and a 20 % dose reduction if neutropenia grade 4 ($g_c = 4$) was observed [52], see also Section 2.1.2. The aforementioned model-informed dosing table (MIDT) (termed *PK-guided dosing* [11]) was evaluated in study arm B, see Section A.1. For dose selection at cycle start T_c , we chose the patient state

$$s_{c-1} = \left(\text{sex, age; ANC}_0, g_1, \dots, g_{c-1}\right), \tag{4.1}$$

with $s_0 = (\text{sex}, \text{age}; \text{ANC}_0)$ the pretreatment state. The covariates sex, age, have previously been identified as important predictors of exposure [11], and baseline absolute neutrophil counts ANC_0 , as a crucial parameter in the PD model [91, 90] (Section 2.1.3). We included the neutropenia grades of all previous cycles $g_{1:c-1} = (g_1, \ldots, g_{c-1})$ based on the nadir concentrations to account for the observed cumulative behavior of neutropenia [90, 170].

We used the term 'model' to refer to Eqs. (2.1)-(2.3), and the term 'model state of the patient' to refer to a model-based representation of the state of the patient, i.e., a distribution of state-parameter pairs (x, θ) , or just a single (reference) state-parameter pair. In the proposed approaches, the model is used to simulate treatment outcomes (in RL called 'simulated experience'), or to assimilate TDM data and infer the model state of the patient, or both. To infer the patient state Eq. (4.1), the grade of neutropenia of the previous cycle g_{c-1} needs to be determined; either directly from the TDM data $(y_{c-1} \mapsto g_{c-1} \mapsto s_{c-1})$ or based on a model simulation of the model state of the patient $((x, \theta) \mapsto c_{\text{nadir}} \mapsto g_{c-1} \mapsto s_{c-1})$. Since generally measurements are not taken exactly at the time of nadir, the model-predicted nadir may provide an improved state estimate.

Overview of the different approaches towards MIPD. We considered three different approaches towards MIPD, see Figure 4.1:

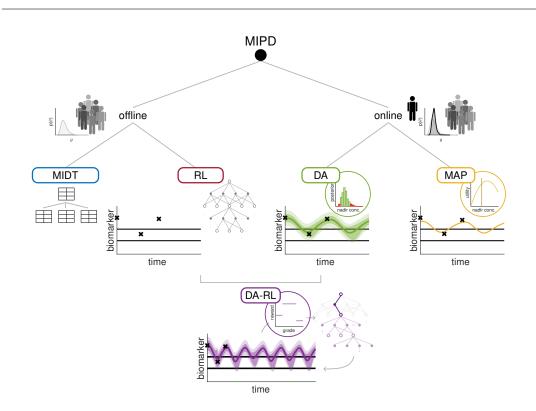


Figure 4.1: Overview of different approaches for model-informed precision dosing (MIPD). The different methods can be categorized according to the time when the computational effort to calculate the optimal dose must be made. *Offline approaches* calculate optimal doses for all possible covariates and state combinations prior to any treatment, like in model-informed dosing tables (MIDT) and reinforcement learning (RL). The physician selects the dosing recommendation in the table/tree based on specific patient information (covariates, observations). While the TDM data (measured drug or biomarker) are used to determine the entry in the table/tree the table/tree itself is static. *Online approaches* solve an optimization problem at every decision time point, i.e., when a dose has to be given. They integrate patient-specific TDM data using Bayesian data assimilation (DA) or maximum a-posteriori (MAP) estimation. *Offline-online approaches* allocate computational resources between offline and online. Pre-calculated dosing decision-trees are individualized during treatment, based on TDM data.

- (i) Offline approaches support dose individualization based on MIDT or dosing decision trees (RL-guided dosing). At the start of therapy, a dose based on the patient's covariates and baseline measurements is recommended. During therapy, the observed TDM data are used to determine a path through the table/tree; While the treatment is individualized to the patient (based on a-priori uncertainties), the procedure of dose individualization itself does not change, i.e., the tree/table is static. As such, it can be communicated to the physician before the start of therapy.
- (ii) Online approaches determine dose recommendations based on a model state of the patient and its simulated outcome. Bayesian DA or MAP estimation assimilate individual TDM data to infer the posterior distribution or MAP point-estimate as model state of the patient, respectively. Online approaches tailor the model (more precisely, the parameters) to the patient, however, clinical implementation requires an IT infrastructure and/or

easy-to-use software. While this might constitute a challenging problem that hinders broad application [171], successful examples of implementation already exist [172].

(iii) Offline-Online approaches combine the advantages of dosing decision trees and an individualized model. The individualized model is used in two ways, to infer the patient state more reliably than sparsely observed TDM data and to individualize the dosing decision tree (using individualized uncertainties, rather than population-based uncertainties).

Key to all approaches is the so-called reward function R (RL terminology), also often termed cost or utility function, defined on the set S of patient states

j

$$R: \mathcal{S} \to \mathbb{R}. \tag{4.2}$$

Ideally, the reward corresponds to the net utility of beneficial and noxious effects in a patient given the *current* state [173]. For neutrophil-guided dosing, a reward function was suggested that maps (MAP-based) nadir concentrations to a continuous score [84] or penalizes the deviation from a target nadir concentration $(c_{\text{nadir}} = 1 \cdot 10^9 \text{ cells/L})$ [76]; we used in this study a utility function but also provide a comparison of the results with the suggested target concentration, see also Section C.3 and Figure C.5. The individualized uncertainties quantified via DA allow to consider the probability of being within/outside the target range in the reward function [CM1], which is more closely related to clinical reality. For the patient state Eq. (4.1) used in RL, we also designed the reward function Eq. (4.2) to account for efficacy and toxicity. We chose to penalize the short-term goal (avoiding life-threatening grade 4) more than the long-term goal (increased median (overall) survival associated with neutropenia grades 1–4 [74]) :

$$R(s_c) = \begin{cases} -1 & \text{if } g_c = 0, \\ 1 & \text{if } g_c = 1, 2, 3, \\ -2 & \text{if } g_c = 4, \end{cases}$$
(4.3)

see also Section 2.1.1 for more details.

4.1 Offline approaches

4.1.1 Model-informed dosing tables

MIDTs provide a simple format to make model-informed dose adjustments accessible. Yet the dose adaptations are very limited in resolution and are usually restricted to a few covariates and previous observations. We considered as an example and also to benchmark the developed approaches the PK-guided dosing algorithm [11], see Figure A.1. The dose is selected according to covariates, a model-informed PK measure, and the neutropenia grade of the previous cycle.

4.1.2 RL-guided dosing

As described in Section 2.3, RL problems can be formalized as MDPs, modeling sequential decision-making under uncertainty. Here, we focus on translating the general RL formulation into a clinical setting. The virtual physician's task is to learn a dosing strategy in order to optimize the expected long-term therapy outcome of the virtual patient, see Figure 4.2 for an illustration of the typical agent-environment interaction for a single dose selection in the clinical setting of interest.

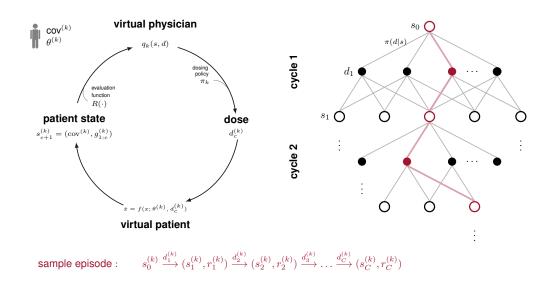


Figure 4.2: Model-based reinforcement learning (planning). The expected long-term return (action-value function) is estimated based on simulated experience (sample approximation Eq. (4.4)). For simulating experience, an ensemble of virtual patients is generated, $k = 1 \dots, K$ (for all covariate classes $\mathcal{COV}_l, l = 1, \dots, L$, covariates $\operatorname{cov}^{(k)}$ are sampled within the covariate class and model parameters $\theta^{(k)}$ are sampled from the prior distribution). At start of each cycle c, a dose $d_{c+1}^{(k)}$ is chosen according to the current policy π_k , and the outcome (grade of neutropenia) is predicted based on the model $\dot{x} = f(x; \theta^{(k)}, d_c^{(k)})$ for the sample parameter vector $\theta^{(k)}$ and chosen dose. The updated patient state $s_c^{(k)}$ is assessed using the reward function R. The sequential dose selections (going through the circle C times (left part)) lead to so-called sample episodes; the entirety of episodes to a tree structure (right part).

The decision time points of the MDP are given by the treatment cycles c. We consider a therapy of six cycles, $c = 1, \ldots, C$ with C = 6, i.e., an episodic task. Every episode corresponds to a path in the tree of possibilities (Figure 4.2). In each cycle, the patient state S_c is inferred and a dose D_c selected. Note that we adapted the general RL formulation, introduced in Section 2.3, to the specific setting, i.e., the decision time instances are the cycles (c instead of t) and the actions are the different doses (D instead of A). Due to unexplained variability between patients (and occasions), transitions between states are characterized by transition probabilities Eq. (2.22), $\mathbb{P}[S_{c+1} = s_{c+1}|S_c = s_c, D_{c+1} = d_{c+1}]$. The reward is defined via the reward function, i.e., $R_c = R(S_c)$ (Eq. (4.2)), and the dosing policy π (see also Eq. (2.23)) models how to choose the dose for the next treatment cycle

$$\pi(d|s) = \mathbb{P}[D_{c+1} = d|S_c = s].$$

A dosing policy is evaluated based on the return G_c at time step c, defined as the weighted sum of rewards over the *remaining* course of therapy, Eq. (2.24) with c = t. The discount factor $\gamma \in [0, 1]$ balances between short-term ($\gamma \to 0$) and long-term ($\gamma \to 1$) therapeutic goals (see Section 2.3 and C.5.2).

As aforementioned model-based RL methods that rely on sampling estimate the expected value in Eq. (2.25) via a sample approximation. To simplify the calculations we have discretized the continuous covariates age and ANC₀ into covariate classes COV_l , l = 1, ..., L. For each class COV_l consider the ensemble

$$\mathcal{E}_{\mathrm{RL}}(\mathcal{COV}_l) := \left\{ \left(x_0(\theta_c^{(k)}), \theta_c^{(k)}, \operatorname{cov}^{(k)} \right) \right\}_{k=1}^K$$

with $\operatorname{cov}^{(k)}$ sampled within \mathcal{COV}_l according to the covariate distributions in the CEPAC-TDM study [52, 95], parameter values sampled from $p_{\Theta}(\theta^{\mathrm{TV}}(\operatorname{cov}^{(k)}), \Omega)$ and initial states according to (2.1). Then, for each $k = 1, \ldots, K$ with K large, a sample episode

$$s_0^{(k)} \xrightarrow{d_1^{(k)}} (s_1^{(k)}, r_1^{(k)}) \xrightarrow{d_2^{(k)}} (s_2^{(k)}, r_2^{(k)}) \xrightarrow{d_3^{(k)}} \dots \xrightarrow{d_C^{(k)}} (s_C^{(k)}, r_C^{(k)})$$

using policy π_k is determined and

$$q_k(s,d) = \frac{1}{N_k(s,d)} \sum_{k'=1}^k \sum_{c=1}^C \mathbf{1}_{\{s_c^{(k')} = s, d_{c+1}^{(k')} = d\}} G_c^{(k')}$$
(4.4)

computed, see also Eq. (2.27). Here, $N_k(s, d)$ denotes the number of times that dose d was chosen in patient state s amongst the first k episodes.

To balance exploration and exploitation, we employed MCTS in conjunction with UCT as policy in the iterative training process [132, 144, 141, 142, 169]:

$$\pi_{k+1}(d_{c+1}|s_c) = \begin{cases} 1 & \text{if } d_{c+1} = \underset{d \in \mathcal{D}}{\operatorname{arg\,max}} \operatorname{UCT}_k(s_c, d), \\ 0 & \text{else} \end{cases}$$
(4.5)

with UCT_k defined based on the current sample estimate $q_k(s_c, d)$

$$\operatorname{UCT}_{k}(s_{c},d) = \underbrace{q_{k}(s_{c},d)}_{\operatorname{exploitation}} + \epsilon_{c} \underbrace{\frac{\sqrt{N_{k}(s_{c})}}{N_{k}(s_{c},d)+1}}_{\operatorname{exploration}}.$$
(4.6)

It successively expands the search tree (Figure 4.2) by focusing on promising doses (exploitation, large $q_k(s_c, d)$), while also encouraging exploration of doses that have not yet been tested exhaustively (small $N_k(s_c, d)$ relative to the total number of visits $N_k(s_c) := \sum_{d'} N_k(s_c, d')$ to state s_c). The parameter ϵ_c balances exploration vs. exploitation; it depends on the range of possible values of the return and current state of the therapy (cycle c), see Eq. (4.9). Finally, we define $\hat{\pi}_{\text{UCT}} = \pi_K$ as an estimate of the optimal dosing policy in the training setting (learning with virtual patients), and $\hat{q}_{\pi_{\text{UCT}}} = q_K$ as an estimate of the associated expected long term return. In a clinical TDM setting (RL-guided dosing), we finally use $\pi^* = \arg \max \hat{q}_{\pi_{\text{UCT}}}$, i.e., $\epsilon_c = 0$ (no exploration) in Eq. (4.6). See Section 2.3.2 for details.

4.2 Online approaches

4.2.1 MAP-guided dosing

MAP-guided dosing is widely applied in various therapeutic areas for online therapy individualization [174, 76] and implemented in assorted software tools, e.g., TDMx [175], InsightRX [176]. The optimal dose is determined in a two-step procedure:

- 1. The MAP-estimate $\hat{\theta}_c^{\text{MAP}}$ is computed according to Eq. (2.10) based on the patientspecific TDM data $y_{1:c}$ collected up to the end of cycle c, see Section 2.2.1 & 3.1.
- 2. Then, the MAP estimate is used to generate model predictions for solving the optimization problem in the dose selection at start of cycle c + 1

$$d_{c+1}^* = \underset{d}{\operatorname{arg\,min}} - R(c_{\operatorname{nadir}}(\hat{\theta}_c^{\operatorname{MAP}}, d)),$$

where we investigated different evaluation functions $R(\cdot)$ based on the MAP-based nadir concentration.

We proposed a utility function that was designed to mirror the essence of the reward function, which we employed in RL-guided dosing, to enable a fair comparison $R(s_c) = -s_c^2 + 3 \cdot s_c$, see Figure C.5 (top panel), where the model state of the patient state was given by

$$s_c = \min_{t \in [T_c, T_{c+1}]} C_{\text{neutr}} \left(t; \hat{\theta}^{\text{MAP}}, d \right)$$
.

In order to also offer a comparison to the often-used concept of a target concentration, we also performed target concentration intervention with a target of $c_{\text{nadir}} = 1 \cdot 10^9 \text{cells/L}$ [76]. For this, we minimized the squared difference, i.e., $R(s_c) = (s_c - 1)^2$, see Section C.3.

MAP-guided dosing depends largely on the reliability of MAP-based predictions which, however, do not necessarily represent the most probable therapeutic outcomes and neglect model uncertainties as discussed in Chapter 3. In particular, any distributional information like, e.g., the tails of the posterior distribution, which describe sub-therapeutic as well as toxic ranges, are completely neglected. These tails provide crucial information for dose selection and might be pronounced due to the often chosen lognormal prior parameter distributions. In addition, the choice of the target or utility function has a crucial impact on the optimal dose selection. While the concept of a utility would be quite desirable, the definition is rather challenging since clinically rather therapeutic ranges are observed.

4.2.2 DA-guided dosing

Sequential DA approaches have been introduced in Chapter 3 as more informative and unbiased alternatives to MAP-based predictions of the therapy outcome since they more comprehensively make use of patient-specific TDM data. The individualized uncertainty in the model state of the patient is inferred and propagated to the predicted therapy time course, allowing to predict the probability of possible outcomes. For this, the uncertainty in the individual model parameters is sequentially updated via Eq. (2.17) for cycle-data, i.e.,

$$p(\theta|y_{1:c}) \propto p(y_c|\theta) \cdot p(\theta|y_{1:c-1}),$$

where $y_{1:c} = (y_1, \ldots, y_c)^T$ denotes the patient's TDM data up to and including cycle c, and $y_c = (y_{c1}, \ldots, y_{cn_c})^T$ the measurements taken in cycle c. Since the posterior distribution $p(\theta|y_{1:c})$ generally cannot be determined analytically, DA approaches approximate it by an ensemble of particles:

$$\mathcal{E}_{1:c} := \left\{ \left(x_{1:c}^{(m)}, \theta_c^{(m)}, w_c^{(m)} \right) \right\}_{m=1}^M$$

In our context, a particle represents a potential model state of the patient (for the specific patient covariates cov) with a weighting factor $w_c^{(m)}$ characterizing how probable the state is (given prior knowledge and TDM data up to c) as presented in Chapter 3. As more TDM data are gathered, the Bayesian updates reduce the uncertainty in the model parameters and consequently in the therapeutic outcome, see Figure 3.1 (reduced width of CrI/PI). Since subtherapeutic as well as toxic ranges, i.e., very low or high drug/biomarker concentrations, are described by the tails of the posterior distribution, the uncertainties provide crucial additional information compared to the mode (MAP estimate) for dose selection.

We chose the *optimal dose* to be the dose that minimizes the weighted risk of being outside the target range; in our context the a-posteriori probability of $g_c = 0$ or $g_c = 4$:

$$d_{c+1}^* = \underset{d \in \mathcal{D}}{\operatorname{arg\,min}} \ \lambda_0 \sum_{m=1}^M w_c^{(m)} \mathbf{1}_{\{g(\theta_c^{(m)}, d)=0\}} + \lambda_4 \sum_{m=1}^M w_c^{(m)} \mathbf{1}_{\{g(\theta_c^{(m)}, d)=4\}}$$
(4.7)

with $g(\theta_c^{(m)}, d)$ denoting the predicted neutropenia grade based on the nadir concentration by forward simulation of the *m*-th particle for dose *d*, see Figure C.6 for an illustration. We penalized grade 4 more severely than grade 0, i.e., $\lambda_4 = 2/3$ and $\lambda_0 = 1/3$, similarly as in Eq. (4.3).

The integration of an ensemble of particles into the optimization problem, instead of a point estimate (as in MAP-guided dosing), increases the computational effort and complexity of the problem. If time or computing power is limited, approximations have to be used, e.g., by solving only for the next cycle dose rather than all remaining cycles at the cost of neglecting long-term effects. Alternatively, the number of particles M could be reduced (we used both approximations in this study). The DA optimization problem is stated in the space of actions (doses). RL, on the contrary, optimizes in the space of states by estimating the expected long-term return as an intermediate step (Eq. (4.4)), thereby promising efficient solutions to the sequential decision-making problem under uncertainty [32].

4.3 Combining offline and online: DA-RL-guided dosing

The particle-based DA scheme and the model-based RL scheme address the problem of personalized dosing from different angles. A combined DA-RL approach, therefore, offers several advantages by integrating individualized uncertainties provided by DA within RL, see Figure 4.3. First, instead of the observed grade (e.g., measured neutrophil concentration on a given day, translated into the neutropenia grade), we may use the smoothed posterior expectation of the quantity of interest (e.g., predicted nadir concentration). This reduces the impact of measurement noise and the dependence on the sampling day, see Section C.6.1 and C.6.2. Second, for model simulations within the RL scheme, we can sample from the posterior $p(\theta|y_{1:c})$ represented by the ensemble $\mathcal{E}_{1:c}$, i.e., from individualized uncertainties, instead of the prior $p(\theta)$, i.e. population-based uncertainties. During the course of the treatment, the ensemble of potential model states of the patient is continuously updated when new patient-specific data are obtained (see Eq. (2.17)). This allows to individualize the expected long-term return during treatment as new patient data are observed, see Figure 4.3, i.e., the dosing decision tree in RL is updated prior to the next dosing decision.

Since the refinement as well as the DA part has to run in real time (online), it has to be performed efficiently. We do not need to take all possible state combinations into account, but only those that are still relevant for the *remaining* part of the therapy. This reduces the computational effort, in particular for later cycles. The proposed DA-RL approach results in a sequence of estimated optimal dosing policies $\hat{\pi}^1, \hat{\pi}^{1:2}, \ldots$ with $\hat{\pi}^{1:c}$ denoting the estimated optimal dosing policy based on TDM data $y_{1:c}$, i.e., based on $\mathcal{E}_{1:c}$. In addition, we do not need to estimate the individualized action-value function from scratch, but can exploit $q_{\pi_0} := \hat{q}_{\pi_{\text{UCT}}}$ as a prior determined by the RL scheme prior to any TDM data (see paragraph following Eq. (4.6)). In PUCT (predictor+UCT [169, 34]), the exploitation vs. exploration parameter

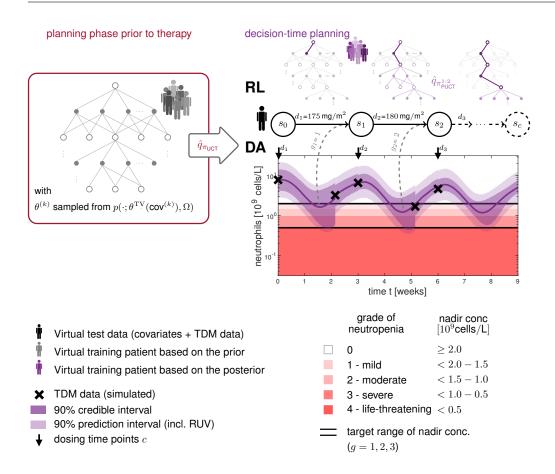


Figure 4.3: The interplay of data assimilation (DA) and reinforcement learning (RL). In the planning phase prior to therapy, the expected long-term return $q_{\pi_0} := \hat{q}_{\pi_{\text{UCT}}}$ is estimated in Monte Carlo Tree search (MCTS) with upper confidence bound applied to trees (UCT) using an ensemble of covariates $\cos^{(k)}$ and parameter values $\theta^{(k)} \sim p(\cdot|\theta^{\text{TV}}(\cos^{(k)}), \Omega)$. The first dose is selected based on $q_{\pi_0} := \hat{q}_{\pi_{\text{UCT}}}$ for the patient specific covariate class. The DA algorithm initializes a particle ensemble given the patient's covariates cov. The ensemble is propagated forward continuously in time, and observed patient TDM data (black crosses) is assimilated when it becomes available. This results in updated uncertainty, visible as 'cuts' in the credible/prediction intervals. In contrast, the RL state evolves in discrete time steps *c* according to the decision time points and only considers selected features/summaries of the model state of the patient, e.g., smoothed posterior expectation of nadir concentrations translated into neutropenia grades. At each decision time point, the posterior model state of the patient is used to refine the prior computed $\hat{q}_{\pi_{\text{UCT}}}$ (grey tree) for future reachable states (light purple tree). This individualizes the tree based on individualized uncertainties ($\mathcal{E}_{1:c}$).

 ϵ_c in Eq. (4.6) is modified to prioritize doses with high a-priori expected long-term return:

$$U_k(s_c, d) = \underbrace{q_k^{1:c}(s_c, d)}_{\text{exploitation}} + \epsilon_c \cdot \underbrace{\frac{\exp(\hat{q}_{\pi_{UCT}}(s, d))}{\sum_{d'} \exp(\hat{q}_{\pi_{UCT}}(s, d'))}}_{\text{prioritizing}} \underbrace{\frac{\sqrt{N_k(s_c)}}{N_k(s_c, d) + 1}}_{\text{exploration}}.$$
(4.8)

Finally, we define $\hat{\pi}_{PUCT}^{1:c} = \pi_K^{1:c}$ based on $\mathcal{E}_{1:c}$ as an estimate of the optimal *individualized* dosing policy in the training setting (using Eqs. (4.5)+(4.8)), and $\hat{q}_{\pi_{PUCT}^c} = q_K$ as an esti-

mate of the associated expected long term return based on $\mathcal{E}_{1:c}$. For individualized dose recommendations in a clinical TDM setting, we again use $\pi^* = \arg \max \hat{q}_{\pi_{\text{PUCT}}^{1:c}}$, i.e., $\epsilon_c = 0$ in Eq. (4.8). See Figure 4.3 for an illustration, a pseudo code is provided in Algorithm 2, Table C.1 summarizes the relevant notation, and further details are given in Section C.6.

Algorithm 2 DA-RL guided dosing

Sample particles to get ensemble \mathcal{E}_0 from the prior $p_{\Theta}(\cdot; \theta^{\mathrm{TV}}(\mathrm{cov}), \Omega)$ Get s_0 based on covariates and baseline measurement Choose optimal dose $d_1^* = \underset{d \in \mathcal{D}}{\operatorname{arg\,max}} \hat{q}_{\pi_{\mathrm{UCT}}}(s_0, d)$ for c = 1 : C do $\mathcal{E}_{1:c} \leftarrow$ update ensemble $\mathcal{E}_{1:c-1}$ by assimilating data $y_c \qquad \triangleright$ DA part: Algorithm 1 $s_c \leftarrow$ posterior expectation under ensemble $\mathcal{E}_{1:c}$ $\hat{q}_{\pi_{\mathrm{PUCT}}^{1:c}} \leftarrow$ MCTS with PUCT using the ensemble $\mathcal{E}_{1:c}$ Choose optimal dose $d_{c+1}^* = \underset{d \in \mathcal{D}}{\operatorname{arg\,max}} \hat{q}_{\pi_{\mathrm{PUCT}}^{1:c}}(s_c, d)$ end for

4.4 Application to manage neutropenia in 3-weekly paclitaxel treatment

We compared our proposed approaches with existing approaches for MIPD based on simulated TDM data in paclitaxel-based chemotherapy. The design was chosen to correspond to the CEPAC-TDM study [52]: neutrophil counts at day 0 and 15 of each cycle were simulated for virtual patients employing the PK/PD model for paclitaxel-induced cumulative neutropenia (BME model, Figure 2.2) [90] described in Section 2.1.3. We also provide a comparison with a sampling design that better matches the typical nadir time of paclitaxel, i.e., considering the second sampling time point at day 12, see Section C.2. The simulation study was performed in MATLAB R2017b/2018b. The generated virtual patient populations for training and testing (N = 1000) were sampled based on the reported covariate ranges in the CEPAC-TDM study [52]. In the offline approaches, the second neutrophil measurement (at day 15 of the cycle) is used to infer the grade of neutropenia (according to the CTCAE [63], see also Figure 2.1).

For the standard dosing approach, we employed the rules applied in the CEPAC-TDM study arm A, i.e., $200 \text{ mg/m}^2 \text{ BSA}$, and a 20 % dose reduction if grade 4 neutropenia was observed [52] (Section 2.1.2). For MAP-guided dosing, a utility function based on a hypothetical survival probability across the different neutropenia grades was investigated in the literature [84].

In MAP-guided dosing, the sensitivities for gradients used in the MATLAB solver fmincon were computed using the Toolbox AMICI [177, 152]. To save computational time, we only optimized over the next cycle (rather than over all remaining cycles) as is also typically done in the literature. In DA-guided dosing, we solved the one-dimensional optimization problem Eq. (4.7) using the fminbnd function in MATLAB (golden section search and parabolic interpolation). Each objective function evaluation corresponds to M model simulations for the corresponding cycle, therefore, we chose a rather small ensemble size M = 100. Note that a larger ensemble size could be chosen for the DA step, while subsequently solving the optimization problem only for a subset of the ensemble.

In RL-guided dosing, we employed MCTS to exploit the characteristics of an episodic task (six treatment cycles) instead of Q-learning, see also Section C.5. For the considered patient state representation Eq. (4.1), we obtained in total L = 32 covariate classes COV_1, \ldots, COV_L (2 genders × 4 age classes × 4 baseline neutrophil count classes), per covariate class we have 19531 possible grade combinations for the 6 cycles, thus leading to a dimension of the discrete state space of $|\mathcal{S}| = 624992$. As we do not need to make a dose decision after the last cycle we can exclude the leaves of the tree (grade of the last cycle), reducing the total number of states to $|\mathcal{S}| = 124992$. The discrete dose steps of 5 mg/m^2 BSA were chosen within the range of given doses in the CEPAC-TDM study $d_{\min} = 60 \text{ mg/m}^2$ BSA and $d_{\max} = 250 \text{ mg/m}^2$ BSA leading to $|\mathcal{D}| = 39$. We chose a discount factor for future rewards: $\gamma = 0.5$, see Section C.5.2. Note that this implies that the current grade of neutropenia was higher weighted than future grades. Yet, γ is sufficiently large to factor in the impact of the current dose choice on the grade of neutropenia in future cycles. The exploration-exploitation parameter ϵ_c was chosen cycle-varying, since the expected return changes over time (cycles), due to the intermediate rewards. Based on Hoeffding's inequality for random variables that can take values in the interval [a, b] this gives,

$$\epsilon(c) = c_{UCT} \cdot \sqrt{\sum_{k=1}^{C-c} \gamma^{k-1} \cdot (b_k - a_k)^2}, \qquad (4.9)$$

with respect to our chosen reward function $a_k = -2$ and $b_k = 1$ for all k. We chose $c_{UCT} = 3$. The choice of all tuning parameters/reward function was further investigated in Section C.5.2.

In the DA-RL guided dosing approach, the q values are further individualized whenever new patient-specific data become available based on a tree search that focuses on relevant doses through the prior probabilities, see Figure C.17.

We focused only on paclitaxel dosing; we did not take into account drop-outs, dose reductions due to non-hematological toxicities, adherence, and comedication. The occurrence of grade 4 neutropenia, therefore, differed between our simplified simulation study and the clinical study (as might be expected), see Section C.1. This should be taken into account when interpreting the results.

4.4.1 Novel individualized dosing strategies decreased the occurrence of grade 4 and grade 0 neutropenia compared to existing approaches

Figure 4.4 shows the predicted neutrophil concentrations—median & 90 % CI—over six cycles of three weeks each. Successful neutrophil-guided dosing should result in nadir concentrations within the target range (grades 1–3, between black horizontal lines). In all cycles, PK-guided dosing prevented the nadir concentrations (90 % CI) to drop as low as for the standard dosing (Figure 4.4 A). However, PK-guided dosing also increased the occurrence of grade 0 at the nadir (Figure 4.5). Choosing as sampling time point day 12, which better reflects the typical nadir time, has not substantially improved the results, see Figure C.3 and C.4.

RL-guided dosing controlled the neutrophil concentration well across the cycles (Figure 4.4 B) and the distribution of nadir concentrations over the whole population was increasingly concentrated within the target range (panel F). The occurrence of grade 0 and 4 neutropenia was substantially reduced compared to standard and PK-guided dosing (Figure 4.5). For MAP-guided dosing, the occurrence of grade 4 neutropenia increased over the cycles (Figure 4.5), showing the typical cumulative trend of neutropenia [90], despite the inclusion of TDM data. In contrast, DA steadily guided nadir concentrations into the target range (Figure 4.4 D and F), thereby substantially decreasing the variance, i.e., the variability in outcome. The occurrence of grade 0 and 4 evaluated at the nadir was reduced considerably in later cycles

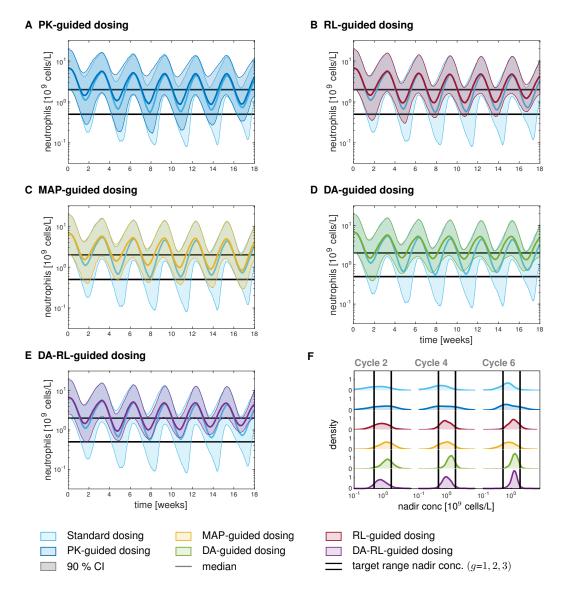


Figure 4.4: Comparison of different dosing policies for paclitaxel dosing. Comparison of the 90 % confidence intervals (CIs) and median of the neutrophil concentration for the test virtual population (N = 1000) using (A) PK-guided dosing (B) RL-guided dosing (C) MAP-guided dosing, (D) DA-guided dosing, and (E) DA-RL-guided dosing, each in comparison to the standard dosing (BSA-based dosing). PK-guided dosing is the only approach that also takes into account exposure ($T_{Cdrug} \ge 0.05 \, \mu mol/L$). (F) Comparison of the distributions of model-predicted nadir concentrations (smooth by kernel density estimation) for the test virtual population at cycles 2, 4, and 6. The black solid lines show the target range of neutropenia grades 1,2, and 3. DA-RL-guided dosing best directs the nadir concentrations into the target range.

60

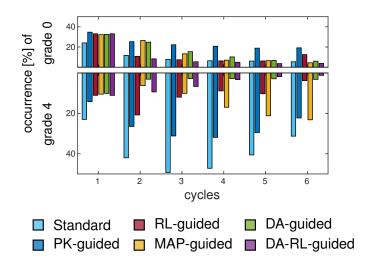


Figure 4.5: Occurrence of grade 0 and grade 4 for the different dosing policies. The percentage is based on a test virtual population (N = 1000) and six cycles (inferred from the model predicted nadir concentration shown in Figure 4.4). The y-axes are scaled according to the different weighting of grade 0 and grade 4 neutropenia to allow to evaluate the total bar length. The shorter the total bar length the better. Additional analysis is provided in Figure C.18.

(Figure 4.5), suggesting that individualized uncertainty quantification played a crucial role in reducing the variability in outcome. Integrating individualized uncertainties and considering the model state of the patient in the RL approach (DA-RL-guided dosing) also moved nadir concentrations into the target range and clearly decreased the variance (Figure 4.4 B+F). The slight differences between DA and DA-RL (Figure 4.5) might be related to the difference in weighting grades 0 and 4 in the respective reward functions (Eq. (4.7) vs. Eq. (4.3)). Note that the different weighting of neutropenia grades 0 and 4 introduces a 'skewness' towards higher neutrophil concentrations. For additional comparisons, see Figure C.18.

In summary, individualized uncertainties as in DA- and DA-RL-guided dosing seemed to be crucial in bringing nadir concentrations into the target range and reducing the variability of the outcome, thus achieving the goal of therapy individualization. For this specific example, both approaches showed comparable results, but DA-RL has the greater potential for long-term optimization in a delayed feedback environment as well as integrating multiple endpoints.

4.4.2 Identification of relevant covariates via investigating the expected long-term return in RL

A key object in RL is the expected long-term return or action-value function $q_{\pi}(s, d)$, see Eq. (2.25), which is the expected long-term return and quantifies the benefit of administering a dose in a given patient state. It plays a critical role in RL as it summarizes and stores the information from accumulated experience. We demonstrate that it contains important information to identify relevant covariates to individualize dosing.

Figure 4.6 A shows the estimated action-value function for RL-guided dosing stratified for the covariates, sex, age, and baseline neutrophil counts ANC_0 (covariate classes are shown in the legend) for the first cycle dose selection. ANC_0 was found to be by far the most important

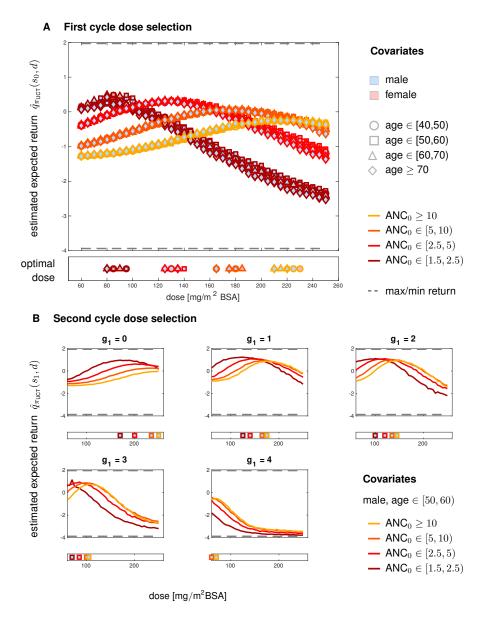


Figure 4.6: Expected long-term return across the dose range for dose selection. (A) across the considered covariate combinations for the dose selection in cycle 1. The symbols plotted below the x-axis show the optimal dose for the corresponding covariate class (i.e., the arg max of the plotted line). (B) for fixed sex and age class (here, male between 50 and 60 years) with different pre-treatment neutrophil values ANC₀ and observed neutropenia grades in cycle 1, i.e., g_1 . The optimal dose for the second cycle depends on the neutropenia grade of the previous cycle and the pre-treatment neutrophil count ANC₀ in [10⁹cells/L]. The grey dashed line shows the maximum and minimum possible return from the first cycle (A) and the second cycle (B) onwards, with $\gamma = 0.5$. The covariate classes were chosen based on the CEPAC-TDM study population: inclusion criteria for the CEPAC-TDM study were ANC₀ > 1.5 $\cdot 10^9$ cells/L; the typical baseline count for male was ANC₀ = 6.48 $\cdot 10^9$ cells/L (arm B). The median age was 63 years ranging from 51 to 74 years (5th and 95th percentile of the population in arm B), see [95, 52].

characteristic for the RL-based dose selection at therapy start. Differences in age and sex played only minor roles. For comparison, the first cycle dose selection in the PK-guided algorithm is only based on sex and age. The steepness of the curves gives an idea about the robustness of the dose selection. For the second dose selection, the grade of neutropenia in the first cycle (g_1) has the largest impact, while larger ANC₀ led to larger optimal doses (Figure 4.6 B). To illustrate the dose selection in RL, we extracted a similar decision tree to the one developed by Joerger et al. [11], see Figure C.9.

Similar investigations are not straightforward for MAP- or DA-guided dosing as no means is provided to investigate dose recommendations for an entire population; these approaches optimize doses for a single patient.

4.5 Discussion

In this chapter, we proposed three promising MIPD approaches employing DA and/or RL that substantially reduced the number of (virtual) patients in life-threatening grade 4 and grade 0 neutropenia, a surrogate marker for efficacy of the anticancer treatment. We have shown that DA and RL techniques can be seamlessly integrated and combined with existing NLME and data analysis frameworks for a more holistic approach to MIPD. Our study demonstrates that the incorporation of individualized uncertainties (as in DA) is favorable over state-of-the-art online algorithms such as MAP-guided dosing. RL provides a flexible framework to account for the uncertainty and delayed effects associated with a dose selection. DA as a means to quantify posterior uncertainty naturally fits into this framework to efficiently further individualize the dose recommendations based on more detailed patient information compared to the reduced state representation used in RL. The integrated DA-RL framework allows not only to consider prior knowledge from clinical studies but also to improve and individualize the model and the dosing policy simultaneously during the course of treatment by integrating patient-specific TDM data. Thus, the combination provides an efficient and meaningful alternative to solely DA-guided dosing, as it allocates computational resources between online and offline and the RL part provides an additional layer of learning to the model (in form of the expected long-term return) that can be used to gain deeper insights into important covariates for the dose selection. Therefore showing that RL approaches can be well interpreted in clinically relevant terms, e.g., highlighting the role of ANC_0 values.

We focused on one specific aspect of the larger decision making problem in oncology: the dosing of a given chemotherapeutic agent. However, RL has also been found to be beneficial for choosing the optimal drugs for first- and second-line treatment [36]. Also, RL-guided dosing in oncology has been proposed before [38], however, only considering the mean tumor diameter. Since only a marker for efficacy was considered this led to a one-sided dosing scheme and resulted in very high optimal doses. The authors, therefore, introduced action-derived rewards, i.e., penalties on high doses. In contrast, neutrophil-guided dosing considers toxicity and efficacy (link to median survival) simultaneously. We limited the application example to neutropenia alone to guide the dosing. Ideally, however, dosing decisions should also include other adverse effects (e.g., peripheral neuropathy [178]), tumor response or long-term outcomes (e.g., overall or progression-free survival) [95], and other concomitant medication (anticancer combination agents, e.g., carboplatin, supportive medication, e.g., G-CSF and other patientspecific co-medications). Notably, RL easily extends to multiple adverse/beneficial effects and co-medication and is especially suited for time-delayed feedback environments [37, 129], as typical in many diseases. Unlike current less complex MIDT, the decision tree of RL is not straightforward to navigate or remember, therefore, an application in clinics would require

the development of easy-to-use software or dashboards, as e.g. for infliximab [172], however, no online computing time is involved for dose selection.

Using MCTS with UCT, we employed an RL framework that exploits the possibility to simulate until the end of therapy and evaluate the return. Consequently, it requires fewer approximations as temporal difference approaches (e.g., Q-learning, used in [38]) that avoid computation of the return via a decomposition (Bellman equation Eq. (2.28)). For comparison, we also applied model-based Q-learning (Q-planning), see Section C.5.3, however, we found that for the specific example MCTS outperformed Q-planning. Yet the decomposition in Q-planning could be advantageous for long-term therapies that cover a large time span. Exploration via UCT allows to systematically sample from the dose range (as opposed to an ϵ -greedy strategy) and allows to include additional information, e.g., uncertainties or prior information (as in predictor+UCT (PUCT)), see also Section C.5.1. This becomes key when combined with direct RL based on real-world patient data, see e.g. [179, 180], which would allow to compensate for a potential model bias. At the end of a patient's therapy, the observed return can be evaluated and used to update the expected return $\hat{q}_{\hat{\pi}}$. Thus, moving from a model-based estimate towards an estimate based on real-world data. This update would even be possible if the physician did not follow the dose recommendation (off-policy learning) and could be implemented across clinics, as it could be done locally without exchanging patient data. Thus, the presented approach builds a basis for continuous learning post-approval, which has the potential to substantially improve patient care, including patient subgroups underrepresented in clinical studies.

Possible future work. The proposed combined DA-RL framework serves as a valuable foundation, to be extended in the future with respect to the considered state/action space, the reward function, and the integration of observational data. In this study, a simplified setting for the RL framework was chosen in terms of the state and action space. A categorical state space was considered with covariate classes as well as neutropenia grades instead of the intrinsically continuous underlying covariates or neutrophil concentration. Continuous action/state spaces require the use of function approximations, e.g., deep neural networks (Deep RL). The possible scope of actions was restricted to fixed dose steps at fixed times. This could be extended to accommodate varying time steps and a more continuous dose space while taking into account the practicability of the design space in terms of dosage form and clinical workflow. An important future aspect to investigate are outliers in the TDM data that potentially have serious consequences for subsequent decision making. There exist robust approaches for parameter estimation in a maximum likelihood context that employ heavier tailed distributions in the error model [181], which should also be investigated for RL and DA-RL-guided dosing.

The choice of the reward function is, in general, a key aspect of the proposed approach. Even if it is clear which adverse events to avoid from a therapeutic standpoint, it is not obvious what the associated numerical reward should be, e.g., grade 4 compared to grade 0 neutropenia. Thus, it is crucial to look into different reward functions within a model-based framework to understand the effect of the chosen numerical values on the resulting dosing policy, see Figure C.13 for the considered setting. RL builds a quantitative framework that allows linking short-term markers with long-term goals, which is lacking in current stand-alone models for the single aspects. RL can be also seen as a means to learn an adaptive reward function for short term markers, e.g., neutropenia with respect to long-term outcomes. Adaptive learning of the reward function could also contribute to the challenge to identify meaningful biomarkers for drugs with a delayed response. Due to the limited time frame of clinical studies, the relationships are often not sufficiently explored and a real-world setting (phase IV/post-approval) would allow to collect more data related to long-term outcomes.

Well-informed and efficient MIPD bears huge potential in drug development as well as in clinical practice as it could (i) increase response rates in clinical studies [8], (ii) facilitate recruitment by relaxing exclusion criteria [7], (iii) enable continuous learning post-approval and thus improve treatment outcomes in the long-term.

5 Continuous learning across patients

In the previous chapters, we have explored how PK/PD models can be used to generate differentiated predictions of the therapy time course and how they can support informed clinical decision-making. The PK/PD models themselves have always been assumed—as commonly done- to be correct and appropriate for the target patient population, i.e., it is assumed that the model sufficiently well represents the drug-patient-disease system, that the variability in outcome is adequately described and that the prior study population (used to develop the model) is representative of the target patient population (to which the model will be *applied*). These assumptions, however, often do not correspond to clinical reality. A certain model misspecification or population shift can be expected due to the limited amount of data the models were built on: data from clinical trials involving only a limited number of patients, selected according to strict inclusion/exclusion criteria within a restricted time frame, or data from different hospitals/study centers [15, 39]. Therefore, models underlying MIPD will inevitably be confronted with deviating data in clinical routine, e.g., center-related differences [182], differences in pathophysiology [183] or differences related to the patient population (patients with comorbidities, comedications, or with special characteristics, e.g., morbidly obese, pregnant, or unusual genotypes) [15, 39, 184, 185, 186]. In this 'imperfect model scenario', the benefits of MIPD approaches may not be clear. It is therefore prudent to improve and adapt a model as clinical routine data on the observed patient population is obtained.

For a given drug-disease-patient-system, there are often numerous models available from literature, but based on different patient populations, e.g., for warfarin therapy [187, 188], vancomycin [183, 189], or ciclosporin [190]. In addition, adjustments to the model used in a MIPD framework were necessitated after treatment of the first patient cohort [191] or in retrospect [192, 90]. As an illustrative example, we focus on models for paclitaxel-induced neutropenia, which build the basis for neutrophil-guided MIPD to individualize chemotherapy dosing as discussed in the previous chapters [75, 73]. Since the publication of the goldstandard model for neutropenia [91], it has been used as a starting point for the development of many model variants, which differ not only in the parameter estimates [67, 94, 9, 11] but also structurally [193, 194, 195, 90], e.g., to account for cumulative neutropenia [90] (see Section 2.1.3).

The challenge to choose between competing models developed in different clinical settings is often approached via model averaging or model selection approaches [189]. In model averaging all candidate models are used, weighting the model predictions with the patient-specific TDM data. In contrast, in model selection, a single model is selected based on a retrospective external evaluation of independent data collected previously in the intended setting (from the same hospital and patient population) [182] and prospective fit-for-purpose verification [183]. None of the approaches, however, does integrate the new data collected during the application of MIPD into the initial models underlying MIPD. In other words: the initial model itself remains unchanged and is not improved with data, which limits its predictive performance in the long run. In this regard, continuous learning approaches based on an ever-growing amount of data have enormous potential to improve the predictive capabilities of MIPD in clinical practice. The problem of transferability is a well-known and studied problem in ML literature, and is often called lifelong learning, continual learning [196, 197], transfer learning [198, 199], or domain adaptation [200]. Contrary to typical ML applications, however, patient data may not be accessible across different hospitals or institutions. This may restrict current approaches based on the pooling of data, which require access to individual TDM data of all patients [201]. Therefore, approaches for model learning are needed that are based on summary information of the data that is extracted locally and can then be shared. In the previous chapter, direct RL, i.e., updating the model-based estimate of the expected long-term return q with observational data, was discussed as a possibility to correct a potential model bias in a dosing policy. It would nevertheless actually be preferable to correct a possible model bias at the level of the model in order to also allow for improved Bayesian forecasting of the therapeutic outcome. Correcting a potential model bias or following a population shift on the level of the model parameters is the focus of this chapter, based on [CM3].

We propose an approach that builds on a sequential hierarchical Bayesian framework for continuous learning. Essentially, the underlying prior model for MIPD is improved on the level of the population parameters as new data from the target patient population are collected. In this way, an individual patient's therapy benefits from every previously treated patient. Importantly, the approach separates the inference of the individual model parameters during a patient's therapy (as described in Chapter 3) from the update of the population parameters across patients by exchanging information via a sample representation of the posterior of the parameters instead of the patient data itself. The proposed approach is based on ideas from Bayesian integration of meta-analyses [202, 203]. First, we demonstrate how a model bias or population shift could affect MIPD in an *in silico* trial setting in terms of misspecified population parameters and structural misspecifications. For the continuous learning framework, we focus in the present thesis on how to correct a model bias on the level of the structural model parameters, i.e., updating the typical or variability parameter values. In particular, we also discuss aspects regarding sampling designs for TDM, which play an important role in model learning. The proposed approach aims at bridging the gap between population analyses in academia or industry and informed individualized dosing in clinical practice, and hence may help to increase the applicability of MIPD approaches in everyday healthcare use.

5.1 Simulation study framework to investigate model bias

In the subsequent simulation study, we considered a normal distribution for p_{Θ} in Eq. (2.3), which can be typically derived via transformation, e.g., log-transformation in case of the log-normal distribution. As MIPD approach we used DA-guided dosing, as presented in detail in the previous chapters (Section 4.2.2).

Paclitaxel-induced neutropenia models We investigate the paclitaxel-induced neutropenia models considered in the context of the CEPAC-TDM study, see Table 2.1. The initial model (hereafter gold-standard [11]) builds on the structure of the gold-standard model for chemotherapy-induced neutropenia [91] with parameter values estimated based on a pooled data set of two prior studies [9, 204] including patients with ovarian cancer, NSCLC, and patients with various solid tumors [11]. Paclitaxel was given either as monotherapy or in combination with carboplatin. In the CEPAC-TDM study, only NSCLC patients were included and paclitaxel was given in combination with carboplatin or cisplatin over six treatment cycles. It was observed that the initial model (gold-standard) overestimated the neutrophil concentration at later cycles since the model does not take into account cumulative neutropenia [90], see Figure 2.3 D. The parameters were re-estimated (hereafter gold-standard R) based on the CEPAC-TDM data, and finally, the structure was modified to account for bone marrow exhaustion, see Figure 2.2. Here, we focus our analyses on the more challenging PD models, while we considered the PK model to be given with parameter values inferred previously based on the CEPAC-TDM study data [90], see Section 2.1.3. In Table D.1 we list additional models proposed in the literature for paclitaxel-induced neutropenia, which illustrates the challenge of choosing a suitable model for MIPD in practice.

Model bias scenarios Model bias denotes a summary term for different specific biases originating from, e.g., a limited number of patients the model was built on, misspecified distributions of model parameters, or differences in analytical methods [15]. Given the described model components (Section 2.1.3), a model bias can result from a misspecified structural model Eqs. (2.1)+(2.2), prior parameter distribution Eq. (2.3), and/or likelihood Eq. (2.4). In the following, we consider two types of model biases:

- Structural bias. A bias in the structural model, e.g., due to the manifestation of phenomena in the target patient population that have not been observed in the prior clinical studies. To study structural bias in the context of paclitaxel-induced neutropenia, we used the BME model [90] (Table 2.1 right column) to generate TDM data, while we used the gold-standard model [11] (Table 2.1 left column) in MIPD. The latter lacks the structural feature of cumulative neutropenia over multiple cycles.
- *Parameter bias.* A bias in the prior parameter distribution. This might include the distributional assumption (normal, lognormal, etc.) as well as the estimated parameter values for a given distribution. Here, we only focus on the latter, e.g., we assume that the type of distribution is the same, but its parameters differ. To study parameter bias, we used the gold-standard R model [95] (Table 2.1 middle column) to generate TDM data, while we used the gold-standard model [11] (Table 2.1 left column) in MIPD. Both rely on the same structural model; the parameter values of the former were re-estimated to the CEPAC-TDM data.

For reference, we compared the performance of MIPD in the presence of structural or parameter bias to (i) the MIPD based on an unbiased model (*unbiased model* scenario) as in Chapter 4, and (ii) the standard dosing [95] (Section 2.1.2).

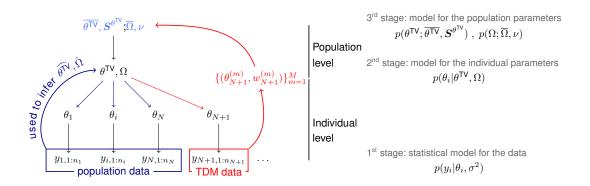


Figure 5.1: Hierarchical Bayesian model framework with separation between the inference on the individual level and the inference on the population level. The standard population analyses typically used to build the prior knowledge based on population estimates $\widehat{\theta^{TV}}$, $\widehat{\Omega}$ is shown in dark blue. To update the prior population estimates, the population parameters are seen as random variables with parametric probability distributions parametrized with hyperparameters $\overline{\theta^{TV}}$, $S^{\theta^{TV}}$ and $\psi = (\nu - n_{\Omega} - 1)\overline{\Omega}, \nu$. A sample representation of the individual posterior is used to update the hyperparameters of the population parameter distributions (red arrow). On the right the corresponding probability distributions are given for the different levels of the hierarchical model.

TDM sampling scenarios The effect of a potentially misspecified prior model on MIPD is dependent on the amount of available TDM data per patient to correct for this bias. Therefore, we considered different TDM sampling schemes:

- 1. *sparse sampling*: neutrophil measurements at day 1 and day 15 of each cycle (sampling design of CEPAC-TDM study).
- 2. intermediate sampling: weekly neutrophil measurements (as in [9]).
- 3. rich sampling: neutrophil measurements are taken every third day.

While the first two sampling schemes correspond to current clinical settings, the third mimics the prospective growing availability of point-of-care devices (e.g., HemoCue[®] WBC Diff for measuring neutrophil counts [30] as discussed in Section 2.1.1), foreseeing richer patient monitoring data to improve and update models.

5.2 Hierarchical modeling

To continuously update and learn population parameters, we considered in addition hyper priors on the population parameters of the NLME models in Section 2.1.3. The hierarchical structure of fully Bayesian population models thus comprises three stages [205, 156], see Figure 5.1:

- 1. Statistical model for the TDM data, given by Eq. (2.4), which describes the deviations between the individual model predictions and the observational data. Often, an additive normally distributed error is assumed (potentially on a log-scale).
- 2. The distributional assumption for IIV, Eq. (2.3) describes parameter differences between individuals, possibly including a covariate model. Often a lognormal distribution is chosen.

3. Distributional assumptions for population parameters (so called hyper priors)

$$p(\theta^{\mathrm{TV}}), p(\Omega)$$

which describe the uncertainty in the population parameters.

Hyper priors Population analyses are typically performed in a frequentist NLME setting, reporting maximum likelihood estimates (MLEs) of the population parameters jointly with their relative standard errors (RSEs) or CV. This leaves the problem of how to determine suitable hyper prior distributions for the population parameters.

The normal-inverse Wishart distribution has been proposed previously as hyper prior for the population parameters ($\theta^{\text{TV}}, \Omega$) [206]. The distribution of the typical values is normal with mean $\overline{\theta^{\text{TV}}}$ and variance $\mathbf{S}^{\theta^{\text{TV}}}$. Using the asymptotic normality of the posterior (Section 2.2.1), we chose the prior $p(\theta^{\text{TV}}|y^{\text{pop}})$ to be normally distributed with mean $\overline{\theta^{\text{TV}}}$ identical to the MLE $\widehat{\theta^{\text{TV}}}$ and variance $\mathbf{S}^{\theta^{\text{TV}}}$ identical to the squared standard error $(\text{SE}_{\theta^{\text{TV}}})^2$, see also [100]. Since we considered the log-transformed parameters, the standard error were transformed appropriately, e.g., using a Taylor expansion at the MLE or a sampling based approach.

The inter-individual variability matrix Ω was assumed to be inverse-Wishart distributed with parameters Ψ , and degrees of freedom ν , i.e., $\mathcal{IW}(\Psi,\nu)$. We considered Ω to be diagonal and chose Ψ such that the population estimate equals the mean $\Psi/(\nu - n_{\Omega} - 1)$, i.e., $\Psi = (\nu - n_{\Omega} - 1)\hat{\Omega}$. The distributions of the typical and variability values were assumed to be independent.

5.3 Novel approach to learn population parameters across patients

To learn and improve a model across patients the information provided by the patient-specific TDM data needs to be included into the hierarchical model. In mathematical terms, we are interested in the marginal posterior (to ease notation, we drop the indices):

$$p(\theta^{\mathrm{TV}}, \Omega|y) \propto \int p(\theta, \theta^{\mathrm{TV}}, \Omega|y) d\theta$$
 (5.1)

with the joint posterior

$$p(\theta, \theta^{\mathrm{TV}}, \Omega | y) \propto p(y | \theta, \sigma^2) p(\theta | \theta^{\mathrm{TV}}, \Omega) p(\theta^{\mathrm{TV}}) p(\Omega)$$
(5.2)

determined from a full hierarchical Bayesian procedure. A sample approximation to the joint posterior Eq. (5.2) allows for a straightforward approximation of the marginal in Eq. (5.1). For our particle filter based inference, this would require to augment the particle state and parameter space by the population parameters (θ^{TV}, Ω), i.e., two additional parameters per structural model parameter. This, however, is computationally expensive and thus, limits real-time inference during the patient's therapy. Also, direct access to the individual patient data would be needed for updating the population parameters, which limits learning across different hospitals.

Therefore, we propose a two-level sequential hierarchical Bayesian approach based on previous approaches for Bayesian inference for meta-analyses [202, 203] to learn across TDM patients, $i = 1, \ldots, N_{\text{TDM}}$. Importantly, this approach does not change the inference on the individual level (see Algorithm 3 for pseudo code):

1. Individual level: Estimate individual parameters of the ith patient

$$p(\theta_i|y_i) \propto p(y_i|\theta_i, \sigma^2) p(\theta_i|\hat{\theta}^{\mathrm{T}\hat{\mathrm{V}}}, \hat{\Omega})$$

e.g., using a PF, SIR or MCMC approach, as described in Section 2.2. We employed in our analysis a particle filter ('DA' in the pseudo-code), as it was shown to be best suited for our requirements (Chapter 3). This gives rise to a sample representation of the posterior, $\{(\theta_i^{(m)}, w_i^{(m)}), m = 1, \ldots, M\}$, summarizing the information provided by the data of the *i*th patient, see also Eq. (2.14).

2. Population level: Update population parameters by sampling iteratively from the joint posterior $p(\theta, \theta^{\text{TV}}, \Omega|y)$ via a Metropolis-Hastings-within-Gibbs sampling scheme [100, 202], i.e., sampling from the full conditionals:

$$p(\theta^{\mathrm{TV}}|\theta_i, \Omega_i, y_{1:i}) \propto p(\theta_i|\theta^{\mathrm{TV}}, \Omega) p(\theta^{\mathrm{TV}}|\theta_{i-1}, \Omega_{i-1}, y_{1:i-1})$$
(5.3)
= $p(\theta_i|\theta^{\mathrm{TV}}, \Omega) p(\theta^{\mathrm{TV}}|\theta_{i-1}, \Omega_{i-1})$

$$p(\Omega|\theta_i, \theta_i^{\mathrm{TV}}, y_{1:i}) \propto p(\theta_i|\theta^{\mathrm{TV}}, \Omega) p(\Omega|\theta_{i-1}, \theta_{i-1}^{\mathrm{TV}}, y_{1:i-1})$$

$$= p(\theta_i|\theta^{\mathrm{TV}}, \Omega) p(\Omega|\theta_{i-1}, \theta_{i-1}^{\mathrm{TV}})$$
(5.4)

$$p(\theta_i|\theta_i^{\text{TV}}, \Omega_i, y_{1:i}) \propto p(y|\theta_i, \sigma^2) p(\theta_i|\theta_i^{\text{TV}}, \Omega_i) \,.$$
(5.5)

Sampling from Eq. (5.3) in iteration l = 1, ..., L corresponds (in our setting) to sampling from a multivariate normal distribution $\mathcal{N}(\mu_i^{\theta^{\mathrm{TV}}(l)}, \Sigma_i^{\theta^{\mathrm{TV}}(l)})$ with parameters

$$\Sigma_i^{\theta^{\mathrm{TV}}(l)} = \left(\left(\mathbf{S}_{i-1}^{\theta^{\mathrm{TV}}} \right)^{-1} + \left(\Omega_i^{(l-1)} \right)^{-1} \right)^{-1}$$
(5.6)

$$\mu_i^{\theta^{\mathrm{TV}}(l)} = \Sigma_i^{\theta^{\mathrm{TV}}} \left(\left(\Omega_i^{(l-1)} \right)^{-1} \theta_i^{(l-1)} + \left(\mathbf{S}_{i-1}^{\theta^{\mathrm{TV}}} \right)^{-1} \overline{\theta_{i-1}^{\mathrm{TV}}} \right), \tag{5.7}$$

and from Eq. (5.4) corresponds to sampling from an inverse-Wishart distribution $\mathcal{IW}(\Sigma_i^{\Omega(l)}, \nu_i^{(l)})$ with parameters

$$\Sigma_{i}^{\Omega(l)} = (\nu_{i} - d - 1)\bar{\Omega}_{i-1} + \left(\theta_{i}^{(l-1)} - \theta_{i}^{\mathrm{TV}(l)}\right) \left(\theta_{i}^{(l-1)} - \theta_{i}^{\mathrm{TV}(l)}\right)^{T}$$
(5.8)

$$\nu_i^{(l)} = \nu_{i-1} + 1. \tag{5.9}$$

Further, sampling from Eq. (5.5) is achieved via a M-H step by using as proposals the posterior samples generated on the individual level $\{\theta_i^{(m)}\}_{m=1}^M$, which are drawn according to weights $w_i^{(m)}$ and accepted with probability

$$\alpha = \frac{p(\theta_i^{*(l)}|\theta_i^{\mathrm{TV}(l)}, \Omega_i^{(l)}) / p(\theta_i^{*}|\overline{\theta_{i-1}^{\mathrm{TV}}}, \overline{\Omega}_{i-1})}{p(\theta_i^{(l-1)}|\theta_i^{\mathrm{TV}(l)}, \Omega_i^{(l)}) / p(\theta_i^{(l-1)}|\overline{\theta_{i-1}^{\mathrm{TV}}}, \overline{\Omega}_{i-1})} \,.$$
(5.10)

Importantly, Eqs. 5.6-5.9 and the acceptance probability in Eq. (5.10) do not contain any patient data y_i . Thus, given the posterior ensemble generated on the individual level, the patient data themselves are not required to perform the population parameter updates. Rather, the patient data y_i implicitly enter via the (weighted) sample representation

of the posterior $p(\theta_i|y_i)$. Finally, to start with a parametric distribution for the next patient a normal and inverse-Wishart distribution is fitted to the MCMC samples (after removing a 'burn-in'), i.e., for the typical values $p(\theta^{\text{TV}}) \approx \mathcal{N}(\overline{\theta_i^{\text{TV}}}, \mathbf{S}_i^{\theta^{\text{TV}}})$ with

$$\overline{\theta_i^{\mathrm{TV}}} = \frac{1}{L} \sum_{l=1}^{L} \theta_i^{\mathrm{TV}(l)}, \quad \mathbf{S}_i^{\theta^{\mathrm{TV}}} = \frac{1}{L-1} \sum_{l=1}^{L} (\theta_i^{\mathrm{TV}(l)} - \overline{\theta_i^{\mathrm{TV}}}) (\theta_i^{\mathrm{TV}(l)} - \overline{\theta_i^{\mathrm{TV}}})^T, \quad (5.11)$$

and for the IIV parameters $p(\Omega) \approx \mathcal{IW}((\nu_i - n_\Omega - 1)\overline{\Omega}_i, \nu_i)$ with

$$\bar{\Omega}_i = \frac{1}{L} \sum_{l=1}^{L} \Omega_i^{(l)}, \quad \nu_i = \nu_{i-1} + 1.$$
(5.12)

For the next patient, the particle ensemble for the individual level is initialized based on the new parameters $\overline{\theta_i^{\text{TV}}}, \overline{\Omega}_i$.

Algorithm 3 Two-level sequential hierarchical Bayesian learning in MIPD

1: Input: $\theta^{\text{TV}}, \text{SE}_{\theta^{\text{TV}}}, \hat{\Omega}, \nu_0, (y_{i,1:n_i} \text{ only for individual level})$ 2: Set hyper prior parameters $\overline{\theta_0^{\text{TV}}} := \widehat{\theta^{\text{TV}}}, \mathbf{S}_0^{\theta^{\text{TV}}} := (\text{SE}_{\widehat{\theta^{\text{TV}}}})^2, \overline{\Omega}_0 := \hat{\Omega}, \nu_0$ 3: for $i = 1 : N_{\text{TDM}}$ do // Individual level 4: initialize particle ensemble $\{\theta_{i0}^{(m)}, x_{i0}^{(m)}, w_{i0}^{(m)}\}_{m=1}^M$ based on $p(\theta | \overline{\theta_{i-1}^{\text{TV}}}, \overline{\Omega}_{i-1})$ 5: for j = 1: n_i do $\{\theta_{ij}^{(m)}, x_{ij}^{(m)}, w_{ij}^{(m)}\}_{m=1}^M \leftarrow \text{DA}(y_{ij}, \{\theta_{ij-1}^{(m)}, x_{ij-1}^{(m)}, w_{ij-1}^{(m)}\}_{m=1}^M)$ 6: 7: \triangleright Alg. 1 end for 8: // Population level 9: initialize Markov chain $\theta_i^{\text{TV}(0)} = \overline{\theta_{i-1}^{\text{TV}}}, \Omega_i^{(0)} = \overline{\Omega}_{i-1}$ and $\theta_i^{(0)}$ sampled from $p(\theta | \overline{\theta_{i-1}^{\text{TV}}}, \overline{\Omega}_{i-1})$ for l = 1 : L do 10: 11: 12:// Gibbs sampling part 13:// Gibbs sampling part draw $\theta_i^{\text{TV}(l)}$ from $p(\theta^{\text{TV}}|\theta_i^{(l-1)}, \Omega_i^{(l-1)}, y_i)$ draw $\Omega_i^{(l)}$ from $p(\Omega|\theta_i^{(l-1)}, \theta^{\text{TV}(l)}, y_i)$ // Metropolis-Hastings part draw proposal $\theta_i^{*(l)}$ from $\{\theta_{in_i}^{(m)}\}_{m=1}^M$ according to $\{w_{in_i}^{(m)}\}_{m=1}^M$ ⊳ Eq. (5.3) 14: ⊳ Eq. (5.4) 15:16: 17:and add rejuvenation 18: Accept proposal with probability α ▷ Eq. (5.10) 19:end for 20: Parametric approximations of hyper priors: 21: $p(\theta^{\mathrm{TV}}) \approx \mathcal{N}(\overline{\theta_{i}^{\mathrm{TV}}}, \mathbf{S}_{i}^{\theta^{\mathrm{TV}}}) \\ p(\Omega) \approx \mathcal{IW}((\nu_{i} - n_{\Omega} - 1)\bar{\Omega}_{i}, \nu_{i})$ ⊳ Eq. (5.11) 22: ▷ Eq. (5.12) 23:24: end for

Continuous learning simulation setting The continuous learning approach was applied to $N_{\text{TDM}} = 100$ virtual patients with available TDM data over six treatment cycles depending on the considered sampling scheme. The continuous learning approach was repeated 10 times to account for statistical variability in the individual patient parameters considered for the

update. The same virtual patients were used in a simulation study using MIPD alone without continuous cross-patient learning (DA-guided dosing) to demonstrate how MIPD could be affected by a model bias. On the individual level, model parameters $(MTT, Slope, ANC_0)^T$ were estimated. We restricted the population updates to 'MTT' and 'Slope', as for 'ANC₀' the baseline method B2 described in [96] is used, i.e., no typical parameter was estimated but the baseline value was used to initialize the (empirical Bayes) prior (see Eq. (2.6)). Besides, we consider a setting that includes γ in the individual level inference as the value differs across the models. In this study, we neither estimated σ on the individual level nor on the population level. However, the values for σ used to generate the TDM data differed from those that the models in the 'imperfect model scenarios' assume. The considered hyper priors, i.e., the distributional assumptions for the population parameters, are summarized in Table 5.1. Since no relative standard errors are available for the gold-standard model [11], the values reported in [9] (one of the two pooled studies) were chosen as conservative choice. The choice of the degrees of freedom ν is generally difficult but was chosen here to balance confidence in the estimated value while still enabling adaptation. The simulation study was performed in MATLAB 2019b.

Table 5.1: Hyper priors for the gold-standard model used in the simulation study.

Parameter	distribution	hyperparameters
TV parameters		
$\log(MTT)$	\mathcal{N}	$\overline{\theta^{\mathrm{TV}}}_0 = \log(2.6)$, $\mathbf{S}_0^{\theta^{\mathrm{TV}}} = 0.0013$
$\log(\text{Slope})$	\mathcal{N}	$\overline{\theta^{\mathrm{TV}}}_0 = \log(141)$, $\mathbf{S}_0^{\theta^{\mathrm{TV}}} = 0.016$
IIV parameters		
$\omega^2_{ m MTT}$	\mathcal{IW}	$\Psi_{\rm MTT} = 0.6561 = (12 - 2 - 1)0.0729, \nu_0 = 12$
$\omega_{ m Slope}^2$	\mathcal{IW}	$\Psi_{\text{Slope}} = 1.8144 = (12 - 2 - 1)0.2016, \nu_0 = 12$

5.4 Application to paclitaxel-induced neutropenia models

5.4.1 Current MIPD approaches may not be beneficial in the presence of model bias

For a performance analysis, we generated TDM data (including RUV) on day 1 & day 15 of each cycle (sparse sampling as in the CEPAC-TDM study) over six treatment cycles. Figure 5.2 illustrates the performance of MIPD with/without model bias in comparison to standard dosing (median and 90% CIs).

The left column illustrates the scenario of parameter bias, i.e., the structural model and the class of prior distributions are identical to the data generating process, but the parameter values of the prior distribution differ. In this case, MIPD performs comparably to the standard dosing (top left), also in terms of occurrence of grade 4 and grade 0 neutropenia (bottom left). For reference, in the corresponding unbiased model scenario, the MIPD approach clearly reduces the occurrence of grade 4 & 0 (bottom and middle panel). It is worth noting that the confidence intervals in all panels show a certain 'skewness' towards higher neutrophil concentrations (lower grade of neutropenia). This is due to the choice of the weighting factors (grade 4 is penalized more strongly than grade 0) in the dose optimization problem Eq. (4.7).

The right column illustrates the more challenging scenario of structural bias, i.e., the structural model differs from the model underlying the data generating process. Of note, in this case, both standard dosing and MIPD perform much worse than in the scenario of parameter bias (bottom panel). In 3 out of 6 cycles, MIPD results in even larger occurrences of grade 4 compared to standard dosing. The gold-standard model underestimates the drug effect on neutrophil concentrations (see Figure 2.3 D) and hence too high doses are selected, especially in presence of cumulative neutropenia. Despite relying on an inappropriate structural model, DA is able to correct this initial bias on the parameter level over the course of a patient's therapy by integrating TDM data, which leads to a decrease in the incidence of grade 4 neutropenia in later cycles. For reference, in the corresponding unbiased model scenario, the MIPD approach clearly and very quickly reduces the occurrence of grade 4 and grade 0 (bottom and middle panel). This scenario corresponds to the setting of Chapter 4, Figure 4.4 D. In comparison to the gold-standard R scenario, the occurrence of grade 4 and grade 0 neutropenia is even further decreased, which might be related to the smaller RUV parameter, see Table 2.1.

In summary, if the underlying model is not consistent with the observational data, MIPD might not be beneficial compared to standard dosing that solely relies on TDM data ('model-free'). As outlined in the introduction, a model bias can be expected, if MIPD is applied in clinical routine, therefore, the top panels might better reflect clinical reality than the middle panels. This is the status-quo of state-of-the-art MIPD approaches. They generally do not exploit the wealth of TDM data used during MIPD to learn and update the underlying models—except for the aforementioned recent approaches based on pooling the data for re-estimating the NLME model [201].

5.4.2 Continuous learning MIPD can counteract parameter bias, but depends on the sampling scheme

The proposed continuous learning MIPD framework is able to adapt the biased parameter distribution over time as TDM patients are observed. Figure 5.3 illustrates the sequential updates of the proposed framework for the posterior distributions of the typical parameters of 'Slope' and 'MTT' across 100 patients for different sampling schemes. For the rich sampling scenario (left), the panel shows how the posterior—95% highest posterior density (HPD) area—evolves over the number of observed patients (displayed after every 5th patient), moving away from the prior estimate (gray star) towards the value used to generate the TDM data (black star). As more patients are observed, uncertainty about the typical 'Slope' and 'MTT' parameters decreases, as indicated by the decreasing size of the HPD area.

Thus, the proposed continuous learning MIPD framework successfully allows to learn the typical values underlying the TDM data from sample representations of the posterior on the individual level. Note that the parameters γ and σ were not estimated, although different values are used to generate the data, which might lead to some deviations. The results for when γ is also included in the individual level inference are shown in Figure D.1.

To what extent the continuous learning MIPD framework is able to counteract parameter bias, however, depends on the sampling scheme, see Figure 5.3 (middle and right panel). For the intermediate sampling scenario (weekly), the posterior distribution moves towards the parameter values that have been used to generate the data. A final parameter bias, however, remains, potentially due to parameter identifiability issues. To assess practical identifiability, we investigated the loglikelihood and logposterior on the individual patient levels, see Figure D.2. To exclude the possibility that the sampling time points of the intermediate scheme (weekly) were chosen unfavorably, we performed an optimal design analysis, see Figure D.3. For the sparse sampling scheme, the TDM data on the individual level are not sufficient to fully remove model bias. Only for the rich sampling, the data

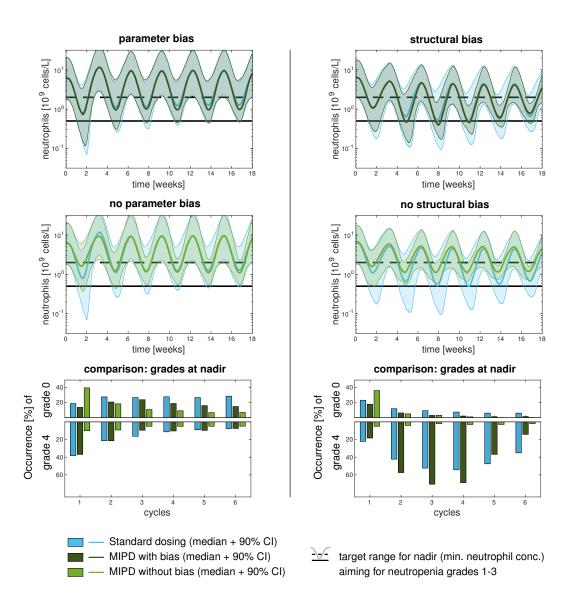


Figure 5.2: MIPD under model bias/population shift. Neutropenia time courses were simulated for $N_{\text{TDM}} = 10 \cdot 100$ virtual patients using MIPD approaches with different underlying models and the standard dosing approach for paclitaxel (200 mg/m^2 with 20% reduction if grade 4 neutropenia was observed in the previous cycle). TDM data were simulated including residual variability on day 1 & day 15 of each cycle (sparse sampling scenario). In the left column, TDM data were generated using the gold-standard R model. The top panel demonstrates the 'imperfect model scenario'; the MIPD approach uses the gold-standard model. The middle panel corresponds to the 'perfect model scenario'; the MIPD approach uses also the gold-standard R model. The model. The median time course is shown along its 90% confidence interval (CI). The bottom panel compares the occurrence of life-threatening grade 4 neutropenia and subtherapeutic grade 0 neutropenia at the nadir for the different dosing strategies. In the right column, the TDM data were generated using the bone marrow exhaustion (BME) model [90]. Note that grade 4 neutropenia is penalized more ($\lambda_4 = 2/3$) compared to grade 0 neutropenia ($\lambda_0 = 1/3$) in Eq. (4.7). This is accounted for in the scale of the bottom panels, which allows to interpret the length of the total bars.

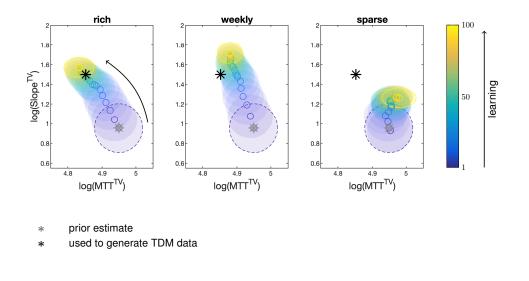


Figure 5.3: Comparison of the sequential updates of the hyper prior for the typical MTT and Slope value for different TDM scenarios. Gray star: prior estimate of the parameters; black star: target parameters, i.e., the ones used to generate the TDM data. Sparse sampling consists of measurements on day 1 & 15 of each cycle, weekly sampling corresponds to an intermediate data situation and for rich sampling it is assumed that neutrophils are monitored every third day. The mean (circle) and 95% highest posterior density (HPD) is shown (shaded ellipse) after every 5th patient.

are sufficiently informative to move away from the (biased) prior estimate towards the data-generating value, resolving the practical unidentifiability.

The sequential updates for the corresponding IIV parameters are displayed in Figure 5.4 for the rich sampling scenario. The IIV parameter for 'MTT' moves from the prior estimate towards zero as no IIV has been estimated in the gold-standard R model; in other words, TDM data were generated with the same parameter value for 'MTT' for all patients. The IIV parameter for 'Slope' initially increased (dark blue) as individually estimated 'Slope' parameters deviated considerably from the biased prior typical 'Slope' parameter. However, as more TDM data were observed and the typical value was increased, also the IIV parameter moved back towards the target value (close to the prior value). Overall, it can be observed that the magnitude of IIV for 'MTT' and 'Slope' is overestimated to a certain extent, which may be due to the increased RUV not being taken into account.

5.4.3 Continuous learning in MIPD has the potential to substantially improve therapy outcome even for structural bias, again depending on the sampling scheme

Finally, we investigated the effects of continuous learning of population parameters on MIPD. We compared the performance of the proposed continuous learning MIPD approach to the MIPD approach without learning (DA-guided dosing) as well as standard dosing across 100 patients. The analysis was repeated 10 times to account for statistical variability in the model parameters used to generate the TDM data. Here, we show only the more challenging scenario of a structural bias. For the parameter bias scenario, we refer to Figure D.4. Figure 5.5 compares the performance of the different approaches. Standard dosing and DA-guided dosing are as in Figure 5.2, but here shown for the intermediate sampling design as continuous

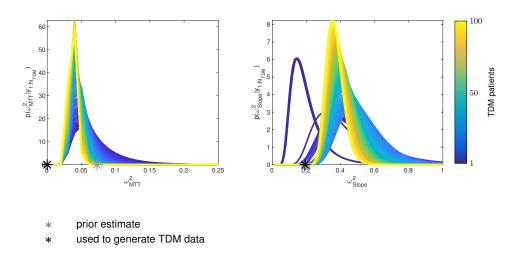


Figure 5.4: Sequential updates of the IIV parameters across patients for the rich sampling scenario.

learning is more effective with more TDM data (as seen above). Here, we also consider uncertainty with respect to the parameter γ .

As already described in Figure 5.2, we found that DA-guided dosing was able to adjust also to some extent for cumulative neutropenia over time (see Figure 5.5 (dark green) for the intermediate sampling scheme). It can be observed that the 'Slope' parameter increased, while parameters 'Circ₀' and γ decreased over the course of the therapy of six cycles, leading to a decrease in the occurrence of grade 4 after cycle 3 and a substantial decrease in outcome variability. Effectively, when considering the data points one at a time, the sequential DA framework allowed to account for changes in the parameters over time. This is potentially a very beneficial property, e.g., to better reflect disease progression and intra-individual variability without explicitly modeling it. While this might be very desirable for MIPD for the individual patient, it could be misleading when learning across patients. When the final parameter estimate (after six cycles) was used to update the population parameter (Slope^{TV}, MTT^{TV} , this introduced a bias for the first cycle of the next patient, resulting in a high occurrence of grade 0 for the first cycle (Figure 5.5 bottom left). Continuous learning was considered across the first 100 patients (blue-green) as well as continued learning of the second 100 patients (yellow) after gaining experience on 100 patients from the target patient population (results are displayed for the same virtual patients, different 100 target patients were considered for the previous experience). It can be observed that the typical 'Slope' parameter increases (green vs. blue-green vs. yellow) as it is continuously learned across patients (initial Slope value at t = 0 in the top right panel). The parameter values are, however, not comparable to the values of the BME model due to the structural differences between the models.

A major improvement can be seen for the continuous learning MIPD approach, which reduced the occurrence of grade 4 substantially across all cycles compared to DA-guided dosing alone as well as standard dosing. This improvement can be already seen for the first 100 patients (blue-green) but is even more pronounced for the second 100 patients. Only the aforementioned time-dependence of the parameters leads to an increased occurrence of grade 0 neutropenia in the first cycle, but this is also (to a lesser extent) visible in the 'perfect model

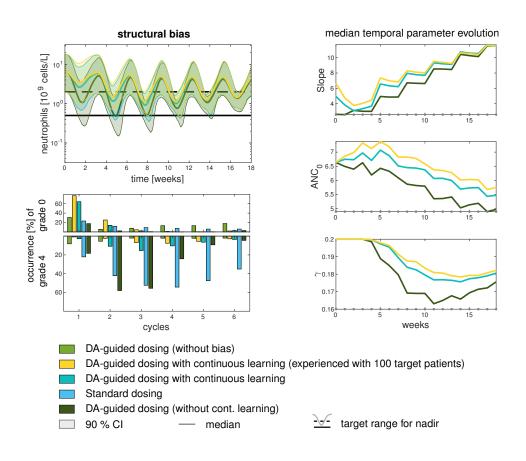


Figure 5.5: Sequential DA allows for temporal parameter changes within the course of a patient's therapy. The TDM data were generated using the BME model and the results are shown for the intermediate sampling scheme (sampling timepoints are indicated as grey dots in the right panels). The median (intra-individual) temporal parameter evolution over the course of a patient's therapy was computed across all virtual patients $(N_{\text{TDM}} = 10 \cdot 100)$.

scenario' Figure 5.2. The results for the rich sampling scenario are comparable (Figure D.6), however, for the sparse sampling scheme, the benefits are not so clear (Figure D.5).

5.5 Discussion

In this chapter, we investigated the transferability of population PK/PD models to different clinical settings, a crucial application hurdle of MIPD in clinical routine. In a relevant application scenario, we first showed that model misspecification might severely impact MIPD, and therefore, models should be adapted to the target patient population. The used MIPD approach, DA-guided dosing, proved to be able to counteract model bias to some extent, but only improved MIPD at later cycles when a certain amount of TDM data was collected. As the used DA approach processed data sequentially, it allowed to account for temporal changes in the parameters and thus adapted the gold-standard model to some extent to cumulative neutropenia. As aforementioned this could be a very valuable property to describe disease progression or time-varying effects, e.g., time-varying drug clearances [207]. We propose a

sequential hierarchical Bayesian approach to update the population parameters using posterior samples as a means to exchange information after every treated patient so that the model better reflects the target patient population. We showed that the approach successfully allowed to learn the underlying population parameters of the PK/PD model used to generate the patient data, however, the results depend very much on the sampling scheme. In addition, we showed that continuous learning has the potential to improve MIPD even in presence of structural bias, again depending on the amount of available TDM data.

Health care data challenge algorithms and mathematical approaches not only with sparsity, noise, and missing data. Crucial aspects are also legal, privacy, and confidentiality issues [208]. These often limit the applicability and implementation of patient data analysis. To continuously improve models and evaluate their performance, however, a large amount of patient data is required. In this work, aspects of practical applicability were considered, for example with respect to data protection. Sample representations of Bayesian posterior distributions are discussed and presented as a means to exchange information instead of patient-specific TDM data. The proposed approach has two levels and allows to learn sequentially over patients without using patient data on the population level. Thus, the patient data themselves do not need to be stored or shared across clinics, which is a big advantage compared to pooling approaches [201]. To account for center related differences, an inter-study variability could be included. Thus, the approach builds a basis to develop more informed models integrating an ever-growing amount of data potentially better reflecting rare covariates. This could be particularly relevant in the rare disease area or in the rapeutic areas where human efficacy trials are unethical [86], which requires that models for MIPD are learned directly in clinical routine.

The selection of the initial model used to start the continuous learning process could be based on a retrospective external evaluation using historical data from the intended patient population [182]. Model selection/model averaging approaches do not adapt/improve the underlying model across patients; the *a priori* forecast remains the same for all patients (based on the covariates). In addition, in their general form, these approaches are implemented in conjunction with MAP estimation which provides potentially biased predictions in the context of nonlinear models (Chapter 3). The proposed DA-guided dosing also naturally extends to model averaging and this extension has been considered (on the individual inference level) previously in the context of Bayesian therapy forecasting [125].

The presented analysis revealed, that an important aspect for practical implementation is to critically assess the quality of the inference on the individual level. The dependence on the sampling design clearly showed that more research is necessary and that caution is needed when updating models based on real-world data, as was demonstrated for time-dependent parameters and the dependence on the amount of TDM data. With the prospect of novel digital health care devices, e.g. point-of-care devices, more frequent monitoring could become clinical reality. This could increase the availability of real-world data, which are currently underutilized [209] but has great potential to improve MIPD as shown in this study.

Possible future work. The current approach is limited to misspecifications or population shifts on the structural model parameter level. An important extension in the future would be to also estimate the RUV parameter σ , as an increased error in measurement precision or reporting can be expected in clinical routine compared to e.g., controlled clinical study settings. Currently, the IIV parameters ω^2 captured the increased RUV of the data to some extent, which, however, also increased the uncertainty on the individual level. Further, covariates observed in the intended study population, which were not included in the population analysis, could be integrated into the model via an assumed covariate model with prior parameter distributions centered at zero. If the covariate is of relevance, the population parameter updates should move the prior away from zero.

An important issue to resolve in future studies is to directly target potential structural model bias. Replacing the typical ODE-based models with SDE-based models would allow to explicitly model the uncertainty in the model dynamics, separating measurement errors from model misspecification [210]. The discussed DA methods for Bayesian forecasting are used in most applications with SDE models, therefore the integration of SDE models is straightforward. With the aforementioned advent of novel digital health care devices that enable more frequent TDM sampling, parameter estimation for SDE-based PK/PD models could become feasible. As discussed in Section 2.2.3, however, care should be taken that biological constraints are preserved. Another means to describe uncertainty in the underlying model could be via physics-informed neural networks, which balance in the cost functional not only the deviation of the neural network output from the data but also the deviation from the (mechanistic) model. Thus, they could bridge the missing or possibly biased physiological or biological knowledge in the model to learn beyond the existing structural model. Especially in the context of sparse data as often present in current MIPD applications physics-informed neural networks have huge potential to improve the predictive capabilities of neural networks ¹.

The approach of a learning model (as coined in [15]) for MIPD could be beneficial not only in clinical practice but also during drug development, where new (clinical) study data are generated sequentially and should be integrated into previously developed models [85]. The proposed approach is an important step towards building the underlying models of MIPD on a growing amount of data and thus make MIPD fit-for-purpose in everyday healthcare use.

¹Results based on a project work by Julia Kirchner on "Physics Informed Deep Neural Networks in Pharmacokinetics" (August 27, 2020) that the author of this dissertation co-supervised jointly with Niklas Hartung, Niels Landwehr, Wilhelm Huisinga and Tobias Scheffer.

6 Outlook

In this thesis, mathematical approaches to advance MIPD were developed and successfully applied in the context of cytotoxic anticancer chemotherapy. The scientific challenges in MIPD with respect to therapy forecasting, clinical decision-making for dose individualization, and continuous learning across patients as identified in Chapter 1 were addressed by integrating, adapting, and combining methodologies from different research fields. Within each chapter, the proposed approaches were discussed and possible extensions that could be the subject of future work were outlined. Finally, in this chapter, we want to broaden the discussion beyond the scientific aspects discussed in this thesis and provide general perspectives on MIPD based on [CM4]. Specifically, we focus on aspects of MIPD that need to be addressed in the future for increasing the implementation of MIPD in everyday healthcare use.

First, there is a lack of uniform terminology in scientific literature to describe MIPD and to differentiate it from alternative approaches towards precision medicine, e.g., pharmacogenomics. Frequently used terms for the same approach (here termed MIPD) comprise Bayesian feedback [211], optimal/adaptive control [168], target concentration intervention [21], Bayesian dashboard [212]. In addition, the exact method used for Bayesian inference is often not outlined in detail. As a result, the literature is unclear and incomprehensible. Harmonization of the terminology could increase visibility, prevent misunderstandings, and eventually accelerate the progress of MIPD. Furthermore, courses and trainings for healthcare professionals are needed to improve the understanding of model predictions and quantitative pharmacology for effective and correct use of MIPD tools in clinical practice. At the same time, software tools need to be user-friendly, e.g., using dashboards as for example for infliximab [172], but also reliable and easy-to-interpret. The present work has made clear that current MAP-based predictions are not well interpretable and has demonstrated how more informative, and reliable forecasting of the therapy outcome could be achieved with full Bayesian DA approaches. The presented DA approaches could be integrated into user-friendly MIPD software tools for predicting the therapy outcome under uncertainty to support the decision-making of the treating physician. A transparent and informative visualization of model predictions based on a-posteriori probabilities contributes to a differentiated understanding of the capabilities, but also the limitations of the underlying model. We have also shown how patient covariates can be identified that drive dose selection for RL-based dose recommendations. These considerations are key to building confidence in mathematical models and MIPD among clinicians. Since it is ultimately the treating clinician who makes the treatment decision, an optimal implementation of MIPD in practice involves interaction between the patient, the

clinician, and the decision-support tool, i.e., clinician/patient-in-the-loop [213]. Additionally, practical implementation of MIPD in clinical routine requires an infrastructure that allows for the seamless and efficient integration of available patient data from different sources, e.g., bioanalytical laboratories, bedside monitors, or point-of-care devices, into MIPD software tools. This environment is not provided in most health care systems and is currently often rather limited to university hospitals or research institutions. In this thesis, RL was proposed as a flexible framework that allows integrating multiple sources of data and models. We also discussed how real-world data could be used for continuous learning across patients to update and improve the models while they are applied to patient data. These aspects could help to build fit-for-purpose MIPD tools that integrate various sources of information and improve the underlying models with new data. However, these MIPD tools are subject to regulatory and reimbursement questions including the registration of MIPD tools as medical devices [214, 215] and the costs of software licenses. A key aspect is also the awareness that not all drugs might benefit from MIPD. Therefore, the selection of the right drugs is decisive and should be based on the therapeutic index, the amount of PK/PD variability, biomarker availability for TDM, the risk of morbidity/mortality associated with the disease state, and also the discrepancy between the patient populations from clinical trials and the real-world [209]. These considerations will be crucial for demonstrating the clinical benefit of MIPD in the long-run across different therapeutic areas.

Finally, for patients to benefit from MIPD in the future, MIPD must be early integrated into drug development, be required by regulatory authorities, and be accepted by healthcare professionals.

Conclusion. The present dissertation contributed mathematical and algorithmic approaches to improve the accuracy, reliability, and informative value of MIPD while taking into account new trends in health care as well as aspects of practical applicability. We demonstrated the importance of uncertainty quantification, the integration of uncertainties into dose selection, and continuous learning of the underlying models within MIPD. The proposed approaches represent an important advancement over the current MIPD approaches, promising decisive benefits for future individualized therapies and contribute to making MIPD practical for everyday healthcare use.

A Appendix related to the background

A.1 PK-guided dosing

For the sake of completeness, we repeat here the algorithm by Joerger et al. [11] (called *PK-guided dosing*), see Figure A.1, as we used it for comparison throughout the thesis. In the *PK-guided dosing*, the dose of the first cycle is determined based on the patient's age and sex. For subsequent cycles, the dose is adjusted according to exposure ($T_{C>0.05}$ time during which the drug concentration is above $0.05 \,\mu$ M in hours [h]) and neutropenia grade observed in the previous cycle (inferred from observation at day 15). Thus, the algorithm is not completely offline, since the exposure measure is inferred online using the PK model. The PD-based adaptations, which were the focus in this thesis, however, were performed offline. In our setting, we did not take PK samples but considered the true $T_{C>0.05}$ to be known.

A.2 PK/PD Models used in this thesis

A.2.1 Docetaxel PK model

As PK model for docetaxel, a published three compartment model with first-order elimination was employed [82], which includes the covariates AAG, AGE, BSA, and albumin (ALB). The individual clearance (CL_i) is computed via

 $CL_{i} = BSA_{i} \cdot (CL^{TV} + \theta_{CL-AAG} \cdot AAG_{i} + \theta_{CL-AGE} \cdot AGE_{i} + \theta_{CL-ALB} \cdot ALB^{TV}) \cdot (1 - \theta_{CL-HEP12} \cdot HEP12),$

where we used as typical albumin value $ALB^{TV} = 41 g/L$ and set HEP12=0 (i.e., no elevated hepatic enzymes (HEP12)). The parameter estimates were taken from Bruno et al. [82], see Table A.1, and the system of ODEs for the PK model is given by

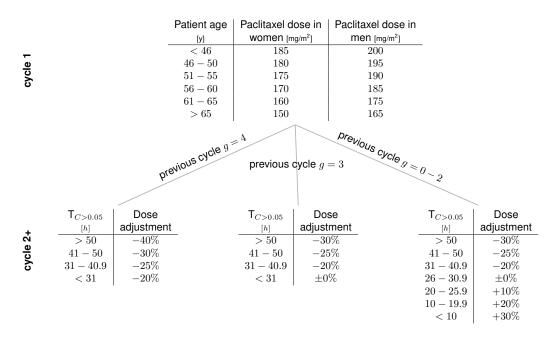


Figure A.1: Model-informed dosing table of the PK-guided dosing algorithm adopted from [11].

$$\frac{d\text{Cent}}{dt} = d(t) - k_{10}\text{Cent} + k_{21}\text{Per1} - k_{12}\text{Cent} + k_{31}\text{Per2} - k_{13}\text{Cent}, \quad \text{Cent}(0) = 0$$

$$\frac{d\text{Per1}}{dt} = k_{12}\text{Cent} - k_{21}\text{Per1}, \quad \text{Per1}(0) = 0$$

$$\frac{d\text{Per2}}{dt} = k_{13}\text{Cent} - k_{31}\text{Per2}, \quad \text{Per2}(0) = 0,$$

with 'Cent' referring to the central compartment, and 'Per 1', 'Per 2' to the first and second peripheral compartment, respectively.

Table A.1: Pharmacokinetic parameter estimates for docetaxel taken from [82].

Structural submodel		bmodel	Covariate submodel	
V	8.31	[L]	$\theta_{\mathrm{CL-AAG}}$ -3.55	
CL	22.1	[L/h]	$ heta_{ ext{CL-AGE}}$ -0.095	
k_{10}	$\mathrm{CL/V}$	[1/h]	$\theta_{\text{CL-ALB}} = 0.225$	
k_{12}	1.07	[1/h]		
k_{21}	1.74	[1/h]		
k_{13}	1.28	[1/h]		
k_{31}	0.0787	[1/h]		

A.2.2 Paclitaxel PK model

We used the PK model in [11], see also Figure 2.2 (left part) for a schematic representation. In the simulation studies in Chapters 3-5, the re-estimated parameters by [90] were used, see

Table A.2. The system of ODEs describing the rate of change of the amount of paclitaxel in $[\mu mol]$ is given by

$$\frac{d\text{Cent}}{dt} = d(t) - \frac{\text{VM}_{\text{EL}} \cdot C_1}{\text{KM}_{\text{EL}} + C_1} + k_{21}\text{Per1} - \frac{\text{VM}_{\text{TR}} \cdot C_1}{\text{KM}_{\text{TR}} + C_1} + k_{31}\text{Per2} - k_{13}\text{Cent}, \quad \text{Cent}(0) = 0$$

$$\frac{d\text{Per1}}{dt} = \frac{\text{VM}_{\text{TR}} \cdot C_1}{\text{KM}_{\text{TR}} + C_1} - k_{21}\text{Per1}, \quad \text{Per1}(0) = 0$$

$$\frac{d\text{Per2}}{d\text{Per2}} = k_{10}C_{10} + k_{10}R_{10}R_{10} + R_{10}R_{10}R_{10}R_{10} + R_{10}R_$$

$$\frac{dt}{dt} = k_{13} \operatorname{Cent} - k_{31} \operatorname{Per2},$$

$$\operatorname{Per2}(0) = 0$$

where Cent refers to the central compartment; Per1, Per2 to the first and second peripheral compartment, respectively; $C_1(t) = \text{Cent}/V_1$ refers to the total paclitaxel concentration in plasma; $k_{13} = Q/V_1$ and $k_{31} = Q/V_3$, where V_3 refers to the volume of Per2; d(t) is the dosing input.

Table A.2: Pharmacokinetic parameter estimates of the previously published PK model [11] for the anticancer drug paclitaxel (re-estimated parameter values taken from [90]).

struct	ural subm	odel	statisti	cal submodel IIV
V_1	10.8	[L]		
V_3	301	[L]	$\omega_{V_3}^2$	0.1639
$\mathrm{KM}_{\mathrm{EL}}$	0.667	[µM]		
$\rm VM_{EL,pop}$	35.9	$[\mu mol/h]$	$\omega^2_{ m VM_{EL}}$	0.0253
$\rm KM_{\rm TR}$	1.44	[µM]	$\omega_{\rm KM_{TR}}^2$	0.3885
$\rm VM_{TR}$	175	$[\mu mol/h]$	$\omega_{ m VM_{TR}}^2$	0.077
k_{21}	1.12	[1/h]	$\omega_{k_{21}}^2$	0.008
Q	16.8	[1/h]	$\omega_Q^{2^{-1}}$	0.1660
covari	ate submo	odel	statisti	cal submodel IOV
$\theta_{\rm VM_{EL}\text{-}BSA}$	1.14		$\pi^{2}_{V_{1}}$	0.1391
$\theta_{\rm VM_{EL}-SEX}$	1.07		$\pi_{\rm VM_{EL}}^2$	0.0231
$\theta_{\mathrm{VM}_{\mathrm{EL}} ext{-AGE}}$	-0.447		statistic	cal submodel RUV
$\theta_{\rm VM_{EL}\text{-}BILI}$	-0.0942		σ^2	0.0317

A.2.3 Docetaxel-induced neutropenia model

Docetaxel-induced neutropenia was previously described using the structure of the goldstandard model for neutropenia [91] with parameter estimates given in Table A.3 as provided in [67].

Structu	ral subm	nodel	Statisti	ical submodel
Circ ₀	5.22	$[10^9 \text{cells/L}]$	$\omega_{\rm Circ_0}^2$	0.0606
MTT	84.2	[h]	$\omega^2_{ m MTT}$	0.0194
Slope	15.6	$[L/\mu mol]$	$\omega_{\mathrm{Slope}}^2$	0.122
γ	0.145	[]	ω_{γ}^2	0.0223
Covaria	ate subm	odel	σ^2	0.180
$\theta_{\text{Circ}_0-\text{AAG}\leq 1.34}$	0.175			
$\theta_{\mathrm{Circ}_0\text{-}\mathrm{AAG}>1.34}$	0.495			
$\theta_{ m Circ_0-SEX}$	-0.121			
$ heta_{ m Circ_0-PERF}$	0.131			
$ heta_{ m Circ_0-PC}$	-0.147			
$ heta_{ ext{Slope-AAG}}$	-0.351			

Table A.3: Parameter estimates for the gold-standard model for docetaxel taken from [67]

B Appendix related to Chapter 3

B.1 Algorithmic details of MAP estimation

We employed gradient descent algorithms to solve the optimization problem Eq. (3.3). These algorithms can often be improved by providing the gradient and the Hessian of the objective function $J(\theta) = -\log p(\theta|y_{1:n})$. The gradient for this specific problem is given by

$$\begin{split} \frac{\partial J(\theta)}{\partial \theta_l} &= -\sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial h_j(\theta)}{\partial \theta_l} \\ &+ \frac{1}{\theta_l} + \frac{(\log(\theta_l) - \log(\theta_l^{\mathrm{TV}}))}{\omega_l^2} \cdot \frac{1}{\theta_l} \,, \end{split}$$

and the Hessian for $l \neq m$

$$\frac{\partial^2 J(\theta)}{\partial \theta_l \partial \theta_m} = -\left(\sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial^2 h_j(\theta)}{\partial \theta_l \partial \theta_m} - \frac{1}{\sigma^2} \frac{\partial h_j(\theta)}{\partial \theta_l} \frac{\partial h_j(\theta)}{\partial \theta_m}\right)$$

and

$$\begin{split} \frac{\partial^2 J(\theta)}{\partial \theta_l^2} &= -\left(\sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial^2 h_j(\theta)}{\partial \theta_l^2} - \frac{1}{\sigma^2} \frac{\partial h_j(\theta)}{\partial \theta_l} \frac{\partial h_j(\theta)}{\partial \theta_l}\right) \\ &+ \frac{1}{\theta_l^2} \cdot \left[\frac{1}{w_l^2} \Big(1 - \log(\theta_l) + \log(\theta_l^{\mathrm{TV}})\Big) - 1\right]. \end{split}$$

Note that $\frac{\partial h_i(\theta)}{\partial \theta_l} = S_l^h$ are the output sensitivities, which are given by

$$S_l^h = \frac{\partial h(x,\theta)}{\partial x} S_l^x + \frac{\partial h(x,\theta)}{\partial \theta_l} \,,$$

using the sensitivities of the states

$$\frac{\partial S_l^x}{\partial t} = \frac{\partial f(x,\theta)}{\partial x} S_l^x + \frac{\partial f(x,\theta)}{\partial \theta_l} , \qquad S_l^x(0) = \frac{\partial x_0(\theta)}{\partial \theta_l}$$

For the computation of the state sensitivities the extended system of ODEs needs to be solved

$$\dot{x} = f(x,\theta), \qquad x(0) = x_0(\theta)$$
$$\dot{S}_l^x = \frac{\partial f(x,\theta)}{\partial x} S_l^x + \frac{\partial f(x,\theta)}{\partial \theta_l}, \qquad S_l^x(0) = \frac{\partial x_0(\theta)}{\partial \theta_l}.$$

Alternatively, the gradient could be computed via adjoint sensitivity analysis which is more efficient for models with a large number of states and parameters [216]. Since the Hessian matrix requires the computation of the second-order sensitivities $\frac{\partial^2 h_i(\theta)}{\partial \theta_i \partial \theta_m}$, which is computationally expensive, often the (expected) FIM is used as approximation

$$\mathcal{I}_{lm}(\theta) = -\sum_{i=1}^{n} \frac{1}{\sigma^2} \cdot \frac{\partial h_i(\theta)}{\partial \theta_l} \cdot \frac{\partial h_i(\theta)}{\partial \theta_m}$$

and

$$\mathcal{I}_{ll}(\theta) = -\left(\sum_{i=1}^{n} \frac{1}{\sigma^2} \cdot \frac{\partial h_i(\theta)}{\partial \theta_l} \cdot \frac{\partial h_i(\theta)}{\partial \theta_l} + \frac{1}{\omega_l^2 \theta_l^2} \cdot \left(\log(\theta_l^{\mathrm{TV}}) - \log(\theta_l) + 1\right)\right).$$

B.2 Additional analyses for the simulation studies

B.2.1 Single cycle study docetaxel

Reference posterior. Since the true posterior distribution is analytically intractable (Section 2.2.1), we employ as reference solution the SIR algorithm with a large number of samples $(M = 10^6)$, as the algorithm is exact for $M \to \infty$. This reference was validated via comparison to the posterior derived by the MCMC algorithm using also $M = 10^6$ samples with a burn-in of 100 samples, see Figure B.1.

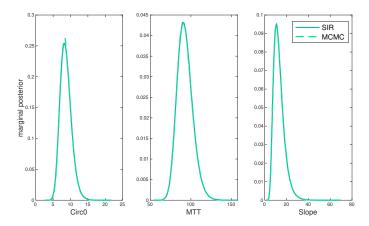


Figure B.1: Comparison of reference posterior. The reference posterior was derived by the SIR algorithm and by the MCMC algorithm using $M = 10^6$ samples.

Potentially biased MAP-based predictions. In Chapter 3, it was described that the MAP estimate does not correctly transform under a nonlinear mapping. As pharmacometric PK/PD models are often nonlinear, this is a major drawback for decision support in MIPD. Figure B.2 (A) shows the posterior of the drug effect parameter 'Slope' on the x-axis and the

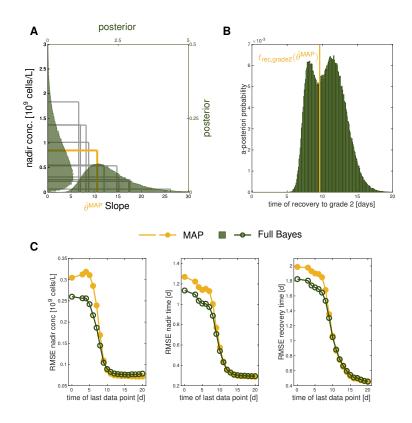


Figure B.2: Illustration of the unfavorable properties of MAP estimation with regard to reliable decision support. (A) The MAP estimate (yellow line) does not correctly transform under a nonlinear mapping to the most probable nadir concentration (based on [103, Figure 5.2]). The posterior of the parameter 'Slope' is depicted for an exemplary patient after four data points $y_{1:4}$ were observed on the x-axis and the corresponding a-posteriori probability of the nadir concentration on the y-axis (same scenario as in Figure 3.4). (B) The mode is not preserved under transformation. Here, shown for the time to recovery to grade 2. The same scenario was considered as for part A. (C) Root mean squared error (RMSE) of selected statistics. Comparison of the accuracy of the computed statistics c_{nadir}, t_{nadir} and t_{rec0} based on MAP estimation and full Bayesian inference (SIR using $M = 10^3$ samples). The RMSE was computed across the whole considered virtual population N = 100.

a-posteriori probability of the nadir concentration on the y-axis. The yellow line shows the MAP estimate for the parameter 'Slope' and links to the MAP-predicted nadir concentration, which clearly does not correspond to the mode of the a-posteriori probability distribution of the nadir concentration (green histogram). In addition, it is shown how some randomly chosen samples transform to the nadir concentrations. Note, however, that the nadir concentration does not only depend on the 'Slope' but also on the other model parameters. Figure B.2 (B) demonstrates the same observation for a different quantity of interest, the time of recovery to grade 2, which shows a bimodal posterior distribution. The MAP-based time to recovery to grade 2 could be misleading as it does not reflect the bimodality of the posterior. Further, we considered as a statistical measure of accuracy the root mean squared error (RMSE) between the model-predicted outcome $T_i(\mathcal{E}_n)$ given data $y_{1:n}$ for individual *i* and the reference outcome

 T^{ref} (for which the data were simulated)

$$\text{RMSE}(T)_n = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_i(\mathcal{E}_n) - T_i^{\text{ref}})^2}$$

Figure B.2 (C) shows the prediction accuracy of the point-estimates of MAP-estimation and of the fully Bayesian approach over time (computed for the N = 100 virtual patients). As more data points are taken into account the RMSE decreases for both estimators. At the beginning of the cycle, the fully Bayesian approach shows increased accuracy across all considered quantities of interest, which is especially relevant for the prediction of the nadir concentration and the nadir time as typically the nadir is around day 9 for docetaxel.

Normal approximation underestimates uncertainty. The delta method leads to a similar underestimation of the uncertainty as the simulation-based approach (NAP sim), see Figure B.3. In addition, it is not straightforward to propagate the uncertainty to quantities of interest (therefore not displayed).

One generally suggested option to overcome the underestimation of the uncertainty is to use the Student's t distribution instead of the normal distribution [154]. We have used quantiles of the Student's t distribution with $\nu = 4$ degrees of freedom (NAP δ t). The CrIs show an increased width, but now overestimate the uncertainties regarding subtherapeutic areas (grade 0), see Figure B.4. This is also not acceptable as underdosing is highly undesirable in oncology.

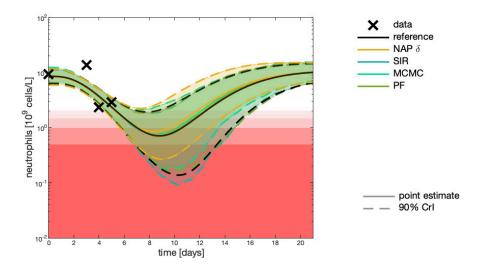


Figure B.3: Propagating uncertainties in the normal approximation (NAP) approach using the delta method instead of the simulation based approach.

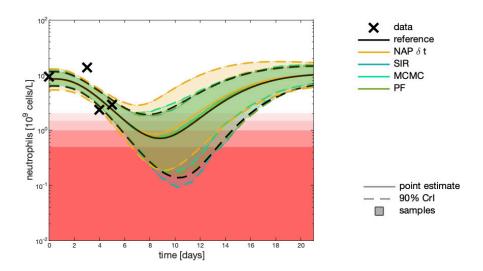


Figure B.4: Propagating uncertainties in the normal approximation (NAP) approach using the delta method with Student t quantiles instead of normal quantiles.

Improved acceptance rate with proposed adaptive Metropolis-Hastings The acceptance rate is an important statistic in MCMC diagnostics to assess the trade-off between exploring the space and efficiently moving the chain. For the M-H algorithm with fixed proposal variance Ω the acceptance rate decreased to very low levels as more TDM data are observed. The proposed adaptive M-H sampler counteracts this decrease and achieves acceptance rates within the suggested range (black horizontal lines) [100], see also Section 2.2.1.

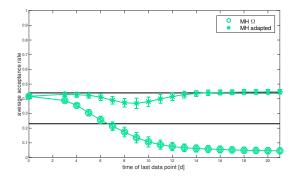


Figure B.5: Acceptance rate of the Metropolis-Hastings algorithm with fixed proposal variance (MH Ω) and with adapted proposal (MH adapted). The black lines mark the area of a good acceptance rate (0.23,0.4).

B.2.2 Multiple cycle study paclitaxel

The data for the simulation study were generated using the BME model prediction including IOV, see Eq. (2.5). The cycle-specific parameters θ_c^{IOV} in Eq. (3.1) are estimated based on the data observed in cycle c, $y_{1:n_c} = (y_1, \ldots, y_{n_c})^T$. The size of the parameter vector that needs to be estimated will, therefore, grow with every occasion, in the case the whole data are processed in a batch (as in MAP, SIR, MCMC).

MAP-estimation including IOV. Assuming independence between the IIV and IOV, the optimization problem for the MAP estimation is given by

$$\begin{split} \hat{\theta}_{n}^{\mathrm{MAP}} &= \operatorname*{arg\,min}_{\theta^{\mathrm{HV}},\theta^{\mathrm{IOV}}} \; \frac{1}{2} \Big(\sum_{c=1}^{C} \sum_{j=1}^{n_{c}} \frac{\left(y_{j} - h_{j}(\theta)\right)^{2}}{\sigma^{2}} \\ &+ 2 \sum_{k=1}^{n_{\theta}^{\mathrm{IIV}}} \log(\theta_{k}^{\mathrm{IIV}}) + \sum_{k=1}^{n_{\theta}^{\mathrm{IIV}}} \frac{\left(\log(\theta_{k}^{\mathrm{IIV}}) - \log(\theta_{k}^{\mathrm{TV}}(\mathrm{cov}))\right)^{2}}{\omega_{k}^{2}} \\ &+ 2 \sum_{c=1}^{C} \sum_{k=1}^{n_{\theta}^{\mathrm{IOV}}} \log(\theta_{k,c}^{\mathrm{IOV}}) + \sum_{c=1}^{C} \sum_{k=1}^{n_{\theta}^{\mathrm{IOV}}} \frac{\left(\log(\theta_{k,c}^{\mathrm{IOV}}) - \log(\theta_{k}^{\mathrm{TV}}(\mathrm{cov}))\right)^{2}}{\pi_{k}^{2}} \Big) \end{split}$$

for the IIV and IOV model $\theta_{k,c} = \theta_k^{TV} \cdot e^{\eta_k + \kappa_{k,c}}$, with $\eta_k \sim \mathcal{N}(0, w_k^2)$ and $\kappa_{k,c} \sim \mathcal{N}(0, \pi_k^2)$, for an additive normal residual error model $y_j = h_j + \epsilon_j$, with $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$ and for data observed up to time point t_n , i.e., $n = \sum_c n_c$, with n_c the number of observations made in cycle c. The gradient with respect to the IIV parameters is given by

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_l^{\text{IIV}}} &= -\sum_{j=1}^n \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial h_j(\theta)}{\partial \theta_l^{\text{IIV}}} \\ &+ \frac{1}{\theta_l^{\text{IIV}}} + \frac{(\log(\theta_l^{\text{IIV}}) - \log(\theta_l^{\text{TV}}))}{\omega_l^2} \cdot \frac{1}{\theta_l^{\text{IIV}}}, \end{aligned}$$

and with respect to the IOV parameters by

$$\begin{split} \frac{\partial J(\theta)}{\partial \theta_{l,c}^{\text{IOV}}} &= -\sum_{j=1}^{n} \frac{(y_j - h_j(\theta))}{\sigma^2} \cdot \frac{\partial h_j(\theta)}{\partial \theta_l^{\text{IOV}}} \\ &+ \frac{1}{\theta_{l,c}^{\text{IOV}}} + \frac{(\log(\theta_l^{\text{IOV}}) - \log(\theta_l^{\text{TV}}))}{\omega_l^2} \cdot \frac{1}{\theta_{l,c}^{\text{IOV}}} \,. \end{split}$$

Comparison of the methods for the multiple cycle study. Figure B.6 shows a comparison of the different methods in forecasting the third cycle. The scenario corresponds to the situation presented in Figure 3.5. All fully Bayesian methods provide almost overlapping CrIs as well as point estimates (median). However, the MAP-based forecasted trajectory deviates significantly from the point estimates (median) of the fully Bayesian methods. The corresponding MAP-based predicted neutropenia grade would be grade 2, although the most probable grade of the reference is grade 3 neutropenia. The a-posteriori probabilities of key quantities of interest related to the neutropenia time course are comparable across the full Bayesian approaches and well match the reference.

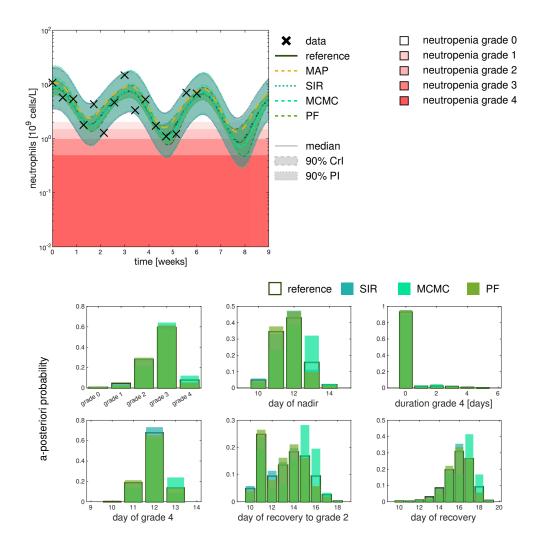


Figure B.6: Comparison of low-cost approximations regarding forecast accuracy. The third cycle is forecasted in case the standard dose is given. For the full Bayesian approaches the median is shown as point estimate, along with the 90% credible interval (CrI) and the 90% prediction interval (PI).

In Figure B.7 the posterior approximations are compared to the reference on the level of the parameters. Also, we can observe the deviation of the posterior from the prior for parameters 'Slope' and 'Circ₀'. As we do not consider PK samples the knowledge gain about the PK parameters is limited. All approximations show good agreement with the reference and the MAP estimate is located at the mode of the posterior on the level of the parameters.

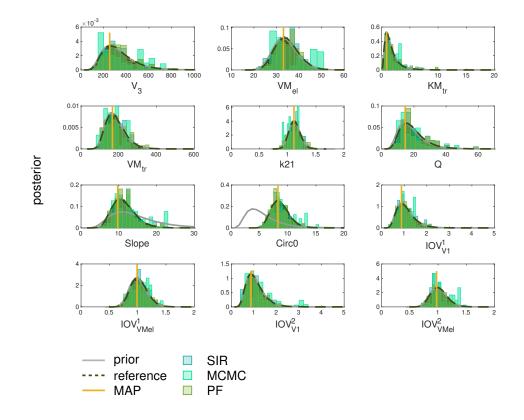


Figure B.7: Comparison of low-cost approximations regarding posterior inference. Approximation of the posterior of the parameters for the scenario considered in Figure B.6

C Appendix related to Chapter 4

Table C.1: Notation related to the reinforcement learning approaches used in RL-guided and DA-RL-guided dosing.

Common notation	
π	generic policy
q_{π}	expected return given generic policy π
Notation specific to MCTS+UCT	
π_k	policy in training phase
q_k	sample approximation of \hat{q}_{π_k} in training phase
$\hat{\pi}_{\mathrm{UCT}} := \pi_K$	policy after training phase (incl. exploration)
$\hat{q}_{\pi_{\mathrm{UCT}}} := q_K$	sample approx. at the end of training phase
$\pi^* = \arg \max \hat{q}_{\pi_{\mathrm{UCT}}}$	RL-guided dosing policy (clinical setting, no exploration)
Notation specific to MCTS+PUCT	
$q_{\pi_0} := \hat{q}_{\pi_{\text{UCT}}}$	prior estimated return
$\pi_k^{1:c}$	policy in training phase using ensemble $\mathcal{E}_{1:c}$
$q_k^{1:c}$	sample approximation in training phase using ensemble $\mathcal{E}_{1:c}$
$\begin{array}{l} q_{\pi_0} \coloneqq \hat{q}_{\pi_{\mathrm{UCT}}} \\ \pi_k^{1:c} \\ q_k^{1:c} \\ \hat{\pi}_{\mathrm{PUCT}}^{1:c} \coloneqq \pi_K^{1:c} \end{array}$	policy after training phase using ensemble $\mathcal{E}_{1:c}$ (incl. exploration)
$\hat{q}_{\pi_{1:c}} := q_{K}^{1:c}$	sample approx. at the end of training phase using ensemble $\mathcal{E}_{1:c}$
$\pi^* = \arg \max \hat{q}_{\pi_{\text{PUCT}}^{1:c}}$	DA-RL-guided dosing policy (clinical setting, no exploration)

C.1 Comparison with reported CEPAC-TDM study outcomes

In the simulation study, we followed the design of the CEPAC-TDM study. To put the simulation results into perspective, we compared the simulated occurrence of grade 4 neutropenia (based on simulated observations on day 15 including RUV) with the observed occurrence in the CEPAC-TDM study for the standard dosing (arm A) and the PK-guided dosing algorithm (arm B), see Figure C.1. We observed that we overpredict the occurrence of grade 4 neutropenia for standard dosing (left panel). This may be attributed to the fact that for the standard dosing in arm A, the dose was also decreased if non-hematological toxicities occurred (see also comment at the end of Section 4.4). Since the applied model did only allow to simulate neutropenia we could not take further aspects into account. For the PK-guided dosing algorithm (right panel), the simulation results were well aligned with

the observed results in the CEPAC-TDM study. The occurrence of grade 4 neutropenia was comparable across all cycles. Additional characteristics of the clinical study that we did not take into account in our simulation study are drop-outs, adherence to the dosing instructions and comedication (e.g., therapeutic G-CSF) [95].

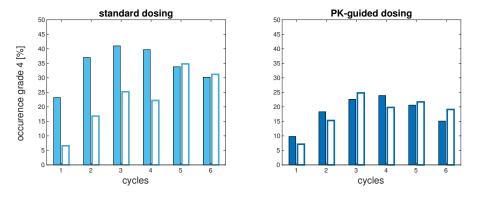


Figure C.1: Comparison of predicted grade 4 neutropenia with observed occurrence in the CEPAC-TDM study. Blue bars show the results from the simulation study (based on day 15 observation) and white bars show the results from the CEPAC-TDM study. The results for the CEPAC-TDM study were retrieved from [95].

C.2 Sampling time points to infer neutropenia grade

The observation time points of neutrophil concentrations were chosen following the CEPAC-TDM study design [52]: the day of the dose administration (day 1) as well as day 15 of each cycle. However, in the evaluation of the dosing algorithm the average model predicted nadir time (based on the gold-standard model) was found to be on day 11.5 [11]. Therefore, we investigated also day 12 as alternative sampling time point.

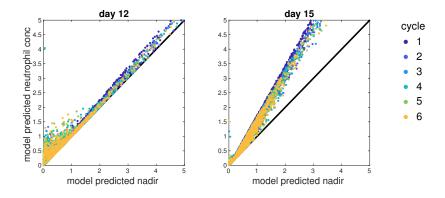


Figure C.2: Comparison of the model predicted nadir concentration compared to the model predicted neutrophil concentration at days 12 and 15. The standard dosing was used for simulation of the neutropenia time courses for 1000 virtual patients.

In Figure C.2, we examined the correlation between the model predicted nadir (based on the BME model) and the simulated neutrophil concentration at day 12 and 15. For larger nadir concentrations (nadir > $1 \cdot 10^9$ cells/L) the neutrophil concentrations at day 15 clearly overpredict the true nadir, i.e., underpredicts the severity of neutropenia. For small nadir concentrations (nadir $\leq 1 \cdot 10^9$ cells/L) the correlation between model predicted nadir and model predicted neutrophil concentration at day 15 seems to be better. This information could be relevant for future studies and demonstrates the importance of optimal sampling time points and the benefit of a model-informed analysis.

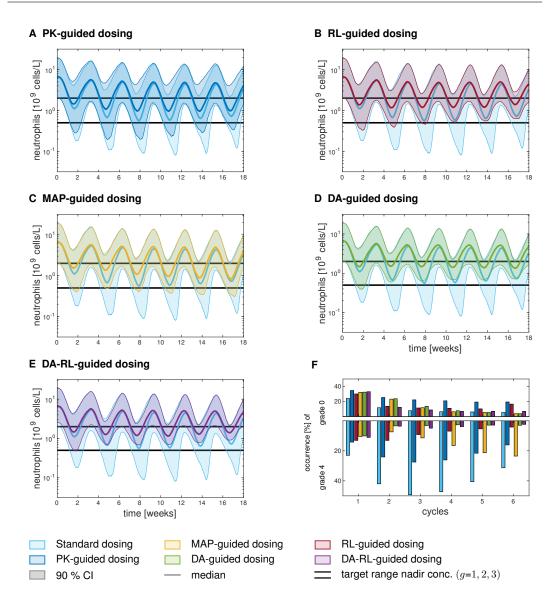


Figure C.3: Comparison of different dosing policies for paclitaxel dosing based on observations at day 12. Comparison of the 90 % confidence intervals and median of the neutrophil concentration for the test virtual population (N = 1000) using (A) the standard dosing and PK-guided dosing (B) RL-guided dosing, (C) MAP-guided dosing, (D) DA-guided dosing and (E) DA-RL-guided dosing. (F) Occurrence of grade 0 and grade 4 across the different dosing policies for the test population over the six cycles.

We also compared the effect of the sampling time point on the PK-guided dosing algorithm by applying the algorithm to the test virtual population, see Figure C.4. For this the neutropenia grade of the previous cycle was inferred either based on simulated neutrophil measurements at day 12 or day 15 in the previous cycle (including RUV). The occurrence of neutropenia grade 4 (evaluated based on model predicted nadir) was slightly higher if the previous cycle grade was inferred from the measurement at day 15 compared to day 12. Thus, the sampling time point affets the PK-guided dosing algorithm and a sampling time point around day 12 is advantageous.

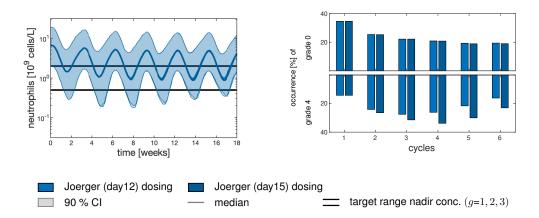


Figure C.4: Comparison of the results when PK-guided dosing was based on the neutrophil measurement at day 12 or day 15.

C.3 Details on MAP-guided dosing

For MAP-guided dosing two possibilities are discussed: either minimizing the deviation to a target concentration or maximizing a utilty function. For our specific setting, the 90% CI of the neutropenia time courses of the virtual test population reached lower neutrophil concentrations for the target concentration intervention compared to the utility function, see Figure C.5, and increased the occurrence of grade 4 neutropenia across all cycles, see Table C.2. This result, however, cannot be generalized and depends on the choice of target/utility.

Table C.2: Occurrence of grade 4 neutropenia across cycles for MAP-guided dosing.

cycle	1	2	3	4	5	6
utility function	10.5%	7.4%	11.4%	16.5%	21.3%	23.5%
target deviation	22%	15.3%	22.3%	29.9%	34.6%	41.5%

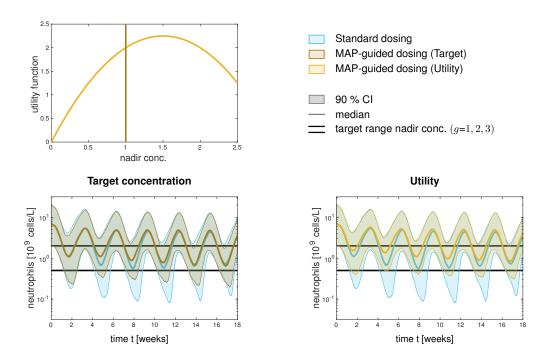


Figure C.5: Comparison of different objective functions for MAP-guided dosing. Two different objective functions were considered for MAP-guided dosing: (i) the least squared differences to a target concentration of $c_{\text{nadir}} = 1 \cdot 10^9 \text{cells/L}$, (ii) a utility function which penalizes low nadir concentration (in the range of grade 4 neutropenia) higher compared to high neutrophil concentrations (in the range of grade 0 neutropenia).

C.4 Details on DA-guided dosing

For the DA-guided approach, we chose the *optimal dose* to be the dose that minimizes the a-posteriori probability of being outside the target range, i.e., the weighted sum of the predicted (a-posteriori) probability of the patient having neutropenia grade $g_c = 0$ or $g_c = 4$ in the next cycle, see Eq. 4.7. We illustrated the DA-guided dosing approach exemplarily for the second cycle dose selection for a virtual patient, see Figure C.6.

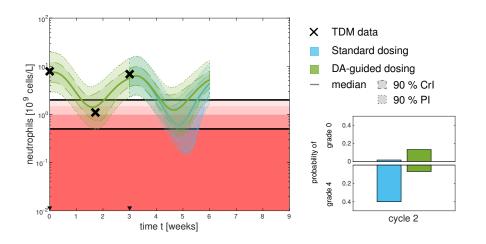


Figure C.6: Exemplary dose selection for cycle 2 in DA-guided dosing. The optimal dose in DA-guided dosing was defined to be the dose that minimizes the weighted sum of risk of grade 0 and grade 4, i.e., probability of grade 0 and grade 4 neutropenia. Note for illustration purposes a larger number of particles was chosen $M = 10^3$. For comparison we also propagated the particle ensemble for the case if the standard dose was chosen (blue).

C.5 Details on RL-guided dosing

The training phase of MCTS is illustrated on the level of the estimate of the action-value function in Figure C.7, and on the level of the therapeutic outcome in Figure C.8.

The computed $\hat{q}_{\pi_{\text{UCT}}}$ -Matrix can be used as a look-up table. For a certain patient state we need to determine the corresponding row in the matrix and then select the dose corresponding to the maximal $\hat{q}_{\pi_{\text{UCT}}}$ -value. This procedure can be visualized in a diagram structure similar to the one developed by Joerger et al. [11], see Figure C.9. Since RL allows to deal with a large amount of information regarding patient state/dose combinations, we just depict a small subtree.

C.5.1 Possible extensions of RL-guided dosing using the variance of the return

The variance of the return

$$\sigma_q = \sqrt{\operatorname{Var}_{\pi}[G_c|S_c = s, D_c = d]}, \qquad (C.1)$$

provides additional information about the associated uncertainty of a dose selection. For estimating Eq. (C.1) for each state-action pair the additional statistic

$$M_{2,k}(s_c, d) = M_{2,k-1}(s_c, d) + (g_c^{(k)} - q_{k-1})(G_c^{(k)} - q_k)$$

is saved, which can be used to calculate the variance via $s_k^2(s_c, d) = M_{2,k}(s_c, d)/(k-1)$ or $\sigma_k^2(s_c, d) = M_{2,k}(s_c, d)/k$ (Welford's online algorithm). Figure C.10 A shows the variance of

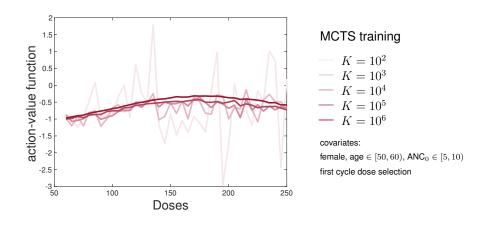


Figure C.7: Training stages of Monte Carlo tree search (MCTS): Approximation of q_{π} . The same virtual test patient population was dosed according to the current estimate of the action-value function q_K after K episodes of the planning steps for the covariate class: female, age between 50 and 60 years, and pre-treatment neutrophil counts ANC₀ \in [5, 10) in [10⁹ cells/L].

the return for certain patient states assuming a normally distributed return, i.e., the 95% CIs are computed via $\hat{q}_{\pi} \pm 1.96 \sigma_q$. The variance could be used similar as the UCB in an exploration strategy, e.g., taking doses less often if the variance is small (see sub-panels for states $g_{1:5} = (0, 0, 0, 0, 0)$ and $g_{1:5} = (4, 4, 4, 4, 4)$). Panel B shows the visiting counts for the current implementation with UCT so that states with large estimated action-value are visited more often.

C.5.2 Investigating the effect of the tuning parameters

First, we investigated the choice of the discount parameter $\gamma \in [0, 1]$. Since no discount is applied to the current cycle, small γ values give the current cycle a much higher weighting compared to later cycles, i.e., prioritizing the short term return rather than the long-term outcome. On the contrary, large γ values put more weight on the long-term goals. In our setting, the long-term goal (median survival) is already included in the immediate reward, since neutropenia grade 0 was also evaluated with -1 in the reward function. Therefore, γ does not have such a strong impact on the results, see Figure C.11. The parameter is expected to be more relevant if efficacy is not 'measured' by the surrogate marker of neutropenia, but rather evaluated based on a tumor growth model or a survival model. This would result in rewards with a larger time lag since the choice of a dose impacts tumor growth and in particular, survival only at (much) later time points.

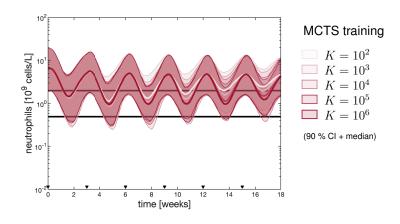


Figure C.8: Training stages of Monte Carlo tree search (MCTS). The same virtual test patient population (as in Figure C.7) was dosed according to the current estimate of the action-value function q_K after K iterations of the planning steps per covariate class.

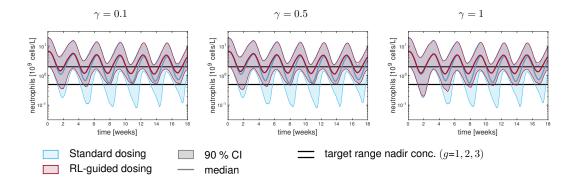


Figure C.11: Choice of the discount parameter γ . The discount parameter $\gamma \in [0, 1]$ weighs the relation between short term and long-term goals. Note, this analysis was done with sampling time points day 0 & 12 and a virtual test population of N = 1000.

We further examined the trade-off between exploration and exploitation. For this, we varied the constant $c_{\rm UCT}$ in Eq. (4.9) in the MCTS with UCT approach. We found that for smaller $c_{\rm UCT}$ values, the algorithm selected only a small number of doses with relatively high probability, see e.g., Figure C.12 for the initial dose selection. This led to an un-smooth action-value function, which is not expected in the considered scenario. Therefore, we chose $c_{\rm UCT} = 3$, as this choice showed a balanced exploration of the dose space while still prioritizing doses with a high expected return.

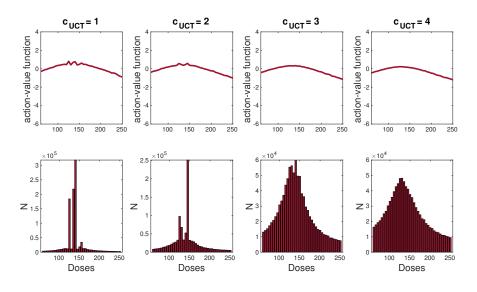


Figure C.12: Monte Carlo tree search: exploration/exploitation parameter. Expectation of the long-term return (action-value function) for exemplary states of the covariate class: male, age $\in [50, 60)$, ANCO $\in [2.5, 5)$ (top panel) and visiting counts of dose selections in the initial state s_0 , $N(s_0, d)$ in training phase for $K = 10^6$ (bottom panel) for different exploration/exploitation parameter $c_{\rm UCT}$ in Eq. 4.9.

Finally, we investigated the effect of changes in the reward function. For this, we exemplary changed the reward—here corresponding to a penalization—of grade 4 neutropenia. In the first scenario, the reward of grade 4 neutropenia was set equal to the reward of grade 0 neutropenia ($R_{c+1} = -1$, if $g_c = 4$). Thus, subtherapeutic and toxic ranges result in the same (negative) reward value. The second scenario, $R_{c+1} = -2$, if $g_c = 4$ corresponds to the scenario presented in Chapter 4. In the third scenario, neutropenia grade 4 was even more strongly penalized, reflecting the potential of exposing patients to immediate life-threatening conditions ($R_{c+1} = -3$, if $g_c = 4$). As expected, the occurrence of grade 4 decreased the stronger grade 4 neutropenia was penalized, see Figure C.13. Due to the uncertainty, at the same time, the incidence of grade 0 is increased. Thus, it is crucial to have a clear therapeutic goal prior to defining the evaluation function. The choice of the evaluation function should be examined in comparison with potential alternatives, as in Figure C.13 and the results should be compared with the desired therapeutic outcome.

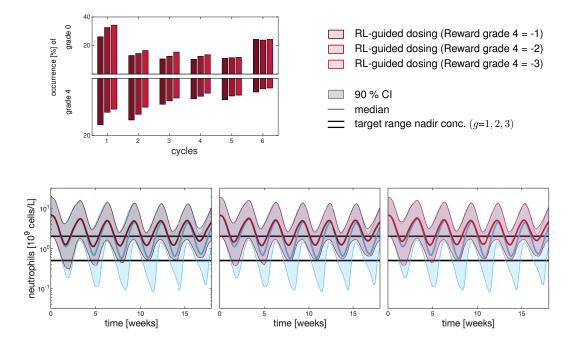


Figure C.13: Comparison of RL-guided dosing results for changes in the reward function. In the three scenarios the rewards for grades 0-3 remained the same and only the reward for grade 4 was changed. In the first panel the reward value of grade 4 was set to -1, thus equal to the reward of grade 0. The second reward function corresponds to the scenario presented in Chapter 4 (-2) and in the last scenario a larger penalty (-3) is put on grade 4 in comparison to grade 0.

C.5.3 Q-planning as an alternative to MCTS

As an alternative to MCTS, Q-planning can be performed to estimate the action-value function. We employed the same state representation and the same reward function as for the MCTS approach. We also visualized the training phase for Q-planning, see Figure C.14. For this specific example and selected patient state representation, the results using MCTS are more promising and could better reduce the incidence of grade 0 & 4 neutropenia in later cycles.

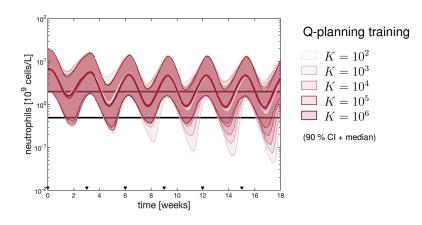


Figure C.14: Training stages of Q-planning. The same virtual test patient population was dosed according to the current estimate of the action-value function q_K after K iterations of the planning steps per covariate class.

C.6 Details on DA-RL-guided dosing

DA can be integrated into RL in two ways, (i) by improving the state representation, and (ii) by using the posterior ensemble in a decision-time planning procedure to update and individualize the estimate of the q_{π} -values reflecting the posterior uncertainty.

First, before any patient is treated, a prior dosing policy $\hat{q}_{\pi_{\text{UCT}}}$ is planned, i.e., determined, via model-based RL, e.g., via MCTS+UCT as in Section 4.1.2. When a patient is to be treated, the ensemble \mathcal{E}_0 for the sequential DA algorithm, e.g., particle filter/smoother, is initialized. The patient-specific TDM data $y_{1:c}$ are integrated, leading to an updated posterior particle ensemble $\mathcal{E}_{1:c}$. At a decision time point t_c , the posterior expectation is computed for an improved estimate of the current patient state, e.g., a sample approximation to the posterior expectation of a nadir concentration

$$\hat{c}_{\text{nadir}} = \sum_{m=1}^{M} w_c^{(m)} \cdot c_{\text{nadir}} \left(x_{1:c}^{(m)}, \theta^{(m)} \right) \,, \tag{C.2}$$

where $c_{\text{nadir}}(x_{1:c}^{(m)}, \theta^{(m)})$ denotes the minimum neutrophil concentration of the *m*-th particle within the cycle. The posterior expected nadir \hat{c}_{nadir} is translated to the corresponding neutropenia grade of the cycle g_c and used to update the current patient state s_c . An MCTS search tree is initialized at the current patient state s_c and the search within the tree is guided by the PUCT algorithm [34], where prior probabilities of choosing a dose are computed from the prior $\hat{q}_{\pi_{\text{UCT}}}$ -values, see Eq. (4.8). For model simulations within each episode in the MCTS the model state parameter vector $x^{(k)}$ and $\theta^{(k)}$ is sampled from the posterior particle ensemble $\mathcal{E}_{1:c}$.

C.6.1 Different approaches to estimate the grade of neutropenia \hat{g}_c in cycle c

In DA-RL-guided dosing, the particle ensemble $\mathcal{E}_{1:c}$ is used to estimate the patient state more reliably than just using the observed neutrophil concentration at day 12 or 15. Figure C.15 shows the RMSE between the estimated neutropenia grade \hat{g}_c and the *true* grade g_c from the underlying 'truth' used to simulate the data. Note, that we neglected the '^' in the main text for ease of notation. Overall, the RMSE is lower for day 12 than for day 15. Moreover, using a model-based state representation reduced the RMSE substantially—and much more than the difference between day 12 and 15. We further compared the posterior expected nadir concentration, see Eq. (C.2), translated into discrete grades, with first computing the probabilities of the different grades and then using the maximum a-posteriori grade, i.e., the grade with the highest sum of weight. The posterior expected nadir concentration performed slightly better and was therefore used in Chapter 4 for approximating the patient state using the particle ensemble $\mathcal{E}_{1:c}$.

C.6.2 RL-guided dosing based on DA state

In the main manuscript, we discussed that DA can be used in two ways to improve RL-guided dosing: (i) providing an improved state estimate (as in the previous section); and (ii) by using the posterior particle ensemble $\mathcal{E}_{1:c}$ to update the $\hat{q}_{\pi_{\text{UCT}}}$ values in relevant and promising dose-state-pairs. In Figure C.16, we investigate the scenario (i) alone, i.e., if we only use the improved state estimate in RL-guided dosing (without decision time planning based on the posterior particle ensemble). We observed a one-sided improvement, only the occurrence of grade 0 was reduced compared to RL alone. This indicates again the key role of individualized uncertainties for MIPD. In short: if the quality of estimating the grade of neutropenia is improved, also the corresponding dosing table should be updated since the RL dosing table accounted for the potential 'bias' in the state estimation. If not, improved estimates are used in decision trees that have been determined based on the less accurate estimate of the grade of neutropenia. Such a mismatch should be avoided.

C.6.3 PUCT algorithm

In the PUCT algorithm, the pre-calculated action-value function values $\hat{q}_{\pi_{\text{UCT}}}$ have to be translated to probabilities. As described in the main text, we used the Boltzmann distribution (see Eq. (4.8) 'prioritizing part') to convert the expectation values in \mathbb{R} to probabilities in [0, 1]. In addition, we performed a kernel density estimation to further smooth the function in case of a rough action-value function due to small visiting counts (this step is more relevant if less pre-training steps were possible, e.g., in larger state spaces), compare small K values (rough) to large K values in Figure C.7. As a result of the updated uncertainties, the action-value function $\hat{q}_{\pi_{\text{PUCT}}}$ (DA-RL-guided dosing) differs from the static $\hat{q}_{\pi_{\text{UCT}}}$ (RL-guided dosing), see Figure C.17. This also led to different optimal doses (markers at the x axis). The purple bars show the visiting counts N of the different doses in the given state, showing that doses are chosen more often that have high $\hat{q}_{\pi_{\text{UCT}}}$ (red line) as enforced via the PUCT algorithm. It can be also seen that the $\hat{q}_{\pi_{PUCT}}$ -curve (purple line) is not very smooth in dose regions which have low $\hat{q}_{\pi_{\rm UCT}}$ values as these values are not chosen often. In PUCT, the search focused more on promising regions of the dose space. In practical applications, deviations from this highly focused search need to be discussed depending on how much one wants to trust the prior knowledge or how much we expect the new patients to deviate (see also Chapter 5).

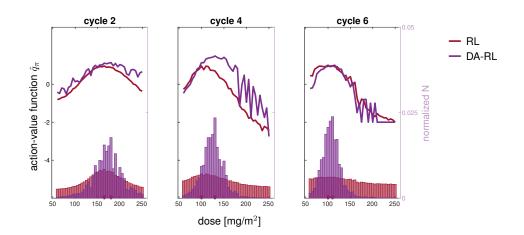


Figure C.17: Comparison of the action-value function values \hat{q}_{π} and visiting counts N (normalized) for RL-guided dosing and DA-RL-guided dosing. The individual action-value function values (DA-RL) differ from the population values (RL). Doses are chosen more frequently in regions where the prior probabilities, defined by the population action-value (red line), are large, see histogram (right axis). Note that it can be observed that the \hat{q}_{π} values are smoother in regions with high number of samples (e.g., cycle 2 for lower doses). Regions with low prior probabilities (computed from RL \hat{q}_{π} values (red line)) are chosen less often. The visiting counts were divided by the maximum number of visits for one dose (i.e., scaled to one, $N/\max(N)$) to allow for comparison since $K_{\rm RL} >> K_{\rm DA-RL}$.

C.7 Comparison across all considered evaluation functions

The different methods towards the optimal dose selection problem considered in the manuscript are based on different evaluation/reward functions. For a more in-depth comparison, we also show in Figure C.18 the results of the different methods for all considered evaluation functions: the utility (MAP-guided dosing), deviation from target concentration (MAP-guided dosing), the weighted sum of occurrence of grade 0/4 (DA-guided dosing), and the total reward (RL-guided dosing).

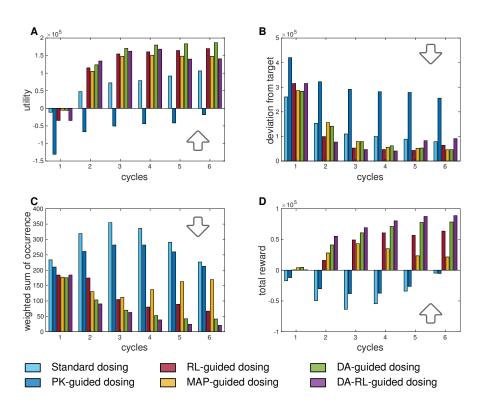


Figure C.18: Comparison of methods across all evaluation functions. (A) Comparison of the utility used in the MAP-guided dosing, see Figure C.5. The higher the utility the better (upward arrow). (B) The deviation from a target concentration $(1 \cdot 10^9 \cdot \text{cells/L})$ (target concentration intervention). The smaller the deviation from the target the better (downward arrow). (C) The weighted sum of occurrence was minimized in the DA-guided approach. The occurrence of grade 4 was penalized more strongly than the occurrence of grade 0. The smaller the weighted sum the better (downward arrow). (D) The total reward as defined for the RL-guided dosing is to be maximized. The higher the total reward the better (upward arrow).

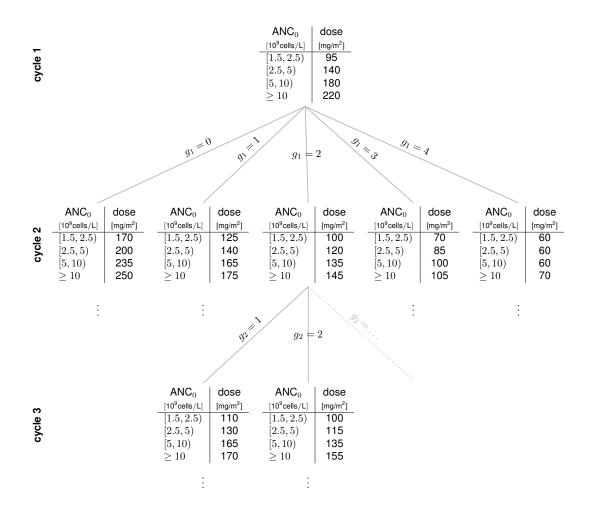
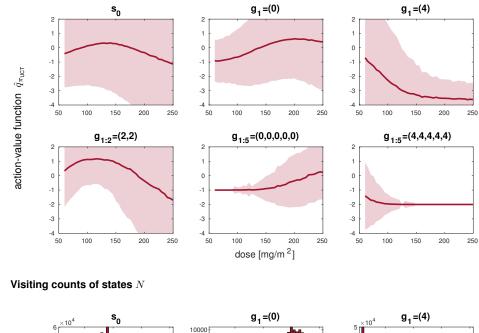


Figure C.9: Illustration of the decision tree procedure for the dose selection. From the action-value function values \hat{q}_{π} a decision tree/look-up table can be extracted. Here shown for fixed covariates: male, age $\in [50, 60)$. For example, the optimal dose for the second cycle depends on the neutropenia grade of the previous cycle and the pre-treatment neutrophil count ANC₀.



A Expectation and variance of long-term return

В

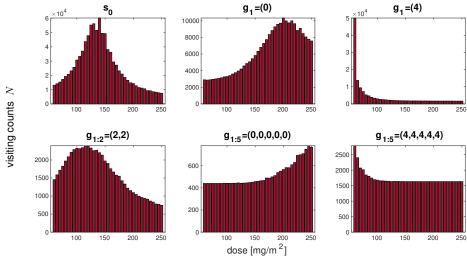


Figure C.10: Statistics of Monte Carlo tree search (MCTS) for different states. (A) Expectation of the long-term return for exemplary states for the covariate class: male, age $\in [50, 60)$, ANC₀ $\in [2.5, 5) \cdot 10^9$ cells/L together with the 95% CIs based on the assumption of a normally distributed return. Note that the CIs are cut off at the minimum/maximum possible return. (B) Visiting counts of states in training phase for $K = 10^6$.

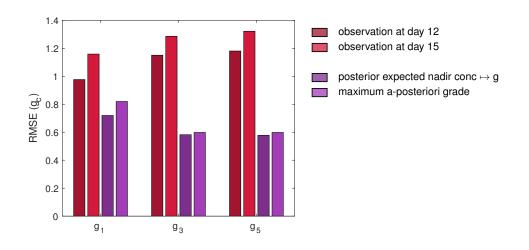


Figure C.15: Root-mean squared error (RMSE) of estimating the grade of neutropenia \hat{g}_c in cycle *c* using different approaches. In RL-guided dosing, the neutrophil measurement is used to infer the neutropenia grade. We investigated the two sampling time points day 12 (typical nadir time) and day 15 (as in the CEPAC-TDM study). For DA-RL-guided dosing, the particle ensemble $\mathcal{E}_{1:c}$ can be used to infer an improved patient state. There are two options: (i) the posterior expected nadir concentration is computed and then translated into a discrete grade; or (ii) the probability of each grade is determined by summing the weights of particle giving raise to that grade; then, the maximum a-posteriori grade is defined as the grade with highest sum of weights.

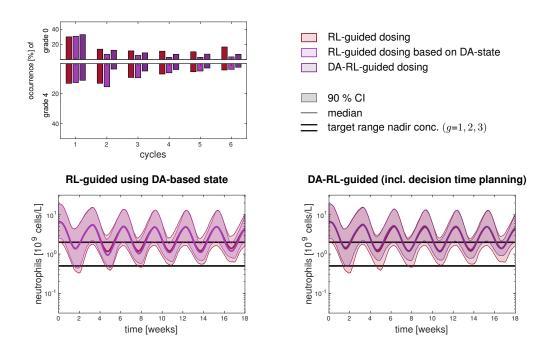


Figure C.16: RL-guided dosing based on the DA-based model state of the patient. The virtual test population was dosed with RL-guided dosing adding the different aspects of DA. Lower left panel: RL-guided dosing using the improved DA-based model state of the patient. Here we used the smoothed posterior expected nadir concentration translated to the discrete neutropenia grades. Lower right panel: DA-RL-guided dosing as presented in Chapter 4. Note, that we used the scenario with sampling time points day 0 & 12 for this analysis.

D Appendix related to Chapter 5

D.1 Paclitaxel-induced neutropenia models

In Table D.1 we provide multiple models that were all proposed for paclitaxel-induced neutropenia based on different or even on the same patient population. This shows the difficulty associated with the selection of a model to use for MIPD. The model structure was introduced by Friberg et al. [91]. In subsequent publications the model structure (gold-standard) remained the same, but either the covariate model changed (Kloft et al. [67]), IOV was modeled (Hansson et al. [94]) or it was fitted to a different patient population (Joerger et al. [11], Henrich et al. [90]. Later, the model was extended to describe also the cumulative behavior of neutropenia (BME) which could be observed over multiple cycle treatment (Henrich et al. [90]).

Parameter	Friberg et al. (2002)	Kloft et al. (2006)	Hansson et al. (2010)	$\begin{array}{c} \text{Joerger et al.} \\ (2007) \end{array}$	Joerger et al. (2012)	Henrich et al. Henrich et al. (2017) (2017)	Henrich et : (2017)
N (Ids)	45	45	45	104	104	366	366
n (samples)	530	530	523		314	3274	3274
cycles	ယ	3(1-18)	3(1-11)	(1-10)	1	6	6
structure	G-S	G-S	G-S	G-S	G-S	G-S	BME
TV							
$ANC_0 [10^9 \text{ cells/L}]$	5.20(3.6)	5.40(7.2)	5.61 (9.4)	$4.35 (1.5-9.4)^*$	ı	6.48*	6.48*
MTT [h]	127(2.1)	126(4.2)	154(4.4)	141(3.7)	141	128(2.03)	145 (2.65)
Slope $[L/\mu mol]$	2.21 (4.5)	2.8(13)	3.48(8.1)	2.08^{\dagger} (12.5)	2.6	4.48(4.55)	13.1(4.56)
γ	0.230(2.8)	0.223	0.270(5.9)	0.26(7.5)	0.2	0.231(6.79)	0.257(5.53)
ftr	I	I	ı	I	I	ı	0.787(2.7)
IIV $(CV\%)$							
MTT	18(30)	17(43)	17(22)	I	27.0	ı	I
Slope	43(32)	36(38)	39(20)	65.5(23.9)	44.9	43.8(8.23)	44.8 (6.54)
ANC_0	35(11)	35(23)	36(13)	41.4(7.56)	31.6	60.3(3.27)	51.5(3.61)
IOV (CV%)							,
MTT	I	I	16(8.5)	I	ı	ı	I
\mathbf{RUV}							
exp. (CV %)	39.9	29.1		41.4	31.6	60.3(3.27)	51.5(3.61)
add. $(\cdot 10^9 \text{ cells/L})$		0.626					
box-cox.			0.431(2.6)				

Table D.1: Parameter estimates for different PD models describing paclitaxel-induced neutropenia. Note that in Kloft et al. (2006) and Hansson et al. (2010), the Slope parameter was reported for unbound drug concentration (Slope = 54.5 and 69.6 L/ μ mol, respectively), which is assumed to be 5% of the total drug concentration; thus we multiplied the values with 0.05 for the Slope parameter that is related to the total drug concentration (for a linear transformation, the relative standard error remains the same). The typical values (TV) describe the fixed effects and the inter-individual variability (IIV) parameters the variability between patients. The residual variability (RUV) describes the deviation between measurements and model predictions accounting for measurement errors and potential model misspecifications. The IIV and RUV parameters are provided as coefficients of variation (CV). Note that baseline method B2 was used for baseline neutrophil counts ANC₀, i.e., the IIV was estimated together with RUV as one single parameter [96].

D.2 Results including an estimation of γ

When γ is included on the individual level inference, the estimation of the typical parameter values for 'Slope' and 'MTT' across patients is improved, see Figure D.1.

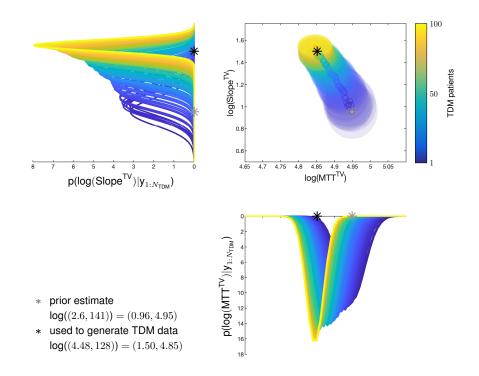


Figure D.1: Exemplary multivariate update of the typical values for MTT (mean transit time) and Slope (drug effect parameter) for the rich sampling scheme (TDM data every third day). The initial model used in the MIPD approach is the gold-standard model and the TDM data are generated using the gold-standard R model. In this case also the parameter γ was estimated on the individual level.

D.3 Parameter identifiability

To investigate the practical identifiability for the intermediate sampling scheme (weekly), we exemplarily computed the loglikelihood for four virtual patients at the end of the therapy, see Figure D.2. In order to exclude effects from other parameters we investigated a simplified setting in which only the parameters 'Slope' and 'Circ₀' were estimated on the individual level. For some virtual patients, the loglikelihood takes the same values for various 'Slope' values, which can be seen from the elongated yellow ranges covering a larger range of 'Slope' values. In addition, the data suggest larger 'Slope' values as the maximum of the likelihood (ML, yellow cross) is reached for larger 'Slope' values than used to generate the data (black cross) in the upper panels. Furthermore, it can be observed that the prior only has a minor influence on the logposterior, when comparing the upper panels (loglikelihood) with the corresponding lower panels (logposterior). The analysis mean (red cross) is close to the MAP estimate

D.4 Optimal design

(yellow cross in bottom panels). Note that for skewed distributions the mean is different from the mode, and therefore, the particle mean should not be directly compared to the MAP. The particles (red circles) cover areas of high posterior probability well.

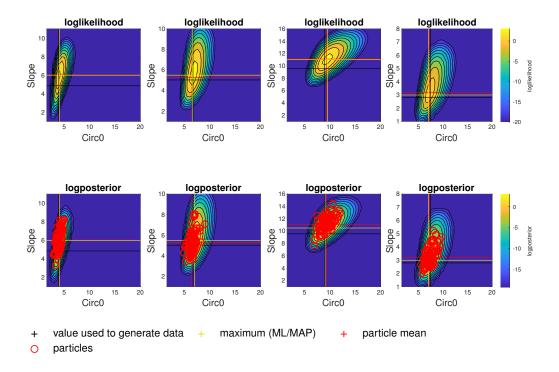


Figure D.2: Practical identifiability for the weekly sampling design. For illustration, M' = 100 particles were resampled from the M = 1000 particles used within the particle filter. ML: maximum likelihood, MAP: maximum a-posteriori

D.4 Optimal design

To investigate whether the reason for the practical identifiability is the choice of sampling time points, we investigated the optimal design for a design with three sampling timepoints where the first sampling time point at day 1 is fixed. To infer the optimal design we used the frequently used criterion of D-optimality, i.e., choosing the design that maximizes the determinant of the FIM. The optimal design was determined for the typical patient (Figure D.3 left) and the whole patient population (Figure D.3 right). The optimal second time point is approximately one day later than in the weekly sampling scheme (day 7) and when only the typical patient is considered, the third time point of the weekly sampling scheme (day 14) is chosen well. However, when the design is chosen based on the entire patient population, an earlier third time point is suggested.

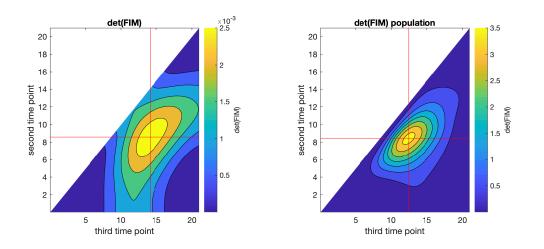


Figure D.3: Optimal design for the gold-standard model. The first sample timepoint at day 1 is fixed and only two more sample points within one cycle were allowed. The left panel shows the D-optimality criterion landscape for the typical patient and the right panel for the population.

D.5 Impact on MIPD: parameter bias scenario

Especially at the beginning of a patient's therapy when the misspecified prior dominated, MIPD benefited substantially from updating the model with every patient, see Figure D.4. The occurrence of grade 4 in the first cycle was considerably reduced compared to DAguided dosing alone. As more patient-specific data were collected, individual parameters were increasingly well estimated with the DA approach, and the influence of the misspecified prior vanished. The occurrence of grade 0 neutropenia is slightly increased in later cycles, which might be related to the overestimation of the IIV parameters ($\omega_{MTT}^2, \omega_{Slope}^2$), see Figure 5.4. We do not estimate the RUV parameter σ , however, the TDM data were generated with an increased parameter value for σ , therefore, the variability parameters capture to some extent the increased variability in the data. This increases the uncertainty on the individual level. DA-guided dosing alone has in this case the advantage that the IIV parameters are fairly similar between the gold-standard and gold-standard R.

The population updates improved the used MIPD approach, especially for the first treatment cycle, when no patient-specific TDM data were available and the dose was solely determined based on *a priori* predictions.

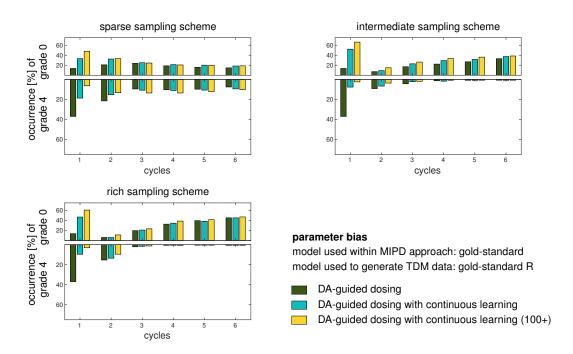


Figure D.4: Impact of continuous learning approach on MIPD outcomes. TDM data were generated for $N_{\text{TDM}} = 100$ virtual patients based on the gold-standard R model. DA-guided dosing based on the gold-standard model was used alone or in conjunction with continuous learning of the population parameters. The analysis was repeated 10 times to account for statistical variability.

D.6 Impact on MIPD: structural bias scenario

In Chapter 5, learning of temporal changes was only shown for the intermediate (weekly) sampling scheme. For completeness, Figure D.5 and Figure D.6 show the results for the sparse and rich sampling schemes, respectively. In the case of sparse TDM data, the continuous learning updates decrease the incidence of grade 4 neutropenia in early cycles but even lead to an increase in later cycles. This might be again related to the overestimation of the magnitude of IIV. For the rich sampling scheme, the results are comparable to the intermediate sampling scheme presented in the main manuscript. Comparison of the parameter estimates could be misleading due to the structural differences of the models.

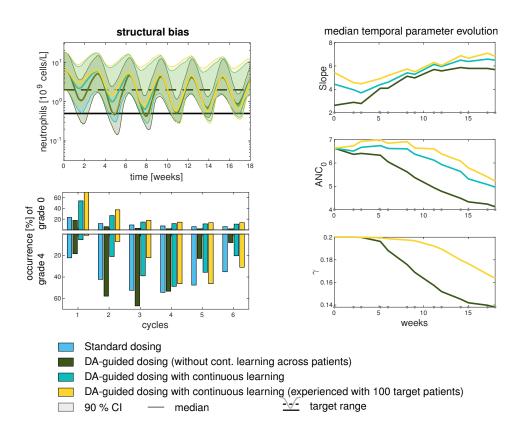


Figure D.5: Impact on MIPD: structural bias scenario with sparse sampling scheme. Learning of temporal changes and adapting to structural model changes considering the sparse sampling scheme.

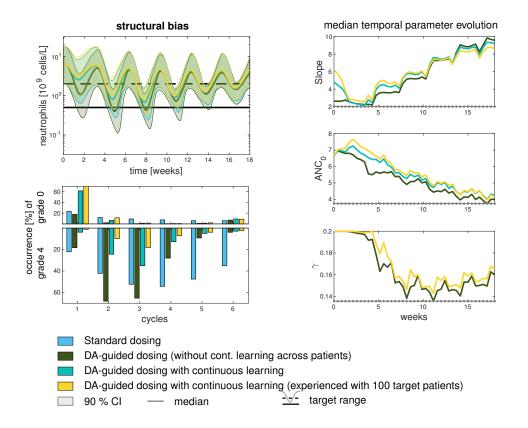


Figure D.6: Impact on MIPD: structural bias scenario with rich sampling scheme. Learning of temporal changes and adapting to structural model changes considering the rich sampling scheme.

Bibliography

- [1] H. Gurney. "Dose calculation of anticancer drugs: a review of the current practice and introduction of an alternative." In: J. Clin. Oncol. 14 (1996), pp. 2590–2611.
- [2] A. Felici, J. Verweij, and A. Sparreboom. "Dosing strategies for anticancer drugs: The good, the bad and body-surface area". In: Eur. J. Cancer 38.13 (2002), pp. 1677–1684.
- [3] M. E. de Jonge, A. D. R. Huitema, J. H. M. Schellens, S. Rodenhuis, and J. H. Beijnen. "Individualised Cancer Chemotherapy: Strategies and Performance of Prospective Studies on Therapeutic Drug Monitoring with Dose Adaptation". In: *Clin. Pharma-cokinet.* 44 (2005), pp. 147–173.
- [4] E. S. Vesell. "Genetic and environmental factors affecting drug disposition in man". In: Clin. Pharmacol. Ther. 22.5part2 (1977), pp. 659–679.
- [5] M. Ingelman-Sundberg. "Genetic and environmental causes for interindividual variability in drug pharmacokinetics". In: Int. Congr. Ser. 1220 (2001), pp. 175–186.
- [6] A. S. Darwich, K. Ogungbenro, O. J. Hatley, and A. Rostami-Hodjegan. "Role of pharmacokinetic modeling and simulation in precision dosing of anticancer drugs". In: *Transl. Cancer Res.* 6.S10 (2017).
- [7] A. S. Darwich et al. "Why has model-informed precision dosing not yet become common clinical reality? lessons from the past and a roadmap for the future". In: *Clin. Pharmacol. Ther.* 101.5 (2017), pp. 646–656.
- [8] R. W. Peck. "The right dose for every patient: a key step for precision medicine". In: Nat. Rev. Drug Discov. 15 (2015), pp. 145–146.
- [9] M. Joerger et al. "Population Pharmacokinetics and Pharmacodynamics of Paclitaxel and Carboplatin in Ovarian Cancer Patients: A Study by the European Organization for Research and Treatment of Cancer-Pharmacology and Molecular Mechanisms Group and New Drug Development Group". In: *Clin. Cancer Res.* 13.21 (2007), pp. 6410–6418.
- [10] C. K. Lee et al. "Carboplatin-paclitaxel-induced leukopenia and neuropathy predict progression-free survival in recurrent ovarian cancer". In: Br. J. Cancer 105 (2011), pp. 360–365.
- [11] M. Joerger, S. Kraff, A. D. R. Huitema, G. Feiss, B. Moritz, J. H. M. Schellens, J. H. Beijnen, and U. Jaehde. "Evaluation of a Pharmacology-Driven Dosing Algorithm of 3-Weekly Paclitaxel Using Therapeutic Drug Monitoring". In: *Clin. Pharmacokinet.* 51 (2012), pp. 607–617.
- [12] H. Gurney. "How to calculate the dose of chemotherapy". In: Br. J. Cancer 86 (2002), pp. 1297–1302.

- [13] D. A. Cameron, C. Massie, G. Kerr, and R. C. Leonard. "Moderate neutropenia with adjuvant CMF confers improved survival in early breast cancer". In: Br. J. Cancer 89 (2003), pp. 1837–1842.
- [14] K. Venkatakrishnan et al. "Optimizing Oncology Therapeutics Through Quantitative Translational and Clinical Pharmacology: Challenges and Opportunities". In: *Clin. Pharmacol. Ther.* 97.1 (2015), pp. 37–54.
- [15] R. J. Keizer, R. ter Heine, A. Frymoyer, L. J. Lesko, R. Mangat, and S. Goswami. "Model-Informed Precision Dosing at the Bedside: Scientific Challenges and Opportunities". In: *CPT Pharmacometrics Syst. Pharmacol.* 7.12 (2018), pp. 785–787.
- [16] T. M. Polasek, C. R. Rayner, R. W. Peck, A. Rowland, H. Kimko, and A. Rostami-Hodjegan. "Toward Dynamic Prescribing Information: Codevelopment of Companion Model-Informed Precision Dosing Tools in Drug Development". In: *Clin. Pharmacol. Drug Dev.* 8.4 (2019), pp. 418–425.
- [17] A. A. Vinks, R. W. Peck, M. Neely, and D. R. Mould. "Development and Implementation of Electronic Health Record-Integrated Model-Informed Clinical Decision Support Tools for the Precision Dosing of Drugs". In: *Clin. Pharmacol. Ther.* 107.1 (2020), pp. 129–135.
- [18] J.-S. Kang and M.-H. Lee. "Overview of Therapeutic Drug Monitoring". In: Korean J. Intern. Med. 24.1 (2009).
- [19] T. Buclin, Y. Thoma, N. Widmer, P. André, M. Guidi, C. Csajka, and L. A. Decosterd. "The Steps to Therapeutic Drug Monitoring: A Structured Approach Illustrated With Imatinib". In: Front. Pharmacol. 11.177 (2020).
- [20] L. B. Sheiner, S. Beal, B. Rosenberg, and V. V. Marathe. "Forecasting individual pharmacokinetics". In: *Clin. Pharmacol. Ther.* 26.3 (1979), pp. 294–305.
- [21] N. Holford. "Pharmacodynamic principles and target concentration intervention". In: *Transl. Clin. Pharmacol.* 26.4 (2018), pp. 150–154.
- [22] S. Reich and C. Cotter. Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge: Cambridge University Press, 2015.
- [23] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. J. Nordlund. "Particle filters for positioning, navigation, and tracking". In: *IEEE Trans. Signal Process.* 50.2 (2002), pp. 425–437.
- [24] P. Bauer, A. Thorpe, and G. Brunet. "The quiet revolution of numerical weather prediction". In: *Nature* 525 (2015), pp. 47–55.
- [25] R. Potthast, A. Walter, and A. Rhodin. "A localized adaptive particle filter within an operational NWP framework". In: Mon. Weather Rev. 147.1 (2019), pp. 345–362.
- [26] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. Kim, and S. Godsill. "Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking". In: *Digit. Signal Process.* 25 (2014), pp. 1–16.
- [27] S. Särkkä. Bayesian Filtering and Smoothing. Cambridge: Cambridge University Press, 2013.
- [28] D. J. Albers, M. E. Levine, A. Stuart, L. Mamykina, B. Gluckman, and G. Hripcsak. "Mechanistic machine learning: how data assimilation leverages physiologic knowledge using Bayesian inference to forecast the future, infer the present, and phenotype". In: J. Am. Med. Informatics Assoc. 25.10 (2018), pp. 1392–1401.

- [29] Q. Li, R. G. Mark, and G. D. Clifford. "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter". In: *Physiol. Meas.* 29.1 (2008), pp. 15–32.
- [30] E. H. Dunwoodie. "Home Testing of Blood Counts in Patients with Cancer". PhD. The University of Leeds, 2018.
- [31] P. Escandell-Montero et al. "Optimization of anemia treatment in hemodialysis patients via reinforcement learning". In: Artif. Intell. Med. 62.1 (2014), pp. 47–60.
- [32] R. S. Sutton and A. G. Barto. *Reinforcement Learning. An Introduction*. Ed. by F. Bach. 2nd. Cambridge, MA: The MIT Press, 2018.
- [33] D. P. Bertsekas. *Reinforcement Learning and Optimal Control*. Athena Scientific, 2019.
- [34] D. Silver et al. "Mastering the game of Go without human knowledge". In: Nature 550 (2017), pp. 354–359.
- [35] D. Silver et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". In: Science 362.6419 (2018), pp. 1140–1144.
- [36] Y. Zhao, D. Zeng, M. A. Socinski, and M. R. Kosorok. "Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer". In: *Biometrics* 67.4 (2011), pp. 1422–1433.
- [37] C. Yu, J. Liu, and S. Nemati. "Reinforcement Learning in Healthcare: A Survey". (preprint on arXiv). 2019.
- [38] G. Yauney and P. Shah. "Reinforcement Learning with Action-Derived Rewards for Chemotherapy and Clinical Trial Dosing Regimen Selection". In: Proc. Mach. Learn. Res. Ed. by F. Doshi-Velez, J. Fackler, K. Jung, D. Kale, R. Ranganath, B. Wallace, and J. Wiens. Vol. 85. Palo Alto, California: PMLR, 2018, pp. 161–226.
- [39] T. M. Polasek, S. Shakib, and A. Rostami-Hodjegan. "Precision dosing in clinical medicine: present and future". In: *Expert Rev. Clin. Pharmacol.* 11.8 (2018), pp. 743– 746.
- [40] C. Wild, E. Weiderpass, and B. Stewart, eds. World Cancer Report: Cancer Research for Cancer Prevention. Lyon, France, 2020.
- [41] A. C. Society. "Cancer Facts & Figures 2020". In: Am. Cancer Soc. Atlanta, 2020, pp. 1–76.
- [42] M. Roy and A. Datta. Cancer: Genetics and Cancer. Springer, Singapore, 1969.
- [43] G. Carioli, P. Bertuccio, P. Boffetta, F. Levi, C. La Vecchia, E. Negri, and M. Malvezzi. "European cancer mortality predictions for the year 2020 with a focus on prostate cancer". In: Ann. Oncol. 31.5 (2020), pp. 650–658.
- [44] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: CA. Cancer J. Clin. 68.6 (2018), pp. 394–424.
- [45] L. A. Emens and G. Middleton. "The Interplay of Immunotherapy and Chemotherapy: Harnessing Potential Synergies". In: *Cancer Immunol. Res.* 3.5 (2015), pp. 436–443.
- [46] C. Zappa and S. A. Mousa. "Non-small cell lung cancer: Current treatment and future advances". In: Transl. Lung Cancer Res. 5.3 (2016), pp. 288–300.

- [47] C. Lazzari et al. "Combination of immunotherapy with chemotherapy and radiotherapy in lung cancer: is this the beginning of the end for cancer?" In: *Ther. Adv. Med. Oncol.* 10 (2018).
- [48] J. N. Bodor, Y. Boumber, and H. Borghaei. "Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC)". In: Cancer 126.2 (2020), pp. 260– 270.
- [49] A. S. Narang and D. S. Desai. "Anticancer Drug Development". In: Pharm. Perspect. Cancer Ther. 1st. New York, NY: Springer US, 2009, pp. 49–92.
- [50] G. A. Orr, P. Verdier-Pinard, H. McDaid, and S. B. Horwitz. "Mechanisms of Taxol resistance related to microtubules". In: Oncogene 22 (2003), pp. 7280–7295.
- [51] K. A. Lyseng-Williamson and C. Fenton. "Docetaxel". In: Drugs 65 (2005), pp. 2513– 2531.
- [52] M. Joerger et al. "Open-label, randomized study of individualized, pharmacokinetically (PK)-guided dosing of paclitaxel combined with carboplatin or cisplatin in patients with advanced non-small-cell lung cancer (NSCLC)". In: Ann. Oncol. 27.10 (2016), pp. 1895–1902.
- [53] N. C. Kampan, M. T. Madondo, O. M. McNally, M. Quinn, and M. Plebanski. "Paclitaxel and its evolving role in the management of ovarian cancer". In: *Biomed Res. Int.* (2015).
- [54] M. J. Piccart and F. Cardoso. "Progress in systemic therapy for breast cancer: An overview and perspectives". In: Eur. J. Cancer, Suppl. 1.2 (2003), pp. 56–69.
- [55] P. Bonomi, K. M. Kim, D. Fairclough, D. Cella, J. Kugler, E. Rowinsky, M. Jiroutek, and D. Johnson. "Comparison of survival and quality of life in advanced non-small-cell lung cancer patients treated with two dose levels of paclitaxel combined with cisplatin versus etoposide with cisplatin: Results of an eastern cooperative oncology group trial". In: J. Clin. Oncol. 18.3 (2000), pp. 623–631.
- [56] A. Roth et al. "Docetaxel (Taxotere®)-cisplatin (TC): An effective drug combination in gastric carcinoma". In: Ann. Oncol. 11 (2000), pp. 301–306.
- [57] S. B. Park, D. Goldstein, A. V. Krishnan, C. S.-Y. Lin, M. L. Friedlander, J. Cassidy, M. Koltzenburg, and M. C. Kiernan. "Chemo\-therapy-induced peripheral neurotoxicity: A critical analysis". In: CA. Cancer J. Clin. 63.6 (2013), pp. 419–437.
- [58] J. Crawford, D. C. Dale, and G. H. Lyman. "Chemotherapy-induced neutropenia: risks, consequences, and new directions for its management." In: *Cancer* 100.2 (2004), pp. 228–237.
- [59] S. Jaillon, M. R. Galdiero, D. Del Prete, M. A. Cassatella, C. Garlanda, and A. Mantovani. "Neutrophils in innate and adaptive immunity". In: Semin. Immunopathol. 35.4 (2013), pp. 377–394.
- [60] S. N. Catlin, L. Busque, R. E. Gale, P. Guttorp, and J. L. Abkowitz. "The replication rate of human hematopoietic stem cells in vivo". In: *Blood* 117.17 (2011), pp. 4460– 4466.
- [61] G. E. Cartwright, J. W. Athens, and M. M. Wintrobe. "The Kinetics of Granulopoiesis in Normal Man." In: Blood 24 (1964), pp. 780–803.
- [62] S. M. Lawrence, R. Corriden, and V. Nizet. "The Ontogeny of a Neutrophil: Mechanisms of Granulopoiesis and Homeostasis". In: *Microbiol. Mol. Biol. Rev.* 82.1 (2018).

- [63] National Cancer Institute. Common terminology criteria for adverse events (CTCAE) version v5.0. 2017.
- [64] D. F. Bainton, J. L. Ullyot, and M. G. Farquhar. "The Development of Neutrophilic Polymorphonuclear Leukocytes in Human Bone Marrow". In: J. Exp. Med. 134.4 (1971), pp. 907–934.
- [65] C. Finch, L. Harker, and J. Cook. "Kinetics of the formed elements of human blood". In: Blood 50.4 (1977), pp. 699–707.
- [66] J. T. Dancey, K. A. Deubelbeiss, and L. A. Harker and Finch. "Neutrophil kinetics in man". In: J. Clin. Invest. 58.3 (1976), pp. 705–715.
- [67] C. Kloft, J. Wallin, A. Henningsson, E. Chatelut, and M. O. Karlsson. "Population Pharmacokinetic-Pharmacodynamic Model for Neutropenia with Patient Subgroup Identification: Comparison across Anticancer Drugs". In: *Clin. Cancer Res.* 12.18 (2006), pp. 5481–5490.
- [68] H. Takatani, H. Soda, M. Fukuda, M. Watanabe, A. Kinoshita, T. Nakamura, and M. Oka. "Levels of recombinant human granulocyte colony-stimulating factor in serum are inversely correlated with circulating neutrophil counts". In: Antimicrob. Agents Chemother. 40.4 (1996), pp. 988–991.
- [69] J. A. Sparano, U. Malik, L. Rajdev, C. Sarta, U. Hopkins, and A. C. Wolff. "Phase I trial of pegylated liposomal doxorubicin and docetaxel in advanced breast cancer". In: J. Clin. Oncol. 19.12 (2001), pp. 3117–3125.
- [70] C. P. Belani, S. Ramalingam, M. C. Perry, R. V. LaRocca, D. Rinaldi, P. S. Gable, and W. J. Tester. "Randomized, Phase III Study of Weekly Paclitaxel in Combination With Carboplatin Versus Standard Every-3-Weeks Administration of Carboplatin and Paclitaxel for Patients With Previously Untreated Advanced Non–Small-Cell Lung Cancer". In: J. Clin. Oncol. 26.3 (2008), pp. 468–473.
- [71] M. Di Maio et al. "Chemotherapy-induced neutropenia and treatment efficacy in advanced non-small-cell lung cancer: a pooled analysis of three randomised trials". In: *Lancet Oncol.* 6.9 (2005), pp. 669–677.
- [72] S. Hamauchi et al. "Neutropenia as a Predictive Factor in Metastatic Colorectal Cancer Treated With TAS-102". In: Clin. Colorectal Cancer 16.1 (2017), pp. 51–57.
- [73] I. Netterberg, E. I. Nielsen, L. E. Friberg, and M. O. Karlsson. "Model-based prediction of myelosuppression and recovery based on frequent neutrophil monitoring". In: *Cancer Chemother. Pharmacol.* 80 (2017), pp. 343–353.
- [74] M. Di Maio, C. Gridelli, C. Gallo, and F. Perrone. "Chemotherapy-induced neutropenia: a useful predictor of treatment efficacy?" In: Nat. Clin. Pract. Oncol. 3.3 (2006), pp. 114–115.
- [75] J. E. Wallin, L. E. Friberg, and M. O. Karlsson. "A tool for neutrophil guided dose adaptation in chemotherapy". In: *Comput. Methods Programs Biomed.* 93.3 (2009), pp. 283–291.
- [76] J. E. Wallin, L. E. Friberg, and M. O. Karlsson. "Model-based neutrophil-guided dose adaptation in chemotherapy: Evaluation of predicted outcome with different types and amounts of information". In: *Basic Clin. Pharmacol. Toxicol.* 106.3 (2009), pp. 234–242.
- [77] J. Y. Vis and A. Huisman. "Verification and quality control of routine hematology analyzers". In: Int. J. Lab. Hematol. 38.S1 (2016), pp. 100–109.

- [78] Hemocue WBC Diff. https://www.hemocue.com/en/solutions/hematology/hemo cue-wbc-diff-system (visited on 06/19/2020).
- [79] H. Russcher, N. Van Deursen, T. Ermens, and R. De Jonge. "Evaluation of the HemoCue WBC DIFF system for Point-of-Care counting of total and differential white cells in pediatric samples". In: Ned. Tijdschr. voor Klin. Chemie en Lab. 38.3 (2013), pp. 140–141.
- [80] M. Karawajczyk, S. Haile, M. Grabski, and A. Larsson. "The HemoCue WBC DIFF system could be used for leucocyte and neutrophil counts but not for full differential counts". In: Acta Paediatr. 106.6 (2017), pp. 974–978.
- [81] M. Patel, S. Palani, A. Chakravarty, J. Yang, W. C. Shyu, and J. T. Mettetal. "Dose schedule optimization and the pharmacokinetic driver of neutropenia". In: *PLOS ONE* 9.10 (2014), pp. 1–12.
- [82] R. Bruno, N. Vivier, J. C. Vergniol, S. L. De Phillips, G. Montay, and L. B. Sheiner. "A Population Pharmacokinetic Model for Docetaxel (Taxotere(R)): Model Building and Validation". In: J. Pharmacokinet. Biopharm. 24 (1996), pp. 153–172.
- [83] Central European Society for Anticancer Research (CESAR) Study of Paclitaxel Therapeutic Drug Monitoring (CEPAC-TDM). 2016. https://clinicaltrials.gov /ct2/show/NCT01326767 (visited on 07/15/2020).
- [84] J. E. Wallin, L. E. Friberg, and M. O. Karlsson. "Model Based Neutrophil Guided Dose Adaptation in Chemotherapy ; Evaluation of Predicted Outcome with Different Type and Amount of Information". In: Page Meet. St. Petersburg. 2009.
- [85] R. L. Lalonde et al. "Model-based Drug Development". In: Clin. Pharmacol. Ther. 82.1 (2007), pp. 21–32.
- [86] Y. Wang, H. Zhu, R. Madabushi, Q. Liu, S.-M. Huang, and I. Zineh. "Model-Informed Drug Development: Current US Regulatory Practice and Future Considerations". In: *Clin. Pharmacol. Ther.* 105.4 (2019), pp. 899–911.
- [87] M. Lavielle. Mixed Effects Models for the Population Approach. New York: Chapman and Hall/CRC, 2014.
- [88] D. R. Mould and R. N. Upton. "Basic concepts in population modeling, simulation, and model-based drug development". In: CPT Pharmacometrics Syst. Pharmacol. 1.9 (2012).
- [89] R. N. Upton and D. R. Mould. "Basic concepts in population modeling, simulation, and model-based drug development: Part 3-introduction to pharmacodynamic modeling methods". In: CPT Pharmacometrics Syst. Pharmacol. 3.1 (2014).
- [90] A. Henrich, M. Joerger, S. Kraff, U. Jaehde, W. Huisinga, C. Kloft, and Z. P. Parra-Guillen. "Semimechanistic Bone Marrow Exhaustion Pharmacokinetic/Pharmacodynamic Model for Chemotherapy-Induced Cumulative Neutropenia". In: J. Pharmacol. Exp. Ther. 362.2 (2017), pp. 347–358.
- [91] L. E. Friberg, A. Henningsson, H. Maas, L. Nguyen, and M. O. Karlsson. "Model of Chemotherapy-Induced Myelosuppression With Parameter Consistency Across Drugs". In: J. Clin. Oncol. 20.24 (2002), pp. 4713–4721.
- [92] N. Macdonald. "Time delay in simple chemostat models". In: Biotechnol. Bioeng. 18.6 (1976), pp. 805–812.

- [93] P. J. Hurtado and A. S. Kirosingh. "Generalizations of the 'Linear Chain Trick': incorporating more flexible dwell time distributions into mean field ODE models". In: J. Math. Biol. 79 (2019), pp. 1831–1883.
- [94] E. K. Hansson, J. E. Wallin, H. Lindman, M. Sandström, M. O. Karlsson, and L. E. Friberg. "Limited inter-occasion variability in relation to inter-individual variability in chemotherapy-induced myelosuppression". In: *Cancer Chemother. Pharmacol.* 65 (2010), pp. 839–848.
- [95] A. Henrich. "Pharmacometric modelling and simulation to optimise paclitaxel combination therapy based on pharmacokinetics, cumulative neutropenia and efficacy". PhD thesis. Freie Universität Berlin, 2017.
- [96] C. Dansirikul, H. E. Silber, and M. O. Karlsson. "Approaches to handling pharmacodynamic baseline responses". In: J. Pharmacokinet. Pharmacodyn. 35 (2008).
- [97] K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation*. Texts in Applied Mathematics. Cham: Springer International Publishing, 2015.
- [98] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. "Variational Inference: A Review for Statisticians". In: J. Am. Stat. Assoc. 112.518 (2017), pp. 859–877.
- [99] D. D. Boos and L. A. Stefanski. *Essential Statistical Inference*. Springer Texts in Statistics. Springer New York, 2013.
- [100] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. Bayesian Data Analysis. 3rd ed. New York: Chapman and Hall/CRC, 2014.
- [101] D. Freedman. "Wald Lecture: On the Bernstein-von Mises theorem with infinitedimensional parameters". In: Ann. Stat. 27.4 (1999), pp. 1119–1141.
- [102] A. F. M. Smith and A. E. Gelfand. "Bayesian Statistics without Tears: A Sampling-Resampling Perspective". In: Am. Stat. 46.2 (1992), pp. 84–88.
- [103] K. P. Murphy. Machine Learning: A Probabilistic Perspective. Cambridge, MA: The MIT Press, 2012.
- [104] A. Doucet and A. M. Johansen. "A Tutorial on Particle Filtering and Smoothing: Fifteen years later". In: Oxford Handb. Nonlinear Filter. Ed. by D. Crisan and B. Rozovsky. Oxford University Press, 2009.
- [105] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equation of state calculations by fast computing machines". In: J. Chem. Phys. 21.6 (1953), pp. 1087–1092.
- [106] W. K. Hastings. "Monte carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (1970), pp. 97–109.
- [107] S. Geman and D. Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6.6 (1984), pp. 721–741.
- [108] A. E. Gelfand and A. F. Smith. "Sampling-based approaches to calculating marginal densities". In: J. Am. Stat. Assoc. 85.410 (1990), pp. 398–409.
- [109] A. Gelman and D. B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: Stat. Sci. 7.4 (1992), pp. 457–472.
- [110] P. Del Moral, A. Doucet, and A. Jasra. "Sequential Monte Carlo samplers". In: J. R. Stat. Soc.: Ser. B (Statistical Methodol.) 68 (2006), pp. 411–436.

- [111] R. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: ASME. J. Basic Eng. 82.1 (1960), pp. 35–45.
- [112] R. Kalman and R. Bucy. "New Results in Linear Filtering and Prediction Theory". In: ASME. J. Basic Eng. 83.1 (1961), pp. 95–108.
- [113] N. Gordon, D. Salmond, and A. Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE Proc. F Radar Signal Process.* 140.2 (1993), pp. 107–113.
- [114] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Trans. Signal Process.* 50.2 (2002), pp. 174–188.
- [115] W. Acevedo, J. de Wiljes, and S. Reich. "Second-order Accurate Ensemble Transform Particle Filters". In: SIAM J. Sci. Comput. 39.5 (2017), A1834–A1850.
- [116] S. Reich. "A non-parametric ensemble transform method for Bayesian inference". In: SIAM J. Sci. Comput. 35.4 (2013), A2013–A2024.
- [117] M. Reinhardt. "Hybrid filters and multi-scale models". PhD Thesis. University of Potsdam, 2019.
- [118] C. W. Tornøe, R. V. Overgaard, H. Agersø, H. A. Nielsen, H. Madsen, and E. N. Jonsson. "Stochastic differential equations in NONMEM®: Implementation, application, and comparison with ordinary differential equations". In: *Pharm. Res.* 22.8 (2005), pp. 1247–1258.
- [119] B. Matzuka, J. Chittenden, J. Monteleone, and H. Tran. "Stochastic nonlinear mixed effects: a metformin case study". In: J. Pharmacokinet. Pharmacodyn. 43 (2016), pp. 85–98.
- [120] J. Leander, J. Almquist, C. Ahlström, J. Gabrielsson, and M. Jirstrand. "Mixed Effects Modeling Using Stochastic Differential Equations: Illustrated by Pharmacokinetic Data of Nicotinic Acid in Obese Zucker Rats". In: AAPS J. 17.3 (2015), pp. 586–596.
- [121] S. Donnet and A. Samson. "A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models". In: Adv. Drug Deliv. Rev. 65.7 (2013), pp. 929–939.
- [122] M. Lavielle. "Pharmacometrics models with hidden Markovian dynamics". In: J. Pharmacokinet. Pharmacodyn. 45.1 (2018), pp. 91–105.
- [123] G. Hripcsak and D. J. Albers. "Next-generation phenotyping of electronic health records". In: J. Am. Med. Informatics Assoc. 20.1 (2013), pp. 117–121.
- [124] E. J. Knobbe and B. Buckingham. "The Extended Kalman Filter for Continuous Glucose Monitoring". In: Diabetes Technol. Ther. 7.1 (2005), pp. 15–27.
- [125] D. J. Albers, M. Levine, B. Gluckman, H. Ginsberg, G. Hripcsak, and L. Mamykina. "Personalized glucose forecasting for type 2 diabetes using data assimilation". In: *PLOS Comput. Biol.* 13.4 (2017).
- [126] L. P. Kaelbling, M. L. Littman, and A. W. Moore. "Reinforcement Learning: A Survey". In: J. Artif. Intell. Res. 4 (1996), pp. 237–285.
- [127] J. Kober, J. A. Bagnell, and J. Peters. "Reinforcement learning in robotics: A survey". In: Int. J. Rob. Res. 32.11 (2013), pp. 1238–1274.

- [128] B. Baccot, R. Grigoras, and V. Charvillat. "Reinforcement Learning for Online Optimization of Banner Format and Delivery". In: Online Multimedia Advertising: Techniques and Technologies. Ed. by X.-S. Hua, T. Mei, and A. Hanjalic. IGI Global, 2011. Chap. 2, pp. 13–31.
- [129] A. Zhavoronkov, Q. Vanhaelen, and T. I. Oprea. "Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology?" In: Clin. Pharmacol. Ther. 107.4 (2020).
- [130] R. Bellman. *Dynamic programming*. Princeton: Princeton University Press, 1957.
- [131] S. Russell and P. Norvig. Artificial Intelligence. A Modern Approach. 3rd ed. New Jersey: Pearson Education Inc, 2010.
- [132] L. Kocsis and C. Szepesvári. "Bandit Based Monte-Carlo Planning". In: Mach. Learn. ECML 2006. Ed. by J. Fürnkranz, T. Scheffer, and M. Spiliopoulou. Vol. 4212. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2006, pp. 282–293.
- [133] A. J. Schaefer, M. D. Bailey, S. M. Shechter, and M. S. Roberts. "Modeling Medical Treatment Using Markov Decision Processes". In: Oper. Res. Health Care (2006), pp. 593–612.
- [134] O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts. "Markov Decision Processes: A Tool for Sequential Decision Making under Uncertainty". In: *Med. Decis. Mak.* 30.4 (2010), pp. 474–483.
- [135] C. Szepesvári. "Algorithms for Reinforcement Learning". In: Synth. Lect. Artif. Intell. Mach. Learn. 4.1 (2010), pp. 1–103.
- [136] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst. Reinforcement Learning and Dynamic Programming Using Function Approximators. Vol. 39. CRC Press, 2017.
- [137] C. J. C. H. Watkins. "Learning from Delayed Rewards". PhD thesis. University of Cambridge, 1989.
- [138] C. J. C. H. Watkins and P. Dayan. "Q-learning". In: Mach. Learn. 8 (1992), pp. 279– 292.
- [139] R. S. Sutton. "Learning to predict by the methods of temporal differences". In: Mach. Learn. 3.1 (1988), pp. 9–44.
- [140] H. Robbins and S. Monro. "A stochastic approximation method". In: Ann. Math. Stat. 22.3 (1951), pp. 400–407.
- [141] R. Coulom. "Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search". In: 5th Int. Conf. Comput. Games. Turin, Italy, 2006.
- [142] D. Silver et al. "Mastering the game of Go with deep neural networks and tree search". In: Nature 529 (2016), pp. 484–489.
- [143] C. B. Browne et al. "A survey of Monte Carlo tree search methods". In: IEEE Trans. Comput. Intell. AI Games 4.1 (2012), pp. 1–43.
- [144] P. Auer, N. Cesa-Bianchi, and P. Fischer. "Finite-time Analysis of the Multiarmed Bandit Problem". In: Mach. Learn. 47 (2002), pp. 235–256.
- [145] W. Hoeffding. "Probability Inequalities for sums of Bounded Random Variables". In: J. Am. Stat. Assoc. Vol. 58. 301. 1994, pp. 409–426.
- [146] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal. "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care". In: Nat. Med. 24.11 (2018), pp. 1716–1720.

- [147] R. Bartolucci, S. Grandoni, N. Melillo, G. Nicora, E. Sauta, E. M. Tosca, and P. Magni. "Artificial intelligence and machine learning : just a hype or a new opportunity for pharmacometrics ?" In: Page Meet. Stock. 2019.
- [148] B. Ribba, S. Dudal, T. Lavé, and R. W. Peck. "Model-Informed Artificial Intelligence: Reinforcement Learning for Precision Dosing". In: *Clin. Pharmacol. Ther.* 107.4 (2020), pp. 853–857.
- [149] B. W. Bequette. "A Critical Assessment of Algorithms and Challenges in the Development of a Closed-Loop Artificial Pancreas". In: *Diabetes Technol. Ther.* 7.1 (2005), pp. 28–47.
- [150] O. Gottesman et al. "Evaluating Reinforcement Learning Algorithms in Observational Health Settings". (preprint on arXiv). 2018.
- [151] O. Gottesman, F. Johansson, M. Komorowski, A. Faisal, D. Sontag, F. Doshi-Velez, and L. A. Celi. "Guidelines for reinforcement learning in healthcare". In: *Nat. Med.* 25.1 (2019), pp. 16–18.
- [152] F. Fröhlich, F. J. Theis, J. O. R\u00e4dler, and J. Hasenauer. "Parameter estimation for dynamical systems with discrete events and logical operations". In: *Bioinformatics* 33.7 (2016), pp. 1049–1056.
- [153] P. Stapor et al. "PESTO: Parameter EStimation TOolbox". In: Bioinformatics 34.4 (2018), pp. 705–707.
- [154] A. Kümmel, P. L. Bonate, J. Dingemanse, and A. Krause. "Confidence and Prediction Intervals for Pharmacometric Models". In: CPT Pharmacometrics Syst. Pharmacol. 7 (2018), pp. 360–373.
- [155] L. Wasserman. All of Statistics. Springer Texts in Statistics. Springer New York, NY, 2004.
- [156] J. Wakefield. "Bayesian individualization via sampling-based methods". In: J. Pharmacokinet. Biopharm. 24.1 (1996), pp. 103–131.
- [157] C. M. Bishop. Pattern recognition and machine learning. Springer New York, NY, 2006.
- [158] I. H. Jermyn. "Invariant Bayesian estimation on manifolds". In: Ann. Stat. 33.2 (2005), pp. 583–605.
- [159] M. Svensén and C. M. Bishop. Pattern Recognition and Machine Learning. Solutions to Exercises Web-Edition. 2009.
- [160] I. Netterberg. Kloft_2006_myelosuppression_docetaxel. 2016. repository.ddmore .eu/model/DDMODEL00000224 (visited on 09/12/2018).
- [161] A. Chaouch, R. Hooper, C. Csajka, V. Rousson, Y. Thoma, and T. Buclin. "Building up a posteriori percentiles for Therapeutic Drug Monitoring". In: 25th Popul. Approach Gr. Eur. (PAGE), Lisboa, Port. 2016.
- [162] S. B. Duffull, E. J. Begg, B. A. Robinson, and J. J. Deely. "A sequential Bayesian algorithm for dose individualisation of carboplatin". In: *Cancer Chemother. Pharmacol.* 39.4 (1997), pp. 317–326.
- [163] T. T. Le, F. Jost, T. Raupach, J. Zierk, M. Rauh, M. Stanulla, M. Metzler, and S. Sager. "A mathematical model of white blood cell dynamics during maintenance therapy of childhood acute lymphoblastic leukemia". In: *Math. Med. Biol. A J. IMA* 36.4 (2018), pp. 471–488.

- [164] T. Bengtsson, P. Bickel, and B. Li. "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems". In: Probab. Stat. Essays Honor David A. Free. Vol. 2. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2008, pp. 316–334.
- [165] J. de Wiljes, S. Reich, and W. Stannat. "Long-Time Stability and Accuracy of the Ensemble Kalman-Bucy Filter for Fully Observed Processes and Small Measurement Noise". In: SIAM J. Appl. Dyn. Syst. 17.2 (2018), pp. 1152–1181.
- [166] J. de Wiljes and X. T. Tong. "Analysis of a localised nonlinear ensemble Kalman Bucy filter with complete and accurate observations". In: *Nonlinearity* 33.9 (2020), pp. 4752–4782.
- [167] J. de Wiljes, S. Pathiraja, and S. Reich. "Ensemble Transform Algorithms for Nonlinear Smoothing Problems". In: SIAM J. Sci. Comput. 42.1 (2020), A87–A114.
- [168] F. Jost, S. Sager, and T. Le. "A Feedback Optimal Control Algorithm with Optimal Measurement Time Points". In: Processes 5.4 (2017).
- [169] C. D. Rosin. "Multi-armed bandits with episode context". In: Ann. Math. Artif. Intell. 61.3 (2011), pp. 203–230.
- [170] M. T. Huizing et al. "Pharmacokinetics of paclitaxel and carboplatin in a doseescalating and dose-sequencing study in patients with non-small-cell lung cancer. The European Cancer Centre." In: J. Clin. Oncol. 15.1 (1997), pp. 317–329.
- [171] R. J. Keizer, E. Dvergsten, A. Kolacevski, A. Black, S. Karovic, S. Goswami, and M. L. Maitland. "Get Real: Integration of Real-World Data to Improve Patient Care". In: *Clin. Pharmacol. Ther.* 107.4 (2020), pp. 722–725.
- [172] M. C. Dubinsky, B. L. Phan, N. Singh, S. Rabizadeh, and D. R. Mould. "Pharmacokinetic Dashboard-Recommended Dosing Is Different than Standard of Care Dosing in Infliximab-Treated Pediatric IBD Patients". In: AAPS J. 19.1 (2017), pp. 215–222.
- [173] L. B. Sheiner. "Learning versus confirming in clinical drug development". In: Clin. Pharmacol. Ther. 61.3 (1997), pp. 275–291.
- [174] N. Bleyzac et al. "Improved clinical outcome of paediatric bone marrow recipients using a test dose and Bayesian pharmacokinetic individualization of busulfan dosage regimens". In: *Bone Marrow Transplant.* 28.8 (2001), pp. 743–751.
- [175] S. G. Wicha, M. G. Kees, A. M. Solms, I. K. Minichmayr, A. Kratzer, and C. Kloft. "TDMx: A novel web-based open-access support tool for optimising antimicrobial dosing regimens in clinical routine". In: Int. J. Antimicrob. Agents 45.4 (2015), pp. 442–444.
- [176] D. M. Hughes, S. Goswami, R. J. Keizer, M. S. A. Hughes, and J. D. Faldasz. "Bayesian clinical decision support-guided versus clinician-guided vancomycin dosing in attainment of targeted pharmacokinetic parameters in a paediatric population". In: J. Antimicrob. Chemother. 75.2 (2020), pp. 434–437.
- [177] F. Fröhlich et al. ICB-DCM/AMICI: AMICI v0.10.19. 2020. https://doi.org/10 .5281/zenodo.3666715 (visited on 09/14/2018).
- [178] F. W. Ojara, A. Henrich, N. Frances, W. Huisinga, N. Hartung, M. Joerger, and C. Kloft. "Time-to-event analysis of paclitaxel-associated peripheral neuropathy in advanced non-small cell lung cancer highlighting key influential treatment/patient factors". In: J. Pharmacol. Exp. Ther. 53.9 (2020).

- [179] R. S. Sutton. "Dyna, an integrated architecture for learning, planning, and reacting". In: ACM SIGART Bull. 2.4 (1991), pp. 160–163.
- [180] D. Silver, R. S. Sutton, and M. Müller. "Sample-based learning and search with permanent and transient memories". In: Proc. 25th Int. Conf. Mach. Learn. - ICML '08. New York, NY, USA: ACM Press, 2008, pp. 968–975.
- [181] C. Maier, C. Loos, and J. Hasenauer. "Robust parameter estimation for dynamical systems from outlier-corrupted data". In: *Bioinformatics* (2017), btw703.
- [182] W. Zhao et al. "External evaluation of population pharmacokinetic models of vancomycin in neonates: the transferability of published models to different clinical settings". In: Br. J. Clin. Pharmacol. 75.4 (2013), pp. 1068–1080.
- [183] R. ter Heine, R. J. Keizer, K. van Steeg, E. J. Smolders, M. Luin, H. J. Derijks, C. P. Jager, T. Frenzel, and R. Brüggemann. "Prospective validation of a model-informed precision dosing tool for vancomycin in intensive care patients". In: Br. J. Clin. Pharmacol. 86.12 (2020), pp. 2497–2506.
- [184] L. B. Sheiner and T. M. Ludden. "Population Pharmacokinetics/Dynamics*". In: Annu. Rev. Pharmacol. Toxicol. 32.1 (1992), pp. 185–209.
- [185] A. N. Deitchman. "The Risk of Treating Populations Instead of Patients". In: CPT Pharmacometrics Syst. Pharmacol. 8.5 (2019), pp. 256–258.
- [186] J. R. Powell, J. Cook, Y. Wang, R. Peck, and D. Weiner. "Drug Dosing Recommendations for All Patients: A Roadmap for Change". In: *Clin. Pharmacol. Ther.* (2020).
- [187] A.-K. Hamberg, M.-L. Dahl, M. Barban, M. G. Scordo, M. Wadelius, V. Pengo, R. Padrini, and E. N. Jonsson. "A PK–PD Model for Predicting the Impact of Age, CYP2C9, and VKORC1 Genotype on Individualization of Warfarin Therapy". In: *Clin. Pharmacol. Ther.* 81.4 (2007), pp. 529–538.
- [188] M. Ohara et al. "Determinants of the Over-Anticoagulation Response during Warfarin Initiation Therapy in Asian Patients Based on Population Pharmacokinetic-Pharmacodynamic Analyses". In: *PLoS One* 9.8 (2014).
- [189] D. W. Uster, S. L. Stocker, J. E. Carland, J. Brett, D. J. Marriott, R. O. Day, and S. G. Wicha. "A Model Averaging/Selection Approach Improves the Predictive Performance of Model-Informed Precision Dosing: Vancomycin as a Case Study". In: *Clin. Pharmacol. Ther.* (2020).
- [190] J.-J. Mao, Z. Jiao, H.-Y. Yun, C.-Y. Zhao, H.-C. Chen, X.-Y. Qiu, and M.-K. Zhong. "External evaluation of population pharmacokinetic models for ciclosporin in adult renal transplant recipients". In: Br. J. Clin. Pharmacol. 84.1 (2018), pp. 153–171.
- [191] D. I. Jodrell, L. M. Reyno, R. Sridhara, M. A. Eisenberger, K. H. Tkaczuk, E. G. Zuhowski, V. J. Sinibaldi, M. J. Novak, and M. J. Egorin. "Suramin: development of a population pharmacokinetic model and its use with intermittent short infusions to control plasma drug concentration in patients with prostate cancer." In: J. Clin. Oncol. 12.1 (1994), pp. 166–175.
- [192] B. A. Conley, A. Forrest, M. J. Egorin, E. G. Zuhowski, V. Sinibaldi, and D. A. Van Echo. "Phase I Trial Using Adaptive Control Dosing of Hexamethylene Bisacetamide (NSC 95580)". In: *Cancer Res.* 49.12 (1989), pp. 3436–3440.

- [193] A. Kaefer, J. Yang, P. Noertersheuser, S. Mensing, R. Humerickhouse, W. Awni, and H. Xiong. "Mechanism-based pharmacokinetic/pharmacodynamic meta-analysis of navitoclax (ABT-263) induced thrombocytopenia". In: *Cancer Chemother. Pharmacol.* 74.3 (2014), pp. 593–602.
- [194] L. Pujo-Menjouet. "Blood Cell Dynamics: Half of a Century of Modelling". In: Math. Model. Nat. Phenom. 11.1 (2016), pp. 92–115.
- [195] M. Craig. "Towards Quantitative Systems Pharmacology Models of Chemotherapy-Induced Neutropenia". In: CPT Pharmacometrics Syst. Pharmacol. 6.5 (2017), pp. 293–304.
- [196] Z. Chen, N. Ma, and B. Liu. "Lifelong Learning for Sentiment Classification". In: Proc. 53rd Annu. Meet. Assoc. Comput. Linguist. 7th Int. Jt. Conf. Nat. Lang. Process. (Volume 2 Short Pap.) Stroudsburg, PA, USA: Association for Computational Linguistics, 2015, pp. 750–756.
- [197] D. L. Silver, Q. Yang, and L. Li. "Lifelong machine learning systems: Beyond learning algorithms". In: AAAI Spring Symp. - Tech. Rep. Vol. SS-13-05. 2013, pp. 49–55.
- [198] S. J. Pan and Q. Yang. "A Survey on Transfer Learning". In: IEEE Trans. Knowl. Data Eng. 22.10 (2010), pp. 1345–1359.
- [199] L. Torrey and J. Shavlik. "Transfer Learning". In: Handb. Res. Mach. Learn. Appl. Ed. by E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, and A. J. Serrano López. IGI Global, 2010. Chap. 11, pp. 242–264.
- [200] Jing Jiang. "A Literature Survey on Domain Adaptation of Statistical Classifiers". 2008.
- [201] J. H. Hughes, D. M. H. Tong, S. S. Lucas, J. D. Faldasz, S. Goswami, and R. J. Keizer. "Continuous Learning in Model-Informed Precision Dosing: A Case Study in Pediatric Dosing of Vancomycin". In: *Clin. Pharmacol. Ther.* (2020).
- [202] D. Lunn, J. Barrett, M. Sweeting, and S. Thompson. "Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis". In: J. R. Stat. Soc. Ser. C (Applied Stat.) 62.4 (2013), pp. 551–572.
- [203] M. B. Hooten, D. S. Johnson, and B. M. Brost. "Making Recursive Bayesian Inference Accessible". In: Am. Stat. (2019), pp. 1–10.
- [204] M. Joerger. "Quantitative Effect of Gender, Age, Liver Function, and Body Size on the Population Pharmacokinetics of Paclitaxel in Patients with Solid Tumors". In: *Clin. Cancer Res.* 12.7 (2006), pp. 2150–2157.
- [205] S. B. Duffull, L. E. Friberg, and C. Dansirikul. "Bayesian Hierarchical Modeling with Markov Chain Monte Carlo Methods". In: *Pharmacometrics*. Ed. by E. I. Ette and P. J. Williams. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2007, pp. 137–164.
- [206] P. O. Gisleskog, M. O. Karlsson, and S. L. Beal. "Use of prior information to stabilize a population data analysis". In: J. Pharmacokinet. Pharmacodyn. 29.5-6 (2002), pp. 473–505.
- [207] J. J. Wilkins et al. "Time-Varying Clearance and Impact of Disease State on the Pharmacokinetics of Avelumab in Merkel Cell Carcinoma and Urothelial Carcinoma". In: CPT Pharmacometrics Syst. Pharmacol. 8.6 (2019), pp. 415–427.
- [208] B. E. Huang, W. Mulyasasmita, and G. Rajagopal. "The path from big data to precision medicine". In: *Expert Rev. Precis. Med. Drug Dev.* 1.2 (2016), pp. 129–143.

- [209] R. J. Tyson, C. C. Park, J. R. Powell, J. H. Patterson, D. Weiner, P. B. Watkins, and D. Gonzalez. "Precision Dosing Priority Criteria: Drug, Disease, and Patient Population Variables". In: Front. Pharmacol. 11 (2020), pp. 1–18.
- [210] I. Irurzun-Arana, C. Rackauckas, T. O. McDonald, and I. F. Trocóniz. "Beyond Deterministic Models in Drug Discovery and Development". In: *Trends Pharmacol. Sci.* 41.11 (2020), pp. 1–14.
- [211] S. Vozeh and C. Steiner. "Estimates of the population pharmacokinetic parameters and performance of Bayesian feedback: A sensitivity analysis". In: J. Pharmacokinet. Biopharm. 15.5 (1987), pp. 511–528.
- [212] D. Mould, G. D'Haens, and R. Upton. "Clinical Decision Support Tools: The Evolution of a Revolution". In: *Clin. Pharmacol. Ther.* 99.4 (2016), pp. 405–418.
- [213] S. Tang, A. Modi, M. W. Sjoding, and J. Wiens. "Clinician-in-the-loop decision making: Reinforcement learning with near-optimal set-valued policies". (preprint on arXiv). 2020.
- [214] U.S. Food and Drug Administration. "Guidance for Industry and Food and Drug Administration Staff: Software as a Medical Device (SAMD): Clinical Evaluation". In: FDA Guid. (2017), pp. 1–32.
- [215] L. Keutzer and U. S. Simonsson. "Medical Device Apps: An Introduction to Regulatory Affairs for Developers". In: *JMIR mHealth uHealth* 8.6 (2020), e17567.
- [216] P. Stapor, F. Fröhlich, and J. Hasenauer. "Optimization and profile calculation of ODE models using second order adjoint sensitivity analysis". In: *Bioinformatics* 34.13 (2018), pp. i151–i159.